Journal of Experimental Psychology:
Learning, Memory, and Cognition
1990, Vol. 16, No. 3, 497–508

0278-7393/90/$00.75

Why Are Conjunctive Categories Overextended?

Nick Chater
Centre for Cognitive Science
University of Edinburgh
Edinburgh, Scotland

Karen Lyon
Department of Psychology
University of Edinburgh
Edinburgh, Scotland

Terry Myers
Centre for Cognitive Science and Department of Psychology
University of Edinburgh, Edinburgh, Scotland

Recent experiments have suggested that subjects tend to overextend conjunctive categories (Hampton, 1988). We present a series of four experiments that confirm this finding. In Experiment 1, we use Hampton's response scale, on which subjects rate both membership and typicality. In Experiment 2, we find that subjects still overextend their categories when they judge membership alone. In Experiment 3, we introduce more response options and conclude that the tendency to overextend is not an artifact of an insufficient range of possible responses. We propose an explanation of overextension that we term the *compensation hypothesis*: The more categories that make up a conjunction, the more leniently membership is judged. We argue that this is a result of having to make "best fit" judgments for multiple constraints in real life. We test this hypothesis in Experiment 4, which uses conjuncts of three categories. As predicted, we find that overextensions are greater with triple conjunctions than in appropriate controls in which two categories are combined or categories are judged alone. We consider the theoretical implications of a compensation strategy for categorization.

Much recent experimental and theoretical work has focused on conceptual combination (Barsalou, 1982; Hampton, 1987; Lakoff, 1975; Osherson & Smith, 1981; Roth & Mervis, 1983; Smith & Medin, 1981). For example, how are judgments of membership of a conjunctive category (e.g., PET FISH) related to judgments of memberships of its constituents (PET and FISH)?

Traditional approaches to formalizing conjunction in natural language semantics (e.g., Dowty, Wall, & Peters, 1981) have typically adopted the Boolean hypothesis. That is, that the set-theoretic intersection of the set of members of each constituent forms the set of members of the conjunction. Thus, for each item, membership of a complex concept is a Boolean function of membership of the constituent concepts. The Boolean hypothesis predicts that an item is judged to be a SPORT WHICH IS ALSO A GAME if and only if it is judged to be both a SPORT and a GAME. In such an account, there is no room for graded membership: An item is either a member of the set corresponding to the category, or it is not.

An alternative view, popular among prototype theorists, is that membership judgments are products of the same underlying processes as typicality judgments. Hampton has called this position the *unitary view*. Because a standard set-theoretic approach is not appropriate for modeling graded structure, the unitary view contradicts a Boolean account of concept combination. Hampton (1988) reports findings that suggest subjects do not judge membership in conjuncts according to the Boolean hypothesis, and thus he argues for a unitary model of concept combination.

Hampton had subjects assess whether or not lists of items are members of a category and how related the items are to that category (Stage 1). A week later subjects assessed the membership and relatedness of the same items to a conjunction of two of the categories (Stage 2). For example, at Stage 1, a subject might decide that chess was a member of the category GAME (+), but not of the category SPORT (−). At Stage 2, the subject might decide that chess is not a SPORT WHICH IS ALSO A GAME (−). This triple of responses is represented (+−−).

A difficulty with the interpretation of these experiments is that the unitary hypothesis is inadvertently built in to the rating scale, assuming the unitary hypothesis rather than confirming it. Using such a rating scale is thus not a fair test of the Boolean hypothesis. The rating scale, design, and the analysis of Hampton's original experiments are now examined in more depth.

Rating Scale

The rating scale used may bias subjects' responses towards those predicted by the unitary hypothesis. In order to assess

---

The first two authors are joint principal authors.

This research was funded by Economic and Social Sciences Research Council Award C00428622023 to Nick Chater and Science and Engineering Research Council Award 85310119 to Karen Lyon. The experiments reported in this article are part of a doctoral thesis by Karen Lyon (Lyon, 1989).

We gratefully acknowledge the help of James Hampton in providing his original stimuli and for helpful comments and suggestions in his review. Thanks are also due to Douglas Medin, James Hampton, and several anonymous reviewers for incisive and helpful comments on earlier drafts of this article. We are indebted to Frances Provan for her advice on statistics, and we thank Malcolm Lyon for formalizing the proof. We would also like to thank members of the Mental Lexicon Workshop for useful discussion.

Correspondence concerning this article should be addressed to Nick Chater, Department of Psychology, University College London, Gower Street, London WC1E 6BT, United Kingdom.

whether or not typicality and membership are aspects of the same phenomenon, it is necessary to have independent measures of both. Because the rating scale measures typicality and membership together, we might expect the membership task to distort the typicality ratings, and the typicality task to distort the membership ratings in the following ways:

1. The requirement to judge typicality may bias subjects to adopt a similarity-based categorization strategy. Hence, the results might be biased towards those predicted by the unitary hypothesis.

2. The scale does not allow items to be rated in a way that violates the unitary hypothesis. Thus it is impossible to give a nonmember a typicality/relatedness rating higher than −1. Similarly, it is impossible to give a member a typicality/relatedness rating lower than 1. Hence, members are necessarily given higher typicality/relatedness ratings than are nonmembers. A violation of the unitary hypothesis would be obtained if some nonmember was rated as more typical/related than some member. For example, subjects might want to rate *tomato* as a typical VEGETABLE, although the subject knows it is a FRUIT. In particular, they might want to give *tomato* a higher typicality/relatedness rather than an atypical vegetable such as *yam*. In such a case, the subject could not simultaneously respect intuitions about typicality and membership. If membership takes precedence, then the subject is forced to give a lower typicality/relatedness rating to *tomato* (say, −1 denoting a related nonmember) than for *yam* (say, +1 denoting an atypical member). Alternatively, if typicality takes precedence, then *tomato* must be given a higher typicality/relatedness rating than *yam*. This means that if *yam* is judged to be a vegetable, then *tomato* will be judged to be a more typical vegetable. On the other hand, if *tomato* is judged to be related nonmember of VEGETABLE, then *yam* must also be a (less related) nonmember. A unitary theorist might respond that there are no such cases. However, this is just the unitary hypothesis.

Design and Analysis

The number of responses expected in each response class (e.g. ++−, −++) is estimated rather than measured directly. Certain non-Boolean responses would be expected even if subjects use a Boolean combination rule, given that (a) subjects remember their previous responses with probability $u < 1$, and (b) each membership judgment for the constituents may be probabilistic. For example, a subject might judge an item to be a SPORT and a GAME at Stage 1. At Stage 2, the subject might not remember the previous judgments and now decide that the item is not in fact a GAME. If the subject is using a Boolean stretegy, the item will not be judged a member of the conjunct. Hence, non-Boolean results do not necessarily imply non-Boolean combination.

Because there is no direct measure of how much subjects remember, Hampton has provided an estimate of a remembering parameter u. This is the probability of subjects remembering their previous judgments. Hampton (1988, p. 17) explains: "The parameter u was estimated from the two observed frequencies for +++ and ++−, and constrained to lie between 0 and 1. If it could be estimated it was set to zero." Using his estimate for u, we derived the probability of each fresh membership judgment. Hampton takes into account the possibility that the order in which items are rated at Stage 1 may influence the probability that an item is judged to be a category member. For example, an item might be more likely to be rated as a MACHINE if this judgment is made before, rather than after, it is rated as a VEHICLE. Given that order was randomized in our experiments, it was possible to ignore the order of rating, which Hampton included as a factor.[1]

Ignoring the order in which categories are rated at Stage 1, we can adapt Hampton's formula for estimating probabilities at Stage 2, according to a Boolean hypothesis. Hence, the probability of classifying an item as a member of the conjunct, when previously classifying it as a member of the first constituent but not of the second (+−+), is

$$p(+-+) = s.(1-g).[u^2.(0) + u.(1-u).g + (1-u).u.(0) + (1-u)^2.s.g]$$

s = probability of an item being rated as a sport
g = probability of an item being rated as a game
u = probability of remembering the classification at Stage 1.

By using formulas of this sort, Hampton was able to estimate the number of responses that should be obtained in each response category, assuming subjects use a Boolean conjunction strategy. Hence, the results obtained could be compared with the estimates on the basis of the Boolean model, allowing an assessment of the degree to which subjects actually used a Boolean strategy.

Hampton estimated the parameters s, g, and u from the data and showed that a Boolean model did not provide a good fit for the frequencies observed in the various response categories. An alternative to estimating s, g, and u from the data and fitting a stochastic Boolean model is to introduce an appropriate control condition. Comparison of the experimental and control conditions gives a direct measure of the degree to which subjects perform according to a Boolean strategy. This controls for within-subject differences over time, which have been shown to be considerable (Barsalou, 1987; McCloskey & Glucksberg, 1978).

In addition to this control, we derive a new prediction of the Boolean hypothesis from the probabilistic model which is independent of the values of u, s, and g. By using Hampton's

[1] Hampton used separate parameters s_1, s_2, g_1, and g_2, which denoted the probability that an item is rated as member when judged first or second at Stage 1. Because there is the possibility that judgments at Stage 1 influence judgments at Stage 2, Hampton uses only s_1 and g_1 when predicting the probability that an item is rated as part of the conjunction of s and g. We, on the other hand, take the mean of s_1 and s_2 and the mean of g_1 and g_2 when predicting the probability that an item will be judged a member of the conjunct. Hampton suggested that we test whether or not this difference in strategy was important by examining the effect of rating order at Stage 1. As detailed below, we found that order effects do not seem to be either large or, except in Experiment 1, significant (i.e., s_1 is not significantly different from s_2). Hence, the two strategies are in practice equivalent.

formulas we find that the probability of an underextension, $p(++-)$, is equal to the probability of an overextension, $p(--+) + p(-++) + p(+-+)$, independent of the particular values of s, g, and u (Appendix A). Hence, if the number of overextensions and underextensions is significantly different, the Boolean hypothesis is disconfirmed. Of course, if the number of overextensions and underextensions is not significantly different, this does not allow us to differentiate the unitary hypothesis from the Boolean hypothesis. However, inasmuch as the frequency of overextensions and underextensions differs, we have a simple measure of the degree of non-Booleanness. In the experiments reported below, we use the difference between overextensions and underextensions to assess the degree of non-Boolean combination. The experiments reported below are designed to address these issues empirically.

Experiment 1

Experiment 1 was a replication of the principle features of Hampton's (1988) Experiment 4. We used the same task, rating scale, and items as used in Hampton's Experiment 4. Because our primary concern was to assess the magnitude of overextension and underextension, rather than effects of constituent order, we counterbalanced the order of constituents in each conjunct. The order of rating of categories at Stage 1 and Stage 2 was random. Our analysis was based on the relative number of overextensions and underextensions, as argued above, rather than following Hampton's original analysis. The purpose of this replication was to provide an appropriate comparison with the other experiments in this series.

Method

Design. The experiment was a within-subjects design. Response type (overextensions and underextensions) was treated as a within-subjects factor. The same subjects rated items at Stage 1 and Stage 2. The same random order of items within a list was used for all subjects at both stages. The order in which the lists were rated was fully randomized across subjects. The role of categories as head and qualifier was counterbalanced.

Subjects. Ten subjects participated in Experiment 1. All subjects were in full-time education in Edinburgh and took part in the experiment as a course requirement. Two subjects were excluded from the experiment because of absence at one of the sessions. The sexes were approximately equally represented in the group, and subjects were naive to the purposes of the experiment.

Materials. Materials were identical to those used by Hampton (1988). Six conjunctions were used: machines–vehicles, furniture–household appliances, pets–birds, buildings–dwellings, food–plants (or part of a plant), and weapons–tools. For each conjunction a list of 16 items was available. This list contained items belonging to both constituents, neither constituent, or only one constituent. Items were typed in random order on a sheet headed by the appropriate category name—for example, machine, vehicle, machine which is also a vehicle, and vehicle which is also a machine. To the right of each item was a 7-point scale from −3 to +3. The instruction sheet included an appropriate completed example not used in the experiment.

Procedure. The experiment was conducted in two stages, a week apart. In the first stage, subjects were given a booklet containing 12 pages of test items, presented in a different random order for each subject. On the front of each booklet were the following instructions, which are taken from Hampton (1988):

> On the following pages you are asked to make a series of judgements about the everyday common usage of words. In each case the question to be answered is whether or not the 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.

Each page was headed by one of the constituent categories—for example, bird. Subjects were asked to decide the membership and typicality of the items. First, subjects had to decide if the item was a member of the category. If the item was judged to be a member, it was given a "+" rating. If the item was judged not to be a member, it was given a "−" rating. If undecided, the subject could use "0," but this was discouraged. Second, subjects had to decide the typicality of members of the category by choosing +1, +2, or +3, where +3 is a very typical member. If the item was a nonmember, subjects indicated how related it was to the category by choosing −1, −2, or −3, where −1 is the most related nonmember. Subjects completed the ratings at their own pace. Stage 2 was the same as Stage 1 save that subjects rated the 6 conjunctions rather than the 12 constituent categories. The ordering of the conjunctions ("pets which are also birds" vs. "birds which are also pets") was balanced across subjects so that half the subjects received one random ordering, and half the reverse ordering.

Results and Discussion

As argued above, a significant difference between the number of overextensions and underextensions subjects make is indicative of a non-Boolean combination strategy. Summed over all subjects, 2.9% of the responses were underextensions and 15.6% were overextensions. That is, subjects were more likely to classify an item as a member of a conjunct when not a member of a constituent than to classify an item as a nonmember of a conjunct when a member of both constituents. The number of overextensions and underextensions for each category is summarized in Table 1. The pattern of overextensions and underextensions may be seen to be similar for all categories except BIRD, for which there are no single overextensions or underextensions. Judgments for the category BIRD were consistent except for one double overextension for "mynah bird" for the category PETS WHICH ARE ALSO BIRDS.

A Wilcoxon signed-rank test was used in the analysis of Experiment 1.[2] There was a significant difference between

[2] Overextensions and underextensions are not strictly independent (i.e., a particular item cannot be both overextended and underextended by the same subject). Because the number of overextensions and underextensions is also frequency data, and thus approximates to a Poisson distribution, this necessitates the use of nonparametric tests. Where multivariate analyses are required, traditional analysis of variance can be replaced by fitting a logistic model. The residuals from the fit of this model provide main effect and interaction information comparable to that provided by analysis of variance without necessitating assumptions of normality.

Table 1
Overextensions and Underextensions for Category Pairs: Experiment 1

Category		Underextensions[a]	Overextensions[b]		
A	B	++−	+−+	−++	−−+
Furniture	House. appl.	5	12	11	1
Food	Plant	2	7	11	1
Weapon	Tool	2	13	13	3
Building	Dwelling	6	9	10	3
Machine	Vehicle	7	4	7	3
Bird	Pet	0	11	0	1

Note. Overextensions and underextensions as percentages of the total number of responses are shown in parentheses below. The triple of pluses and minuses denotes rating for Category A, rating for Category B, and rating for conjunction, irrespective of the order in which the categories were presented. House. appl. = household appliance.
[a] Total underextensions = 22 (2.9%). [b] Total overextensions = 120 (15.6%).

overextensions and underextensions ($W = 0$, $N = 8$, $p < .05$). This replicates Hampton's finding that subjects exhibit non-Boolean responses to membership decisions in conjuncts. We found that subjects preferred to overextend rather than underextend their concepts. The pattern of overextensions and underextensions is summarized in Table 2. Each entry in the table indicates the mean response to the conjunction for the corresponding values of the head and qualifier. For example, we consider together all the responses in which the head was rated −2 and the qualifier was rated +3 at Stage 1. The mean value of the corresponding conjunctions at Stage 2, in these cases, was found to be 0.47. A strictly Boolean model predicts that all and only the entries in the top left quadrant of Table 2 should be positive (i.e., only items that are members of the conjunct are members of both constituents). If underextensions were dominant, negative values would encroach upon the top left quadrant. If the overextensions dominated, positive values would expand into the other quadrants of the table (where one or more of the constituents was deemed to be a nonmember). As can be seen from Table 2, the positive values, in fact, expand out of the top left quadrant.

As Hampton found, there is a marked asymmetry between the influence of the categories, depending on whether they are in head or qualifier position. The rating of the qualifier category is more closely related to the rating of the conjunctive category. This is shown in Table 2 by the fact that positive values encroach into the lower left quadrant—that is, on average, the positive rating of the qualifier "overrides" the negative rating of the head. However, throughout the top right quadrant, which represents negative qualifier ratings and positive head ratings, the average rating for the conjunction is negative; that is, the negative rating for the qualifier "overrides" the positive rating for the head. This indicates that the qualifier is more important in determining conjunct membership than the head. If an item is a member of the qualifier category, it is likely to be rated as a conjunct member even if it is a (related) nonmember of the head category. However, if an item is not a member of the qualifier category, it is unlikely to attain membership of the conjunct. This effect is independent of the particular categories rated because all categories were presented in both qualifier and head positions.

Table 2
Mean Ratings for Conjunctions as a Function of Head and Qualifier Rating: Experiment 1

	Qualifier						
Head	+3	+2	+1	0	−1	−2	−3
+3	2.77	2.08	1.62	0.83	−0.03	−0.79	−1.9
+2	1.94	1.11	1.64		−0.8	−0.29	−1.5
+1	1.87	0.6	0		−0.6		−1.45
0							
−1	0.2	0.54	0.11		−1		−2.71
−2	0.47	−1.57					−2.11
−3	−1.34		−1.18	−1.33	−1.78	−1.71	−2.82

Note. Only values for $n > 4$ are shown.

Hampton notes that the *guppy* effect (a guppy is a better PET FISH than it is a PET or a FISH) is hard to find. That is, membership rating in the conjunction is seldom greater than in either constituent. We found 23 instances (3%) of this effect in our data, similarly a low figure.

As argued above, it is as yet unclear whether the source of these non-Boolean effects is the use of a non-Boolean combination strategy. Experiment 2 was designed to remove the possibility that typicality judgments may bias membership decisions to be less Boolean.

Experiment 2

In this experiment, subjects were required to make membership judgments only. This should eliminate any non-Boolean effects due to the confounding of membership with typicality. In addition, we introduced a control group who judged single categories rather than conjunctive categories at Stage 2. This allowed us to predict how they would have classified the conjunct if they had been using a Boolean rule. This provided a baseline estimate of overextensions and underextensions against which the results of the experimental group might be compared. This second condition provided a direct estimate of spurious non-Boolean effects, thus circumventing the problems of estimating u, s, and g as outlined above.

Method

Design. The experiment was a 2 × 2 mixed design. Response type (overextensions and underextensions) was treated as a within-subjects factor. The between-subjects factor was experimental group. Otherwise, the design was the same as that of Experiment 1.

Subjects. Eighteen subjects took part in the experiment. All subjects were full-time students in Edinburgh. Two of these subjects were excluded from the analysis because they were absent at one of the sessions. The sexes were approximately equally represented, and subjects were naive to the purposes of the experiment.

Procedure. The procedure was identical to that of Experiment 1, except that subjects rated membership only by indicating yes or no for each item. The instructions were adapted from Experiment 1:

> On the following pages you are asked to make a series of judgments about the everyday common usage of words. In each case the question to be answered is whether or not the general category name can be applied to a particular example. For each

example, decide whether you will answer "Yes" or "No," and ring the appropriate response. If you are unable to decide, do not ring either response, but avoid doing this as much as possible. If you are unfamiliar with any of the examples, cross them out.

The control subjects completed Stage 1 in the same way as the experimental subjects, and at Stage 2 they simply repeated Stage 1. The control group's responses at Stage 2 were conjoined Booleanly, after the experiment, to yield response triples comparable to those in the experimental group. Suppose that a subject rated an item as a member of both constituent categories at Stage 1 (++), then rated the same item as a member of just one of the constituents at Stage 2 (+−). A Boolean model of the conjunction would imply that the subject would say that the item was not a member of the conjunct at Stage 2 (−). We now construct the response triple (Constituent 1, Constituent 2, conjunction). In this case it is (++−), which is an underextension. For example, at Stage 1 a subject might judge snooker to be a SPORT and also a GAME (++). At Stage 2, the same subject might now say that snooker is not a SPORT (−) but is a GAME (+). Booleanly conjoining the responses at Stage 2 yields (−). Hence, the subject is deemed to have judged snooker to be a SPORT, a GAME, but not a SPORT WHICH IS ALSO A GAME (++−). This is an underextension. This procedure provided an estimate of overextensions and underextensions due to forgetting and changing one's mind. By directly comparing the two groups we could decide if the overextensions and underextensions in the experimental group resulted from a non-Boolean method of conjunction. This obviated the need to estimate u, s, and g.

Results and Discussion

The number of overextensions and underextensions for the experimental and control conditions for each category is shown in Table 3. Whereas in the control condition the number of overextensions and underextensions is similar (5.4% vs. 4.2%), in the experimental condition there are more than three times as many overextensions as underextensions (12.0% vs. 3.1%). A Spearman's rank correlation between items overextended in both conditions was significant (r = .52, $n = 12$, $p < .05$). This indicates that some categories are consistently more flexible that others, and this leads to greater variability in categorization.

The frequency of overextensions and underextensions was fitted to a 2 × 2, Response Type (overextension vs. underextension) × Group (control vs. experimental condition), logistic model. There was a significant main effect of group (χ^2 = 9.9, $df = 1$, $p < .05$) and response type ($\chi^2 = 32.3$, $df = 1$, $p < .05$). We also found a significant interaction ($\chi^2 = 11.3$, $df = 1$, $p < .05$). Thus subjects in the experimental condition made significantly more overextensions.

An examination of the residual errors from the fit of the logistic model shows[3] that, as expected, there was no difference between the number of overextensions and underextensions in the control condition (residual = 0.756, $df = 1$, $p > .05$). However, there was a significant difference between the frequency of overextensions and underextensions in the experimental condition (residual = 3.127, $df = 1$, $p < .05$). This is in contradiction to the Boolean hypothesis, which predicts equal numbers of overextensions and underextensions in both conditions.

Table 3

Overextensions and Underextensions as a Function of Category and Experimental Condition: Experiment 2

Category		Underextensions[a]	Overextensions[b]		
A	B	++−	+−+	−++	−−+
		Experimental group			
Furniture	House. appl.	3	13	15	3
Food	Plant	4	6	12	0
Weapon	Tool	6	15	2	2
Building	Dwelling	6	0	3	0
Machine	Vehicle	4	2	8	1
Bird	Pet	1	10	0	0
		Control group			
Furniture	House. appl.	7	6	3	0
Food	Plant	6	2	7	0
Weapon	Tool	13	3	1	0
Building	Dwelling	3	3	1	1
Machine	Vehicle	2	0	10	1
Bird	Pet	1	3	0	0

Note. In the experimental group, subjects judged membership of conjunctive categories at Stage 2. In the control condition, subjects rated single categories both at Stage 1 and Stage 2. Overextensions and underextensions as percentages of the total number of responses are shown in parentheses in the footnotes. The triple of pluses and minuses denotes rating for Category A, rating for Category B, and rating for conjunction, irrespective of the order in which the categories were presented. House. appl. = household appliances.
[a] Total underextensions in experimental group = 24 (3.1%), in control group = 32 (4.2%). [b] Total overextensions in experimental group = 92 (12.0%), in control group = 41 (5.4%).

Table 4 shows the number of overextensions for binary judgments as a function of position in the conjunction in Experiment 2. For example, if an item was not judged to be a PET but was judged to be a PET WHICH IS ALSO A BIRD, then it counted as a overextension in the first position (head noun). If an item was not judged to be a BIRD but was judged to be a PET WHICH IS ALSO A BIRD, then it counted as an overextension in the second position (qualifier noun). In the case where an item was judged to be neither a PET nor a BIRD, but judged to be a member of the conjunct a PET WHICH IS ALSO A BIRD, it counted an overextension in the first and second positions. As can be seen from the table, there are a greater number of overextensions in the first position (61) than in the second position (48). Hampton called this effect the *head-qualifier effect*. However, in this experiment the differences between the number of overextensions in first and second position are not significant ($W = 3$, $N = 4$, $p > .05$).

We conclude that there is a strong remaining non-Boolean effect even when membership is rated independently of typicality. There remains the possibility that the non-Boolean effects observed may be due to the restricted range of responses available to the subjects. This possibility was examined in Experiment 3.

[3] In fitting a logistic model, residual errors may be treated as χ values with 1 degree of freedom. Therefore, absolute values above 2.6 reach significance at $p < .05$. Because all the residuals are generated by a single logistic model, problems with multiple testing do not apply.

Table 4
Number of Overextensions for Binary Judgments as a Function of Position in the Conjunction: Experiment 2

	Position	
Category	1st	2nd
Furniture, household appliance	20	16
Food, plant	11	11
Weapon, tool	18	5
Building, dwelling	3	0
Machine, vehicle	4	11
Bird, pet	5	5
Total	61	48

Experiment 3

Experiment 3 was a variant of Experiment 2, with a wider choice of response options at Stage 2. The previous experiment addressed the possibility that membership judgments are biased by typicality judgments. A further possible problem is that subjects may be forced to overextend because of an insufficient range of response options. Subjects might feel unhappy about giving an item a negative rating if it was a good member of one of the constituent categories. For example, *chess* might be judged not to be a member of SPORT, but because it is such a good member of GAME, it might be included in the category SPORT WHICH IS ALSO A GAME. Thus, it may be that categories are overextended in order to express that an item which is a member of one constituent category (such as chess) is a better member of the conjunct than an item which is not a member of either constituent category (such as watching television). To obviate this possible spurious non-Boolean effect, it is necessary to introduce a wider range of response options: A GAME WHICH IS ALSO A SPORT, JUST A GAME, JUST A SPORT, or NEITHER A GAME NOR A SPORT. Hence, subjects may judge *chess* to be JUST A GAME.

Method

The method and procedure were the same as that used in Experiment 2, except that subjects had a wider choice of responses at Stage 2. The task was amended as follows: As before, at Stage 1, subjects had two response options (e.g., "is a VEHICLE" and "is not a VEHICLE"). At Stage 2 four response options for the conjunction were given (e.g., "is a VEHICLE WHICH IS ALSO A MACHINE," "is just a VEHICLE," "is just a MACHINE," and "is neither a VEHICLE nor a MACHINE"). Subjects indicated their response by indicating the appropriate number. As argued above, the greater choice of responses at Stage 2 should reduce any non-Boolean effects due to a lack of appropriate response options.

Subjects. Ten subjects participated in Experiment 3. They were students in full-time education in Edinburgh who participated in the experiment as part of their course work. Two subjects were excluded from the experiment because they were not native English speakers. The sexes were approximately equally represented in the group, and subjects were naive to the purposes of the experiment.

Results and Discussion

Overextensions and underextensions for each category are presented in Table 5. Giving the subjects more choice did not significantly alter the number of overextensions (10.0%) and underextensions (3.4%) from Experiment 2 (12.0% vs. 3.1%). Mann-Whitney U tests showed no significant difference between underextensions in Experiments 2 and 3 ($U = 17.5$ [8], $p > .05$) or between overextensions in Experiments 2 and 3 ($U = 30$ [8], $p > .05$).[4] However, a significant difference still remained between Experiment 3 and the control group of Experiment 2 ($\chi^2 = 6.68$ [1], $p < .05$). Hence, there is a large residual non-Boolean effect. The disparity between overextensions and underextensions cannot be attributed to the lack of response choices. Comparison with the control shows that there is a real non-Boolean component to membership judgments for conjunctions.

Although the results here are significantly more Boolean than in Experiment 1 ($\chi^2 = 4.44$ [1], $p < .05$), the preference for overextension appears to be robust under this manipulation of response option. This finding argues against a response bias explanation of overextensions. Further, Hampton (1988) reports that manipulation of the proportion of members and nonmembers, and therefore of the proportion of positive and negative responses made by subjects, appears to have little influence on the tendency to overextend. These findings argue that overextension is a result of conceptual combination per se.

Discussion

To summarize, Hampton's non-Boolean results were replicated in Experiment 1 (overextensions 15.6%, underextensions 2.9%). The tendency to overextend is found to be robust when the rating of membership is separated from the rating of typicality. Thus the results of Experiment 2 were still significantly different from those predicted by a Boolean model, as shown by comparison with the control condition. This difference was not significantly diminished by introducing a wider range of response options in Experiment 3.

Why is it that subjects are so strongly biased in favor of overextending their categories? The way in which conjunctive categories are actually used in everyday life suggests that subjects are more lenient in their membership judgments the more factors they have to take into account. In finding the best fit for a list of properties, items narrowly failing on one component are not usually excluded. For example, suppose that you have the task of casting the lead role in a local play. The character is a tall, handsome, blond male with a convincing Scottish accent, and a retentive memory. Given that such people are rather rare, you will, of course, compromise on certain dimensions—you might settle for a sandy haired man of average height. This amounts to overextending the

[4] These between-experiments comparisons seem justified because the experiments were carried out simultaneously under the same conditions.

Table 5
Overextensions and Underextensions (Underext.) as a Function of Category: Experiment 3

Category A	Category B	Underext.[a] ++−	Overextensions[b] +−+	−++	−−+
Furniture	House. appl.	5	6	12	1
Food	Plant	8	3	12	1
Weapon	Tool	5	8	8	2
Building	Dwelling	0	1	3	0
Machine	Vehicle	5	6	7	1
Bird	Pet	3	6	0	0

Note. Overextensions and underextensions as percentages of the total number of responses are shown in parentheses below. The triple of pluses and minuses denotes rating for Category A, rating for Category B, and rating for conjunction, irrespective of the order in which the categories were presented. House. appl. = household appliance.
[a] Total underextensions = 26 (3.4%). [b] Total overextensions = 77 (10.0%).

category. You would be especially likely to do this if he fulfilled the rest of your criteria. We call this intuitive explanation of overextensions the *compensation hypothesis.* Such effects can be seen in all uses of real conjunctions of properties: for example, when recognizing a stranger from a description (tall, dark, with a red carnation). When buying a house, a person may have a long list of criteria that must be simultaneously fulfilled. For instance: affordable price, two bedrooms, separate kitchen, central location, garage, garden. Because there are so few houses that meet all of these criteria, most people have to compromise. It is interesting to note that there is a tendency not to admit that the initial criteria have not been met. Instead, people extend their categories to incorporate instances not previously included. If subjects overextend conjunctive categories in order to find a best match, the more criteria that must be simultaneously fulfilled, the more overextensions we should expect.

Experiment 4

Experiment 4 was designed to test the compensation hypothesis by examining the way subjects conjoin three categories. This experiment built on Experiment 2. In Experiment 2, the control worked as follows: At Stage 1, all subjects gave categorization judgments for single categories. At Stage 2, the experimental group gave judgments for the conjunctions; the control group repeated Stage 1. The control group responses at Stage 2 were then Booleanly conjoined. This provided an estimate of the overextensions and underextensions that may be a result of forgetting and changing one's mind, rather than a result of a non-Boolean combination strategy. Thus, this condition provided a direct estimate of spurious non-Boolean effects. Using this idea, we designed controls appropriate to the conjunction of three categories. These controls enabled the results from triple conjunctions to be directly compared with single category judgments and binary conjunctions.

The compensation hypothesis predicts that the greater the number of constituent categories conjoined, the greater the extent to which subjects overextend their categories. Thus, judging membership of the category A FOOD, A PLANT, AND A FLAVORING should lead to more overextensions than judging membership of, say, A FOOD AND A FLAVORING.

Method

Design. The design was a 2 × 3 mixed design. Response type (overextensions and underextensions) was treated as a within-subject factor. The between-subjects factor is experimental group (1, 2, or 3 constituents). For each set of categories conjoined (e.g. FOOD, PLANT, FLAVORING) the order of the constituents was counterbalanced across subjects in such a way that 2 subjects rated each of the six possible orders in Groups 2 and 3. Otherwise, the design was identical to that of Experiment 2.

Subjects. Thirty-six subjects took part in the experiment, 12 in each of three groups. Subjects were students resident in Edinburgh who participated in the experiment voluntarily.

Materials. Stimuli were adapted from the items and categories used in the previous experiments. We added a third category to every pair of categories used previously. Six category triples were used: machine–vehicle–toy, furniture–household appliance–luxury, pet–bird–predator, building–dwelling–business, food–plant–flavouring, weapon–tool–farm equipment. The parenthetical "or part of a plant" used in the previous experiments, following Hampton, was omitted as unwieldy. A slight difference in syntax was employed. Extending the previous syntactic form (A WEAPON WHICH IS ALSO A TOOL) would have been cumbersome (A WEAPON WHICH IS ALSO A TOOL WHICH IS ALSO FARM EQUIPMENT). Hence, we adopted the simpler form A WEAPON AND A TOOL for the binary case, and A WEAPON, A TOOL, AND FARM EQUIPMENT for triple conjunctions. A list of 16 items was adapted from the previous list so that there were approximately equal numbers of items likely to produce each kind of response (i.e., we included 2 items that we considered to be likely members of all categories [+++]; 2 items that we considered to be likely members of the first two categories [++−], and so on, for all 8 possibilities). Thus, we designed the materials so that roughly half of the items were judged to be members of each single category (although, as noted above, Hampton, 1988, found that this proportion does not seem to be important). Items were typed in random order and headed by the appropriate category name—for example, a machine; a vehicle; a toy; a machine and a vehicle; a vehicle and a machine; a machine and a toy; a toy and a machine; a vehicle and a toy; a toy and a vehicle; a machine, a vehicle, and a toy; a machine, a toy, and a vehicle; a vehicle, a toy, and a machine; a vehicle, a machine, and a toy; a toy, a machine, and a vehicle; a toy, a vehicle, and a machine. We prepared lists of all possible permutations to counterbalance for order effects.

The lists and categories used in Experiment 4 overlapped significantly with those in the previous experiments. However, because new categories and items were introduced, and thus different numbers of items were to be judged to be a member of each category, the percentages of overextensions and underextensions were not directly comparable with those of Experiments 1–3. To avoid confusion, we quote overextensions and underextensions as numbers rather than percentages.

Procedure. The design of the experiment at Stage 2 is summarized in Table 6. Each of the three groups made single category judgments at Stage 1 (e.g., A FOOD, A FLAVORING, A PLANT). Group 1 was a control group and also performed single-category judgments at Stage 2. The Stage 2 judgments for each category triple were Booleanly conjoined, as in Experiment 2. For example, at Stage 1, a subject might judge *tomato* to be a FOOD, A PLANT, but not a FLAVORING (++−). At the second stage they might judge *tomato* to be a member of all three categories (+++). The Boolean conjunction of the re-

Table 6
Task Summary for Experiment 4

Group	Stage 1	Stage 2
Group 1	A B C	A B C
Group 2	A B C	(A B) C
Group 3	A B C	(A B C)

Note. Categories within parentheses are conjoined by the subjects; all other conjunctions are Boolean and are performed by the experimenters. Note that in Group 2, (A B) denotes any possible pair of categories.

sponses at Stage 2 is "+," whereas the Boolean conjunction of the responses at Stage 1 is "−." This is an overextension. This control provides an estimate of the number of over- and underextensions expected, even if the subject always makes Boolean judgments. Not all changes between responses at Stage 1 and Stage 2 will result in an over- or an underextension. Suppose that a subject responds (+−+) at Stage 1 and (++−) at Stage 2. The Boolean conjunction of the responses at Stage 2 is "−." This is consistent with the Boolean conjunction of the Stage 1 judgment.

At Stage 2, Group 2 combined two categories of the triple—for example, A FOOD AND FLAVORING—and judged the remaining category, A PLANT, separately. Equal numbers of lists were headed by each possible pair of categories (in this case, food–plant, food–flavoring, plant–flavoring). Some subjects judged the conjunct A FOOD AND A PLANT and the single category A FLAVORING, some judged the conjunct A FLAVORING AND A PLANT and the single category FOOD, and so on. Thus the results are not biased by choosing one particular pair of categories.

The judgment for the conjunct (e.g., +) and for the single category (e.g., −) at Stage 2 were Booleanly combined (−) to provide an estimate of the number of overextensions and underextensions resulting from the combination of two categories by the subjects. It is necessary to have subjects judge the single category so that the results of all three groups are directly comparable. This is so because some of the overextensions in the binary cases would not be picked up in the triple conjunction.

If we are to compare between groups to assess non-Boolean effects, it is crucial that the task be such that a strictly Boolean model of conjunction predicts the same number of overextensions and underextensions in each group. Hence, any difference in the number of overextensions or underextensions between groups is evidence for non-Boolean combination. For this reason we included single category judgments for the remaining category of the triple, in addition to the binary judgements. Thus, in Group 2, subjects judged, for example, the binary category A TOOL AND A WEAPON and also the single category FARM EQUIPMENT. It may seem more natural to have subjects judge the binary conjunction alone and compare the number of overextensions and underextensions directly with those in Group 3. However, if this strategy were used, a purely Boolean model would predict a greater number of overextensions and underextensions in Group 2 than in Group 3. This is so because some of the overextensions and underextensions will be counted in Group 2 but hidden in Group 3. Suppose that a subject uses a Boolean strategy. Thus all overextensions and underextensions are due to forgetting and changing one's mind and so on. At Stage 1, this subject might judge *hammer* to be A TOOL, but not A WEAPON or FARM EQUIPMENT (+−−). At Stage 2, the subject might judge *hammer* A TOOL and A WEAPON but not FARM EQUIPMENT. We now show that using this apparently attractive method of counting overextensions and underextensions leads to this perfectly Boolean subject's being classified as an overextender if in Group 2 but not in Group 3. If the subject is in Group 2, *hammer* will be judged to be A TOOL AND A WEAPON at Stage 2, though it was not judged to be A WEAPON at Stage 1. Hence, the subject would have overextended. On the other hand, the same change of mind does not lead to an overextension in Group 3. *Hammer* is rated as both A TOOL and A WEAPON at Stage 2, but because it is not FARM EQUIPMENT, *hammer* is judged not to be A TOOL, A WEAPON, AND FARM EQUIPMENT. This is illustrated in Table 7. To make the conditions comparable, it is necessary to incorporate the judgment about the third category (FARM EQUIPMENT) in all groups. This is done by Booleanly conjoining the binary judgment (A WEAPON AND A TOOL) with the subject's judgment of the third category (FARM EQUIPMENT) at Stage 2. The outcome of this procedure (in this case "−") is then compared with the subject's Stage 1 judgments to assess the number of overextensions and underextensions appropriately. On a Boolean model of conjunction, the number of overextensions and underextensions should be the same across groups.

The compensation hypothesis predicts that the more categories a subject conjoins, the greater the number of overextensions will be made. Thus, we predict there will be more overextensions in Group 3 than in Group 2, and more in Group 2 than in Group 1. In contrast, as we have noted, the Boolean hypothesis predicts that there will be the same number of overextensions and underextensions in all three groups. Furthermore, considerations analogous to those outlined above show that the number of overextensions and underextensions should be equal in each group.

At Stage 2, Group 3 judged the membership of the triple conjunction (e.g., A FOOD, A PLANT, AND A FLAVORING).

In all other ways the procedure was the same as that of Experiment 2.

Results and Discussion

The frequency of overextensions and underextensions in each category for the three experimental groups are shown in Table 8. The table shows that overextensions increase in Group 3, where subjects judge triple conjunctions. The numbers of overextensions and underextensions reported in this table are not comparable with Experiments 1–3. This is so because additional stimulus materials have been added and because many overextensions are "hidden" in triple conjunctions, as noted above. Table 9 summarizes the number of overextensions for each category when subjects judged membership of triple conjuncts. As can be seen from this table, some categories were overextended far more often than others. For example, PET was overextended 22 times, while PREDATOR was overextended twice, and BIRD was not overextended at all. In some cases, overextensions were more equally distributed across categories. For example, VEHICLE was overextended 8 times, while MACHINE and TOY were overextended 7 times. The stability of judgments for BIRD is understandable because BIRD is a natural kind, and judgments of whether or not something is a BIRD are often intuitively all or none. However, this kind of explanation does not seem to apply to intuitively graded categories such as LUXURY, which was not overextended. A possible explanation for the lack of overextensions in this case is that LUXURY is so flexible that it is very rarely given a "−" rating at Stage 1, and thus that the possibility of overextension does not arise. However, reexamination of the data showed that 60% of items were judged to be LUXURY at Stage 1, a figure comparable with percentage

Table 7
Overextensions and Underextensions as a Function of Category: Experiment 4

Category			Underextensions[a]	Overextensions[b]						
A	B	C	+++−	−+++	+−++	++−+	+−−+	−+−+	−−++	−−−+
				Group 1[c]						
Furniture	House. appl.	Luxury	12	1	4					
Food	Plant	Flavoring	10	3	5	2				
Weapon	Tool	Farm equip.	4		1	2	1	1		
Building	Dwelling	Business	4		5	1				
Machine	Vehicle	Toy	2	1		2		1		
Bird	Pet	Predator			3	1				
				Group 2[d]						
Furniture	House. appl.	Luxury	3	1						
Food	Plant	Flavoring	2	10	2	6			3	
Weapon	Tool	Farm equip.	4	6	1	3				
Building	Dwelling	Business	1		2	1	1			
Machine	Vehicle	Toy	8	4	3	3		1	2	
Bird	Pet	Predator	1		3	1				
				Group 3[e]						
Furniture	House. appl.	Luxury	5	11					1	
Food	Plant	Flavoring	13	9	5	6	3	1	7	
Weapon	Tool	Farm equip.	3	4	6	7	3	1	2	2
Building	Dwelling	Business	4	1	3	2	3		1	
Machine	Vehicle	Toy	2	3	4	4	1	1	2	1
Bird	Pet	Predator			21	1	1			

Note. Overextensions and underextensions as percentages of the total number of responses are shown in parentheses in footnotes. The quadruple of pluses and minuses denotes rating for Category A, rating for Category B, rating for Category C, and rating for conjunction, irrespective of the order in which the categories were presented. House. appl. = household appliance; equip. = equipment.
[a] Total underextensions in Group 1 = 32 (2.8%), in Group 2 = 19 (1.6%), and in Group 3 = 27 (2.3%).
[b] Total overextensions in Group 1 = 34 (3.0%), in Group 2 = 53 (4.6%), and in Group 3 = 117 (10.2%).
[c] Single category judgments at Stage 2.
[d] Binary category judgments combined with single category judgments at Stage 2.
[e] Membership judgments for triple conjunctions at Stage 2.

judgments for the other categories. Thus there seems to be no single obvious reason why some categories overextend more than others. In Experiments 1 and 2 we noted the possible influence of the position of categories in the conjunction. As Hampton found, categories in the first position appear to be more overextended than categories in the second position. It is interesting to wonder how this effect might transfer to the case of triple conjunctions. Table 10 shows the number of overextensions for each category group as a function of position in the triple conjunction. Although there is considerable variation within categories, there appears to be no overall effect of position (48 overextensions in first position, 48 overextensions in second position, 47 overextensions in third position).

Table 8
How an Overextension in Two Categories May Not Count as an Overextension in Three Categories

Group and stage	A weapon	A tool	Farm equipment
Group 2			
Stage 1	+	−	−
Stage 2[a]	(+	+)	−
Group 3			
Stage 1	+	−	−
Stage 2[b]	(+	+	−)

[a] Stage 2 (++) is Booleanly conjoined to give (+) for "a weapon and a tool." This is an overextension.
[b] Stage 2 (++−) is Booleanly conjoined to give (−) for "a weapon, a tool, and farm equipment." This is not an overextension.

One possible explanation for the lack of position effect in triple judgments is the fact that a slightly different syntax was used in this experiment from that used earlier. In Experiment 2 the syntax "A machine which is also a vehicle" was used, after Hampton (1988). In Experiment 4 the syntax "A machine, a vehicle, and a toy" was adopted. This raises the possibility that the apparent loss of position effects is due to a change in syntax rather than the number of items in the conjunction. In Experiment 4, Group 2, subjects made membership judgments for binary categories with the revised syntax ("A machine and a vehicle"). Overextensions for binary judgments for each category group (e.g., machine, vehicle, toy) are shown in Table 11. Additional judgments for single categories at Stage 2 are ignored. As in Experiment 2 (using Hampton's original syntax), there are a greater number of overextensions in the first position (72 vs. 56). However, as before, this did not reach significance in a Wilcoxon signed-rank test ($W = 3.5$, $N = 6$, $p > .05$). Whether or not the head-qualifier effect is present in binary conjunction, but not in triple conjunctions, as suggested by our data, and to what extent, if any, the head-qualifier effect is syntax dependent remain open questions.

The data were fitted to a 2 × 3 Response Type (overextensions and underextensions) × Experimental Group (1, 2, and 3) logistic model. There was a main effect of response type

Table 9
Number of Overextensions by Category for Judgments of Triple Conjunctions (Group 3, Experiment 4)

Category	Overextensions
Bird	0
Pet	22
Predator	2
Building	2
Dwelling	7
Business	5
Machine	7
Vehicle	8
Toy	7
Weapon	9
Tool	13
Farm equipment	13
Furniture	12
Appliance	1
Luxury	0
Food	17
Plant	15
Flavoring	10

($\chi^2 = 56.4$, $df = 1$, $p < .05$). Inspection of the mean number of overextensions and underextensions in each experimental group shows that this is due to the tendency of subjects to overextend rather than underextend their categories (see Table 8). There is also a significant main effect of experimental group ($\chi^2 = 37.6$, $df = 2$, $p < .05$). A post hoc Jonckheere trend test showed that there was a significant trend in overextensions from Group 1 (34 overextensions, 33 underextensions) to Group 3 (117 overextensions, 27 underextensions) ($S = 160$ [12], $p < .05$).

There is a significant interaction between response type (overextensions vs. underextensions) and group (single judgments, binary judgments, triple judgments) ($\chi^2 = 20.2$, $df = 2$, $p < .05$). An examination of residual errors of the fit of the logistic model shows that as expected there was no difference between the number of overextensions and underextensions in Group 1 (residual = 0.086, $df = 1$, $p > .05$). In Groups 2 and 3, however, there were significantly more overextensions than underextensions (residual = 2.769, $df = 1$, $p < .05$ in Group 2; residual = 5.303, $df = 1$, $p < .05$ in Group 3). As argued above, if subjects used a Boolean combination strategy, the number of overextensions and underextensions should be the same for all groups. These results show that the more constituents the subjects have to conjoin, the greater their propensity to overextend. This finding is exactly what the compensation hypothesis predicts.

Table 10
Number of Overextensions for a Category Group as a Function of Position in the Triple Conjunction (Group 3, Experiment 4)

	Position		
Category	1st	2nd	3rd
Furniture, household appliance, luxury	4	6	3
Food, plant, flavoring	16	6	17
Weapon, tool, farm equipment	5	15	11
Building, dwelling, business	6	5	3
Machine, vehicle, toy	9	6	7
Bird, pet, predator	8	10	6
Total	48	48	47

General Discussion

Possibility of a Suppression Effect

Hampton (personal communication, July, 1989) suggests that even on a Boolean model, more overextensions than underextensions might result from a suppression effect at Stage 1. This is based on the finding (Hampton, 1988) that in some cases if an item has already been rated a member of one category, it is less likely to be rated a member of a second category. This means that an item which would normally be rated as, for example, both a VEHICLE and a MACHINE might not be rated as a MACHINE if it has already been rated as a VEHICLE at Stage 1. At Stage 2, however, because both categories appear together, this suppression effect might not be present, and so the item would now be rated as a member of the conjunctive category. This would result in an overextension.

If the suppression effect was operating at Stage 1, then there should be more positive judgments when a category was rated first than when it was rated second. In Appendix B, the percentages of *yes* responses at Stage 1 when rated first or second are shown for Experiments 1–4. There is a significant effect of rating order in Experiment 1 ($W = 0$, $N = 6$, $p < .05$), although inspection of the means reveals that the absolute value of this difference is quite small (58.6% when rated first, 54.2% when rated second). However, this is not observed in Experiments 2–4 (Experiment 2: $W = 6$, $N = 6$, $p > .05$; Experiment 3: $W = 7$, $N = 6$, $p > .05$; Experiment 4: $\chi r^2 = 2.25$, $df = 1$, $N = 6$, $p > .05$). Hence, the suppression effect cannot be responsible for the non-Boolean effects observed in Experiments 2–4.

Summary of Results

This series of experiments has been concerned to outline the consequences of a Boolean model of concept combination and to assess their empirical validity. The first three experiments demonstrated that overextensions significantly exceeded underextensions in direct contradiction to the prediction of the Boolean hypothesis. We concluded that overextension is not an artifact of the particular task used, nor a result of a suppression effect, but is a genuine property of categorization decisions. We discussed a possible explanation of this tendency to overextend and proposed that this was the result of a best fit strategy. Such a strategy appears to be employed

Table 11
Number of Overextensions for Binary Judgments as a Function of Position in the Binary Conjunction (Group 2, Experiment 4)

Category	Position 1st	Position 2nd
Furniture, household appliance, luxury	13	7
Food, plant, flavoring	19	17
Weapon, tool, farm equipment	9	12
Building, dwelling, business	12	6
Machine, vehicle, toy	11	9
Bird, pet, predator	8	5
Total	72	56

in many everyday situations where we must fulfill many criteria together. This compensation hypothesis predicts that the greater the number of criteria to be met, the greater the tendency to overextend. We tested this hypothesis in Experiment 4 by comparing subjects' ratings of triple conjunctions, binary conjunctions, and single categories. We found that subjects overextended most often when judging triple conjunctions, and least often when judging single categories. These findings are just what is predicted by the compensation hypothesis: that categorization of conjunctions is a result of a best fit strategy.

Implications of a Compensation Model

We have postulated that conjunctive categorization judgments are a result of a best fit strategy. When subjects have a number of criteria to meet simultaneously, they tend to interpret these criteria more leniently than when they are judged independently. Thus when a simple category is in the context of being a constituent of a complex category, it is overextended by the subject. The compensation hypothesis is not incompatible with Hampton's composite prototype explanation of overextension (Hampton, 1988). However, we see concept combination as a particular sort of context effect, in which each constituent of the compound acts as a context for the other. For example, in understanding the complex concept PET FISH, the word *pet* influences the kind of fish we expect, and the word *fish* influences the kind of pet we expect because of what we know about fish and pets. In this light, the phenomenon of overextension can be construed as another example of the context dependency of category judgments.

We have explained contextual influence in context combination as a matter of making best fit judgments to a number of criteria. This view may be extended to the general problem of contextual influence on categorization. Categorization judgments may be seen as "best fit" matches to the constraints context has imposed. In such a view, just as the word *pet* may provide a context that alters what is judged to be a fish, the real life context of a pet shop alters what is judged to be a fish.

If concept combination is a product of best fit matching, we shall not be able to understand when and why particular concepts may be overextended without adverting to world knowledge. Overextension and conceptual flexibility are constrained by what we know about the situation. In our example of choosing someone for a theatrical role, it may be allowable to construe 5′8″ as tall, and an ordinary looking man as handsome. On the other hand, in the context of a play about Edinburgh, it may not be possible to overextend the criterion that the actor has a Scottish accent. That is so because a Scottish audience would find such a compromise unacceptable. However, this requirement, too, is subject to context. If the production is touring Japan, then an Irish accent might do just as well. The compromises made in best fit matching appear to be a product of our understanding of the situation we are in. Thus in order to have a full account of conceptual combination, it may be necessary to take account of subjects understanding of the world. A compensation account of concept combination fits well with the view that concepts must be understood in the context of the theories of the world (e.g., Lyon & Chater, 1990; Medin & Shoben, 1988; Medin & Smith, 1984; Murphy, 1988; Murphy & Medin, 1985).

References

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

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development* (pp. 101–140). Cambridge: Cambridge University Press.

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

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition, 15,* 57–71.

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

Lakoff, G. P. (1975). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic, 2,* 458–508.

Lyon, K. (1989). *Context effects and the structure of concepts.* Doctoral thesis, University of Edinburgh.

Lyon, K., & Chater, N. (1990). Localist and globalist approaches to concepts. In K. Gilhooly, M. Keane, R. Logie, & G. Erdos (Eds.), *Lines of thinking: reflections on the psychology of thought* (pp. 41–57). Chichester, England: Wiley.

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

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology, 20,* 158–191.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology, 35,* 113–138.

Murphy, G. L. (1988). Comprehending complex concepts. *Cognitive Science, 12,* 529–562.

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

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

Roth, E. M., & Mervis, C. B. (1983). Fuzzy set theory and class inclusion relations in semantic categories. *Journal of Verbal Learning and Verbal Behavior, 22,* 509–525.

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

(Appendix follows on the next page)

Appendix A

Proof That the Probability of Overextensions and Underextensions is Equal

In Hampton's original formulas, there are separate parameters (s_1 and s_2) that denote whether the category was rated first or second.

Underextensions

$$p(++-)$$
$$=s.g.[u^2.(0) + u.(1-u).(1-g) + (1-u).u.(1-s)$$
$$+ (1-u)^2.(1-s.g)]$$
$$=s.g.[(u-u^2).(2-s-g) + (1-2.u+u^2).(1-s.g)]$$
$$=s.g.[(u^2.(s+g-1-s.g)$$
$$+u.(2.s.g-s-g)$$
$$+(1-s.g)].$$

Overextensions

$$p(+-+) + p(-++) + p(--+)$$
$$=s.(1-g).[u^2.(0) + u.(1-u).g + (1-u).u.(0) + (1-\mathbf{u})^2.s.g]$$
$$+(1-s).g.[u^2.(0) + u.(1-u).(0) + (1-u).u.s + (1-u)^2.s.g]$$
$$+(1-s).(1-g).[(1-u)^2.s.g]$$
$$=(u-u^2).[s.g.(1-g) + s.g.(1-s)]$$
$$+(1-2.u+u^2).[s^2.g.(1-g) + s.g^2.(1-s) + s.g.(1-s).(1-g)]$$
$$=s.g.[(u^2.(s+g-1-s.g)$$
$$+u.(2.s.g-s-g)$$
$$+(1-s.g)].$$

Appendix B

Percentage of "Yes" Responses at Stage 1 as a Function of the Order in Which Categories Were Rated

Category	Rated 1st	Rated 2nd	Rated 3rd
	Experiment 1		
Furniture, household appliance	53.9	47.7	
Food, plant	71.9	65.6	
Weapon, tool	53.9	50.0	
Building, dwelling	55.5	54.7	
Machine, vehicle	61.7	54.7	
Bird, pet	54.7	52.3	
Average	58.6	54.2	
	Experiment 2		
Furniture, household appliance	45.1	37.5	
Food, plant	67.8	70.1	
Weapon, tool	58.7	43.4	
Building, dwelling	54.9	54.2	
Machine, vehicle	48.3	45.1	
Bird, pet	47.2	53.5	
Average	52.8	50.6	
	Experiment 3		
Furniture, household appliance	45.3	29.7	
Food, plant	60.2	60.9	
Weapon, tool	45.3	50.8	
Building, dwelling	50	46.9	
Machine, vehicle	51.6	53.1	
Bird, pet	46.9	57.1	
Average	49.9	49.7	
	Experiment 4		
Furniture, household appliance, luxury	46.2	51.7	46.4
Food, plant, flavoring	49.8	53.5	47.7
Weapon, tool, farm equipment	47.7	49.8	46.7
Building, dwelling, business	45.8	45.5	45.0
Machine, vehicle, toy	44.6	41.1	41.1
Bird, pet, predator	43.8	42.7	43.8
Average	46.3	47.4	45.1

Note. Data are shown for each pair (or triple) of categories, so that the total number of items rated in each position is balanced.

Received December 16, 1988
Revision received August 8, 1989
Accepted September 28, 1989 ■