

Conjunctions of Visually Based Categories: Overextension and Compensation

James A. Hampton
City University London

The degree to which overextension effects found with conjunctions of semantic categories, such as sports and games (J. A. Hampton, 1988), would generalize to categories of visual stimuli was tested in 4 experiments. Overextension occurs when participants categorize a stimulus in the conjunction of 2 categories but fail to categorize the same stimulus as belonging to 1 of the 2 constituent categories considered individually. Stimuli for the present experiments were ambiguous colored letter shapes and cartoon faces that could vary along dimensions of happiness and either apparent intelligence or apparent age. Overextension was found with both stimulus sets, thus showing that the phenomenon is not restricted to categorization in superordinate semantic categories. There was also evidence that typicality in 1 category could compensate for borderline membership of the other. More overextension was found for faces than for letters, and there was evidence for asymmetric compensation between category dimensions.

The classical definition of the conjunction of two concepts is the overlap or intersection of their set membership. Thus, the conjunctive concept "triangle which is red" is composed of shapes that are both triangular and colored red. Recent evidence (Hampton, 1991), however, has suggested that when the two concepts are less well defined and are of the type identified by Rosch (1975) as having a prototype structure, then the psychological representation of their conjunction may not be defined so simply. In one series of experiments, Hampton (1988) studied a range of conjunctive concepts such as "sports which are also games" or "pets which are also birds" and found that membership of the conjunctive set was often overextended. That is, participants placed items in the conjunction, which they had previously judged as not belonging in one of the sets considered alone. Chess, for example, was judged to be a typical game but not to be a sport. However, a majority of participants having made these judgments then placed it in the category of "sports which are games." Similar levels of overextension occurred regardless of whether the conjunction was expressed as "sports which are games" or as "games which are sports," although these two forms were by no means identical in their pattern of membership.

Hampton (1987, 1988) argued that these data reflected an intensionally based process by which participants attempt to form a novel composite prototype concept to represent the conjunction. Using data from an attribute generation and rating study, Hampton (1987) showed that those attributes considered to be generally true of the conjunction were only partly formed from the union of the attributes of the two

constituent concepts. Although overall the importance of an attribute for the conjunction could be predicted as a weighted average of its importance for each constituent, there were also two kinds of noncompositional effect. Attributes that were important for one constituent could sometimes be omitted from the conjunctive concept (e.g., birds *migrate*, whereas "birds which are pets" do not). Other attributes were found that were true of the conjunction but were not true of either constituent (e.g., *talks* and *lives in cages* were true of "birds which are pets" but not of either *pet* or *bird* considered alone). Hampton's composite prototype model proposes that participants integrate the two sets of constituent attributes into a single coherent composite prototype for the conjunction by initially forming the set union of the two sets of attributes and weighting those attributes with the average of their weights for the two constituents. This initial stage is also subject to two constraints. The *necessity constraint* requires that any attribute that is necessarily true of a constituent is also true of the conjunction, and the *impossibility constraint* requires that any attribute of one constituent that is impossible for the other will also be impossible for the conjunction. Hampton (1987) found evidence for both these constraints in attribute inheritance by conjunctions. Membership of the conjunctive category is then computed by using the similarity of any item to the new prototype—rather than by reference to category membership in each constituent separately. Noncompositional effects are then considered to be the result of further processes acting on the union of the two attribute sets to render them *coherent* and to take account of additional exemplar knowledge that a person may have (see Hampton, 1987, for details). The application of coherence requirements will tend to lead to *inheritance failure*. For example, pets are warm and cuddly, but because this is incompatible with being a fish, these attributes are not inherited in the conjunction "*fish which are pets*." Additional exemplar knowledge will tend to lead to *emergent attributes*. For example, birds that are pets tend to live in cages and may even talk, although neither of these attributes are

I thank Lawrence Barsalou, Nick Braisby, Paul De Boeck, Bradley Franks, John Gardiner, Dedre Gentner, Janellen Huttenlocher, Iven van Mechelen, Vincenz Morger, Gert Storms, and Jean-Pierre Thibaut for helpful discussion of this work. I also thank Fay Millichap and Caroline Venables for help with data collection.

Correspondence concerning this article should be addressed to James A. Hampton, Department of Psychology, City University London, Northampton Square, London EC1V 0HB England. Electronic mail may be sent via Internet to j.a.hampton@city.ac.uk.

ever mentioned, or judged to be true, of the individual constituent categories of birds or pets considered alone.

The effect of judging conjunctive category membership with respect to a composite prototype is to introduce the possibility of "nonlogical" responses. Assuming that similarity to the composite prototype is some combinatorial function of similarity on each of the attribute dimensions in the prototype (Hampton, 1979, 1993; Tversky, 1977), similarity to the composite prototype will increase monotonically with the degree of match on each attribute. Hence, an increase in the match to the set of attributes derived from one constituent concept will compensate for a possible decrease in the match to those from the other concept in determining overall similarity to the composite prototype. Because membership in the constituent concept categories is also presumed to depend on degree of match to a similar set of attributes, it follows that the probability of membership in the conjunction will be some monotonically increasing function of the probability of membership in each constituent. An increase in the degree of attribute match to either constituent will generally make an item a better member of the conjunction.¹ This process for judging conjunctive membership will yield nonlogical categorizations for the following reason. To be consistent with Boolean set intersection, once an item matches one constituent well enough to be a clear member, the degree of match to that constituent should no longer influence membership in the conjunction. Judging similarity to a composite prototype concept, however, will entail that any increase in similarity will render an item more likely to belong in the conjunction, unless it is already at ceiling.

The composite prototype model for conjunction applied in its most simple form (i.e., without any mechanisms for attribute inheritance failure or for emergent noninherited attributes) is an example of a general class of model described by Ashby and Gott (1988) as integration models. Ashby and Gott considered a range of decision models for discriminating classes of stimuli lying in a multidimensional feature space. They distinguished independence models in which each dimension (or feature) is decided on independently, and the resulting decisions then combined, from integration models in which the dimensions are first integrated, and then a decision function is applied to the result. It is only independence models that partition a space in rectangular Boolean regions. It may not, however, be a necessary aspect of independence models that the category boundaries are rectangular. A recent stochastic model of conceptual combination (Huttenlocher & Hedges, 1994) suggests that conjunctions of fuzzy categories could be modeled by considering the bivariate distribution of membership values, which together with an independence assumption generates circular borderlines for conjunctive concepts. Other examples of integration models for conceptual combination are to be found in Smith, Osherson, Rips, and Keane (1988) and in fuzzy logic approaches (e.g., Jones, 1982; Oden, 1977). The issue of independence versus integration of information is also critical in the current debate on categorical speech perception (see Massaro, 1989, and commentaries).

The integrative nature of the composite prototype model predicts non-Boolean responses but is not in itself a sufficient account of overextension. Integration models predict inconsis-

tencies between constituent and conjunctive membership decisions but do not necessarily imply any tendency to overextend rather than to underextend the conjunction. Overextension could depend on a number of possible factors, some of which are tested by the experiments reported in this article.

The purpose of the present article is to explore the generality of the phenomenon of overextension of conjunctions to understand better its basis. Previous research on conjunctions of natural concepts (Chater, Lyon, & Myers, 1990; Hampton, 1987, 1988) has relied exclusively on verbal materials. Stimulus sets in these experiments were composed of lists of nouns describing types of object or activity that were then categorized in superordinate categories. A possible problem for this approach is that the objects being categorized are not actual individual entities but are in fact themselves categories. For example, in deciding if chess is a sport, the word *chess* could be understood as referring to a class of different occasions and situations in which the game chess could be played. Some of these situations would perhaps fit the sport category; whereas others would not. This variety of possible instantiations of the noun concepts to be categorized suggests an account of overextension; specifically, that the instantiation brought to mind for the element may be affected by the category with which it is being compared. Hence, the stimulus being categorized may not remain constant across the three tasks (judging each constituent and then their conjunction). (Another problem with semantic category materials is that the design is inherently correlational because no experimental control can be exercised over the exact nature of the items to be categorized.)

To test this account of overextension based on the possible vagueness of using words as the elements to be categorized, two novel stimulus sets using visual stimuli were created for the present experiments. The first consisted of letter shapes that could vary between an *H* and an *A* by small graded steps in the angle of the two side bars of the shape from parallel vertical bars (for a prototypical *H*) to converging bars just touching at the apex (for a prototypical *A*). Orthogonal to this variation in shape, the stimuli were displayed in a range of color that varied between blue and green (Experiment 1) or between red and orange (Experiments 2-4). Participants could therefore make judgments about the letter category and the color category that a stimulus belonged to and could also make judgments about a conjunction, such as blue *A*. The second stimulus set consisted of simple schematic faces drawn with an ellipse for a head and then simple line drawings of eyes, nose, and mouth. For the first experiment the faces were categorized along two dimensions: the facial expression being happy or sad, and the overall appearance of the face as being intelligent or stupid. For the later experiments, intelligence was replaced by age, with categories of adult and child.

The two stimulus sets obviously differed from each other in many ways. The purpose of including both sets in the experiments was first to provide generality and second to start to explore the possibility that complexity of stimulus structure might be a factor in determining the degree of overextension

¹ This will not, however, be true in more complex conjunctions, if the matched attribute has not been inherited by the conjunction.

observed. In her description of semantic prototypes, Rosch (1973, 1975) identified two rather different kinds of prototype concept. In her earlier research, Rosch identified unidimensional prototypes like the concept red or square as *cognitive reference points* on simple dimensions of color or shape.² In her later work with Carolyn Mervis, she identified a second kind of prototype concept that is based on the idea of *family resemblances* (Rosch & Mervis, 1975). These concepts, like fruit or furniture, are composed of sets of correlated attributes. The two different stimulus sets in the present experiments represented one each of these two different conceptions of prototype-based categories. The colored letter stimuli were based on two unidimensional categories of shape and color with a correspondingly simple feature structure akin to the cognitive reference point definition of a prototype. By contrast, the cartoon face stimuli were composed of a number of independently varying features corresponding to different parts of the face and their configural relations and so were more akin to the family resemblance form of prototype concept. I tentatively hypothesized that this more complex feature structure might encourage the formation of a composite prototype to a greater extent than the simple structure of the colored letters and hence lead to more overextension.

A second account of overextension that is addressed by the use of visual stimuli is an account based on nonmonotonic effects in conceptual combination. Hampton (1987) analyzed the attributes considered to be true of constituents (like bird or pet) and conjunctions (such as pet which is a bird) and found two types of nonmonotonicity, which were described above as inheritance failure and emergent attributes. These effects were seen as being characteristic of semantic domains that have a rich structure of known exemplars and of background theories that place constraints on which combinations of attributes can easily be placed together (see also Hampton, 1990; Kunda, Miller, & Clare, 1990; Murphy, 1988, for further examples). Inheritance failure and emergent attributes are also a key part of the account of how it is that exemplars may be better examples of a conjunction than of either constituent (Osherson & Smith, 1981) and hence of why conjunctions may be overextended (Hampton, 1988). If the semantic intension for a conjunction is not simply composed of the union of the intensions of each constituent, but contains emergent attributes of its own, then similarity of an exemplar to the conjunction may be greater than its average similarity to each constituent. Hence, overextension may occur without a marked loosening of the criterion used for similarity-based categorization to the conjunctive prototype.

By contrast with semantic categories, the relatively simple stimuli used here are not *prima facie* liable to suffer from either inheritance failure or from emergent attributes. Thus, if overextension is the result solely of nonmonotonicity effects, no overextension with either of the present stimulus sets would be expected to be found.

Experiment 1

In common with earlier research (Hampton, 1988), in Experiment 1 I obtained three categorization responses for the items in each set of stimuli. The first two categorizations were

with respect to two constituent categories—defined as opposite poles on a particular dimension: the shape and color of the letter shapes—and the happiness and the intelligence of the faces. The two dimensions were arranged to be orthogonal. The third categorization was then for a conjunction of two poles of the constituent dimensions, for example, a green *H* or a happy–stupid face. The data of interest were the extent to which participants would overextend the conjunctive categorizations. If little or no overextension occurs, then this would suggest that overextension as found previously may be restricted to the categorization of noun classes (as in judging that a hammer is a weapon or that chess is a sport) or to categorization involving semantic categories that show non-monotonic conjunction effects. Given individuals to classify in simply defined classes in which intensions do not interact, participants may adopt a more intersective approach to judging conjunctions. Conversely, if overextension is found with these materials, then the implication would be that participants overextend categorization of both classes and individuals, for both rich and simple semantic domains, and the results would lend support to the composite prototype model as a general model of conjunction formation.

Method

Participants. Participants were 48 students at City University London who acted as unpaid volunteers.

Apparatus. The apparatus for all the experiments was the same. Stimuli were presented under program control on the VGA display screen of an IBM compatible computer. The program was written in Microsoft QuickBasic. Participants made classification responses and typicality judgments by typing on the keyboard. There were two sets of stimuli: colored letter shapes and cartoon faces.

Materials: Letter shapes. Twenty-nine letter shapes were created by orthogonally varying two dimensions. The letters had a common central horizontal bar 100 pixels long and 10 pixels thick. At either side were two bars 200 pixels high and 10 pixels thick that were vertical for the *H* stimulus. The vertical bars could be rotated in opposite directions symmetrically so that after 30° of rotation they touched in the center above the horizontal bar to make an *A*. Eleven angles of rotation were used to create 11 steps along the shape dimension. The angles were 0°, 5°, 8°, 10°, 13°, 15°, 17°, 20°, 22°, 25°, and 30°. Zero degrees was a good *H*, and 30° was a good *A*. Color of the shapes was also varied in 11 steps along a dimension from blue to green. The VGA graphics system creates colors on the basis of three parameters for the red, green, and blue (RGB) components of the color, which each can vary from 0 to 63. The best blue used had RGB values of 0, 20, and 50 respectively. The best green had RGB values of 0, 50, and 20, respectively. Intervening steps had different proportions of blue and green with a constant sum of 68 for the two parameters, plus 5 or 10 units of red in the center of the scale to maintain the apparent luminance (as judged intuitively). Apart from the two extremes, neighboring steps on the scale differed by two or three increments on the blue and green values. For both angle and color, pilot studies were used to identify the category boundary, and the stimuli were placed symmetrically around the boundary. The 29 stimuli were symmetrically distributed around the stimulus space according to the scheme shown

² Although color and shape are not strictly unidimensional, they are here described as such to differentiate them from family resemblance prototypes, in which the multidimensionality of the prototype representation is more explicit.

in Figure 1. Examples of different angles between *A* and *H* are shown in Figure 2.

Materials: Faces. The 29 faces were constructed as cartoonlike drawings with a circle of radius 45 pixels for a head, a nose that was constant across drawings, and different features for eyes, eyebrows, and mouth. The apparent intelligence was manipulated by varying the height of the center of the circle defining the head shape relative to the features within it. Thus, intelligent faces had high brows and weak chins; whereas stupid faces had low brows and broad chins. There were 11 steps on the scale, each separated by 2-pixels distance in the height of the head outline. Eleven steps of variation in emotional expression between happy and sad were achieved as a result of pilot studies testing different combinations of features for eyes, brows, and mouth. The final scale involved combinations in which two different eyes, four different eyebrow shapes, and five different mouth shapes were used. As for the letter shape stimuli, the boundary between the two categories on each dimension was estimated in a pilot study, and the stimuli were designed to fall equally either side of the boundary. The distribution of stimuli in the stimulus space was the same as that used for the colored shapes (see Figure 1). The four extreme stimuli prototypical of each conjunction are shown in Figure 3.

Procedure. The participant responded to six blocks of trials. Stimulus sets were alternated between blocks, and the tasks followed one of two fixed orders, balancing the order of judging constituent

concepts. Instruction screens appeared at the start of each block of trials. For the first fixed order, in Blocks 1, 3, and 5, the same set of 29 colored letter shapes was presented in a newly randomized order for each block. In Blocks 2, 4, and 6, the set of 29 cartoon faces was presented similarly randomized. The instructions varied from block to block. In Block 1, the participant was asked to categorize stimuli on the basis of a blue-green distinction. The instruction screen displayed a scale with square color patches at either end corresponding to the prototypical blue and green shades that would be presented, with a labeled scale beneath ranging from +10 (*blue*) through 0 (*borderline*) to +10 (*green*). The instructions stated the following:

The task is to decide whether different colored shapes look more like blue or more like green. For each color, decide if the color is blue or green, and type a *B* or *G* to indicate your choice. You will then be asked to rate the strength of your choice. Use a number 10 to indicate a very typical blue (or green), and a 1 to indicate a color that is only just blue (or green). Use numbers in between to indicate the range of typicality of the colors. The squares above show clear examples of blue and green colors.

The set of colored letter stimuli was then presented in the center of the screen one at a time and "Blue or Green? (B/G)" was displayed. Participants had to press the appropriate key to indicate how they would categorize the stimulus. Their response stayed on the screen

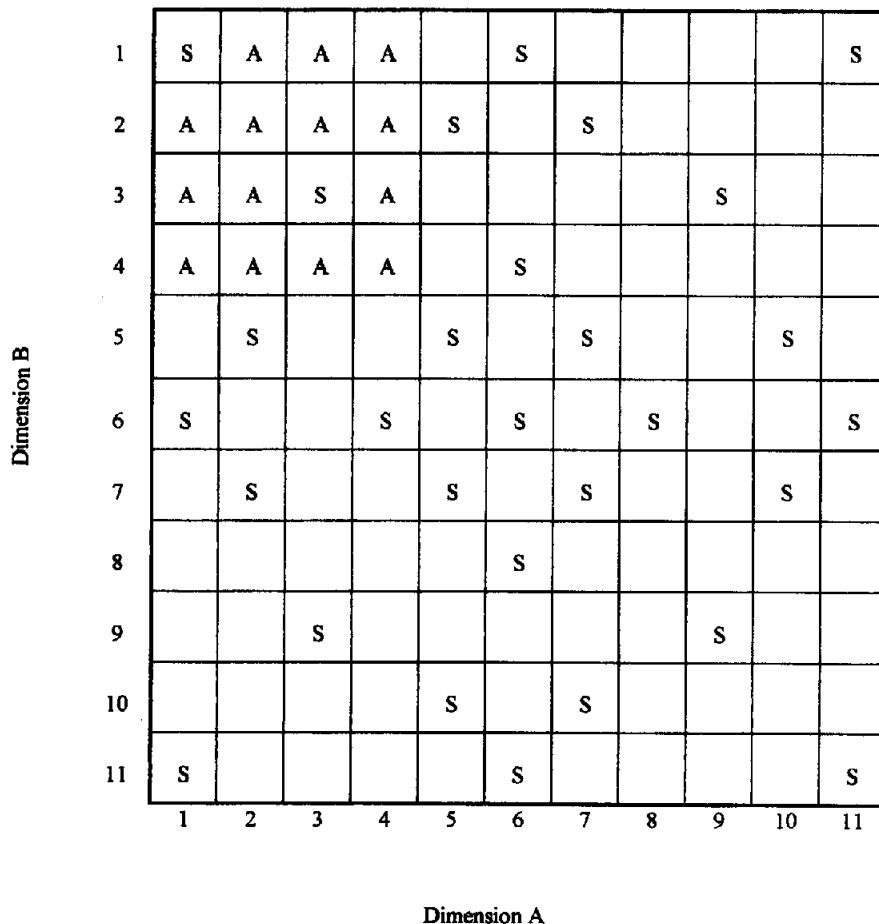


Figure 1. The distribution of standard target stimuli across the two dimensions for both stimulus sets in Experiments 1 and 2 is shown by the letter *S*. The additional stimuli used in the augmented condition of Experiment 2 for the conjunction represented by 1,1 on the two dimensions are shown by the letter *A*.



Figure 2. Examples of the letter shapes used in all experiments.

while they were asked "How typical a blue (or green) do you feel it is? Enter a number from 0 = neither blue nor green, to 10 for a maximum value." (The choice of the word blue or green depended on the participant's first response.) Once the 29 stimuli had been seen, the instructions for Block 2 followed. In Block 2 the procedure was identical except that no prototypical examples of face stimuli were shown with the instructions,³ in which participants were asked to decide whether each face looked happy or sad, by pressing the *H* or *S* key, followed by a typicality rating. Block 3 returned to the letter shape stimuli. The instructions displayed a typical *A* and a typical *H* (in white on black) and participants were instructed that they would see the same set of shapes as before, but they now had to categorize them as *As* (by pressing the *A* key) or *Hs* (by pressing the *H* key). Typicality ratings were again collected. Block 4 was identical to Block 2, except that the dimension to be judged was now how intelligent (*I* key) or stupid (*S* key) the faces looked.

In Blocks 5 and 6, the question asked referred to a conjunction of two constituent categories such as, for example, "Is this a Blue *H*?" (Block 5) or "Is this face both happy and intelligent?" (Block 6), and participants had to respond yes or no by pressing the *Y* or *N* keys, respectively. A typicality response was then obtained for yes responses as before, asking, for example, "How typical an example of a happy and intelligent face is it?" and for no responses, participants were asked "How clearly do you feel that it is not a happy and intelligent face?" In both ratings, a scale with responses ranging from 0

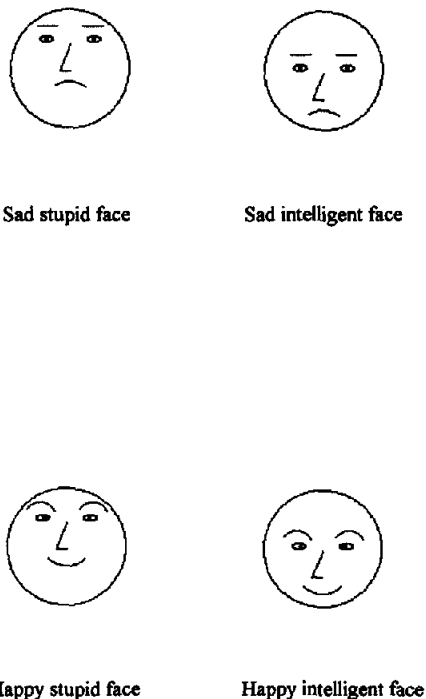


Figure 3. Examples of the prototypical faces for the four conjunctions used in Experiment 1.

(borderline) to 10 (for a very typical or very clear example) was used. In Block 5 (the letter shape stimulus set), a typical example of the conjunction was displayed with the instructions, but no example was given in Block 6 for the faces.

An alternative fixed order of blocks was used for half of the participants to counterbalance the order of categorizing the constituents within each stimulus set and involved switching the orders of Blocks 1 and 3 and Blocks 2 and 4. The order of categorizing the stimulus sets was, however, kept constant, with letter shapes first, followed by faces, in each pair of blocks, 1 and 2, 3 and 4, and 5 and 6. The complete procedure took about 30 min.⁴

Design. The comparison of categorization responses to constituent and conjunction categories was within subjects and within stimuli. Participants were divided into four different groups, each allocated to one of the four possible conjunctions in each stimulus set. Group 1 judged blue *A* and happy and intelligent, Group 2 blue *H* and happy and stupid, Group 3 judged green *A* and sad and intelligent, and Group 4 green *H* and sad and stupid. Each of these groups was further subdivided into two subgroups to counterbalance the order of presentation of the constituent concepts in Blocks 1–4. One group of 24 participants received the tasks in the following order: color, emotion, letter, intelligence; the other group received the tasks in the following order: letter, intelligence, color, emotion.

Results

Consistency and overextension. The first analysis considered the question of whether participants were consistent in their categorization of conjunctive concepts. Each participant made three responses to each stimulus: one for each constituent and one for the conjunction. This triple of responses was coded as one of eight possible types: from --- for three negative categorizations to +++ for three positive categorizations. (The conjunction group that a participant was assigned to was used to label the constituent categorizations as positive, indicating membership of a conjunctive constituent, or negative, indicating membership of the alternative category.) If each triple is defined as Constituent A, Constituent B, and

³ The reason for not providing participants with examples of the faces here, and in Blocks 4 and 6, was that the two dimensions could not be as easily separated as for the letter shapes, in which patches of color and monochrome shapes could be shown. In retrospect, this lack of a prototype could have contributed to less consistency in the responding, and steps were taken in subsequent experiments to remedy the problem by providing participants with a view of the complete set of faces as well as showing opposing prototypes for each dimension, with the other dimension set to the midpoint of the scale.

⁴ The design of this experiment, and of subsequent experiments, used a constant order of judging constituents first and then conjunctions. I felt that this sequence from simple to complex was a more intuitive order in which to make the judgments and would be less likely to incur interference between the decisions than the reverse order. Hampton (1988), whose design formed the starting point for the present research, used the same fixed order for the judgments. An earlier study by Hampton (1986), in which a between-subjects design for measuring the categorization of items in constituents and in conjunctions of semantic categories was used, found no less a degree of overextension as measured by probability of constituent and conjunctive categorization for each item than was observed with the within-subject design used by Hampton (1988).

Table 1
Frequencies of Eight Possible Response Triples and Percentages of Inconsistency for the Two Stimulus Sets in Experiment 1

Stimulus	Response triple							
	+++	++-	+ - +	+- -	- + +	- + -	- - +	- - -
	Underextension (13%)		Overextension of letter (36%)		Overextension of color (10%)		Overextension of both (6%)	
Colored letters ^a	312	46	111	198	34	309	18	298
	Underextension (20%)		Overextension of intelligence (55%)		Overextension of emotion (16%)		Overextension of both (8%)	
Cartoon faces ^b	199	51	170	141	49	252	20	242

Note. Zero ratings were excluded from frequencies.

^aResponse triple refers in order to color, letter, and conjunction responses. ^bResponse triple refers in order to emotion, intelligence, and conjunction responses.

Conjunction A and B, then of the eight possible types, four are consistent with the Boolean definition of conjunction (---, -+-, +--, and +++), and four are inconsistent with it. Three of the inconsistent triples indicate overextension (- - +, - + +, and + - +), and one indicates underextension (+ + -). Responses (5% for letters and 20% for faces) in which a zero typicality rating was given were excluded from the analysis. Table 1 gives the frequency of occurrence of each of the eight types of triple as a function of the different stimulus sets and the different conjunctions.

For the colored-letter stimuli there were 163 overextensions and only 46 underextensions. For the cartoon faces there were 239 overextensions and only 51 underextensions. Clearly, a certain level of inconsistent responding (as measured by the frequencies of inconsistent within-subject, within-stimulus triples) is likely to reflect noise in participants' judgments. Many of the stimuli presented in this experiment were deliberately placed close to the borderline of the categories, and so it would be expected that participants might, on occasion, change their response (McCloskey & Glucksberg, 1978). If a participant initially thinks a stimulus is blue (in Block 1) but when seeing it again in Block 5 decides that it is actually green, then even though the participant is responding intersectively, an inconsistent triple may be produced (depending on what is happening with the other constituent decision). If it is assumed that there is no shift in the criterion for categorization between the first (constituent) and later (conjunctive) judgments, then categorization responses to any one stimulus should be equally likely to change from yes to no as from no to yes. As a result it can be shown that if participants are responding intersectively under these circumstances, any instability in responses should throw up as many overextension responses (caused by a change in a constituent categorization from no to yes) as underextension responses caused by the reverse change from yes to no (see Chater et al., 1990, for a mathematical derivation of this result).⁵ As the analysis holds true for any individual stimulus, it also holds true for the raw total of responses. It is clear from Table 1 that many more overextensions than underextensions were made to both stimulus sets, so a simple response instability account is insufficient to explain the data. For the colored letter shapes, 35 of the 48 participants made more overextensions than underextensions; whereas only 10 gave the

reverse pattern ($z = 3.87, p < .001$, sign test). For the faces, 34 participants made more overextensions than underextensions, whereas only 8 gave the reverse pattern ($z = 3.86, p < .001$, sign test).

It is also evident from Table 1 that the letter shape dimension (37%) was more often overextended than the color dimension (12%). Likewise, intelligence (51%) was more often overextended than emotional expression (13%).

An alternative account in terms of random responding might argue that the overextension was the result of participants "guessing" on some of the trials for the conjunctive task, with a 50% probability of saying yes or no. Because only one quarter of stimuli would be positive for both constituents, then such a strategy would be more likely to generate overextensions than underextensions. Furthermore, overextensions should be more common when a stimulus is clearly not in the constituent category than when it is borderline to the category because there will be more opportunity for overextension if the constituent response is negative. Evidence against this account can be seen in Table 2, which shows the mean frequencies of overextension responses for each of the colored letter and face stimuli as a function of whether they were positive (Scale Positions 1-4, in which 1 was the prototype for the conjunction), borderline (Scale Positions 5-7), or negative (Positions 8-11). It is clear in Table 2 that the incidence of overextensions was systematically distributed across the stimulus space in a way that rules out this random guessing account. With the exception of the intelligence dimension, overextensions were most frequently found in the region of the constituent borderlines; whereas a random guessing strategy would predict that

⁵ Although it may appear that there are three times as many opportunities for overextension as underextension (given three possible overextension triples compared with one underextension triple), it has to be remembered that a + + categorization of a stimulus for the two constituents has three possible reasons for changing to a minus for the conjunction (i.e., a change in either constituent or both); whereas a + - constituent pair will give an overextension only if one of the constituents changes and not if the other does or both do. These two intuitive arguments, in fact, balance out to yield the result (proved mathematically by Chater et al., 1990) that noise in responding should predict equal levels of overextension and underextension overall.

Table 2
Mean Incidence per Stimulus of Overextension as a Function of Scale Position on Each Dimension in Experiment 1

Scale position	<i>n</i>	Overextension			
		Color	Letter	Happiness	Intelligence
Positive (1-4)	8	0.00	0.25	0.87	6.12
Borderline (5-7)	13	2.31	6.31	2.31	5.23
Negative (8-11)	8	0.50	3.00	1.50	6.62

Note. *n* = number of stimuli in each band.

they would be more common in the negative part of the scale. The intelligence dimension apparently reversed the pattern for the other dimensions for reasons that are discussed in the *subsidiary results* section.

A similar comparison for underextensions for the letter stimuli showed that there were also more inconsistent responses when either constituent was in the border region than when both constituents were clearly positive (where most underextensions that were due to random responding in the conjunctive task should occur). For the colored letters, the mean rates of underextension were 2.1 for borderline stimuli (*n* = 21) and 0.5 for clear-cut stimuli (*n* = 2). For the faces, the data were less clear, with respective rates of 1.86 for borderline stimuli and 2.0 for clear-cut stimuli. Problems with the intelligence dimension, discussed below, are probably also responsible for this pattern.

A final argument against the random response model is that even with complete guessing at the conjunction stage, overextensions should not exceed underextensions by a factor of more than three; whereas in Table 1, it can be seen that the ratio of overextension to underextension is closer to four in each case.

Subsidiary results. The purpose of collecting typicality ratings was originally to enable a scale of category membership to be constructed and a regression model fit to the data. Hampton (1988) fit a regression model to predict average rated typicality and membership in a conjunction from average rated typicality and membership in the constituent concept categories. The reported fit for item means (averaged over participants) was excellent when an interaction term was included in the regression.

Similar regression models were fit to the present data, but their interpretation was rendered equivocal by differences in the sharpness and reliability of the constituent dimensions. The analysis is not therefore reported. The overextension shown in Table 1 showed an asymmetry between the dimensions of each conjunction, with letter color and facial emotion being less overextended than letter shape and facial intelligence, respectively. Similar dominance effects were reported by Hampton (1988) and have been confirmed by Storms, De Boeck, van Mechelen, and Geeraerts (1993; see also Storms, van Mechelen, & De Boeck, 1994). The color dimension of blue-green used here displayed a particularly sharp category border. As a measure of the sharpness of a dimension, a statistic, *S*, was calculated for each of the four scales, which was the percentage of participants giving the less frequent (nonmodal) categorization response to each stimulus averaged

across the stimuli for each scale (see McCloskey and Glucksberg, 1978, for a similar use of this measure to indicate the fuzziness of category boundaries). If participants had been divided evenly in categorizing all items in the stimulus set, then *S* would have a value of 50%. On the other hand, if all participants agreed on every item, then *S* would be 0%. For the letter stimuli, the *A-H* scale had an *S* of 16%, and the color scale had an *S* of 9%. For blue versus green responses, there was, in fact, only a narrow region of three increments in the center of the scale where any disagreement occurred amongst participants. Thus, although the boundaries of color categories have often been used as an example of vagueness in semantic categorization, it appears that the color categories used in this experiment were more clear cut than was expected. Because the conjunction borderline was strongly tied to this clear-cut border between blue and green, the color dimension came out as strongly dominant in the regression predicting conjunctive membership, and there was little overextension of the color dimension.

For the face stimuli, the regression model showed stronger predictive power for the emotion dimension. However, this dominance could in turn be attributed to a problem with the intelligence dimension for the face stimuli, which was alluded to in the *Consistency and overextension* section. For the faces, the happy-sad scale had an *S* of 13%, and the intelligent-stupid scale had an *S* of 30%. Although the emotional expression scale showed a sufficient spread of vagueness around the category boundary, the intelligence scale suffered from a considerable amount of disagreement amongst participants across the whole scale. Although some participants accepted the manipulation of the height of the features relative to the head outline as an indication of intelligence, others rated the center of the scale as more intelligent than either extreme—perhaps because it appeared more normal. Note that this unreliability in the scale across subjects does not affect the analysis of response triples because that analysis was conducted entirely within subject so that whatever definition of intelligence a participant adopted, overextension was defined relative to that individual definition. The unreliability does, however, seriously compromise the interpretation of the regression data because unreliable variables obviously do not make good predictors, and the data are based on subject means.

Discussion

Experiment 1 demonstrated that overextension effects are not confined to semantic categories of verbal concepts. A considerable amount of overextension was found with visually presented stimuli. The possibility that there would be differences between the relatively simple letter shape stimuli and the more complex faces was also born out in the higher levels of overextension of faces, but there appeared to be even stronger differences between the two dimensions within each stimulus set. The differences in sharpness of the different dimensions, however, make any clear conclusion on relative degrees of overextension difficult to draw. The finding of overextension with these materials argues against two possible accounts of overextension in semantic categories discussed

previously. The ambiguity of instantiation that is involved in categorizing noun classes, such as apple or chess, cannot account for overextension in Experiment 1, in which participants categorized individual instances displayed in front of them. Nor is it possible that overextension in Experiment 1 was caused by nonmonotonic feature inheritance by conjunctions (Hampton, 1987) because the feature structure of the stimuli did not require access to knowledge of instances or background domain theories.⁶

How can the strong degree of overextension observed here be explained? The first account, following Hampton (1987, 1988), is that participants are forming a composite prototype representation of the typical conjunctive member—a typical blue *A* or a typical happy–stupid face—and are then judging similarity of stimuli to this prototype. This model does not of itself predict net overextension. The model needs a further assumption about exactly how participants set the similarity criterion for membership of the conjunction. How similar must an instance be to the composite prototype for it to be considered a member of the conjunctive class? A possible answer comes from data collected by Hampton (1988), in which it was observed for semantic categories such as sports and games that participants appeared to set the similarity criterion for conjunctive membership so that an instance that is a borderline case of each constituent would also be a borderline case for the conjunction. According to this account, overextension is the result of an integrative decision model, in which the criterion for conjunctive membership is anchored to the center of the stimulus space. For example, if judging the conjunctive concept “a blue *A*,” then a stimulus that is both equally likely to be blue or green and equally likely to be an *A* or an *H* would have a 50% chance of being judged to be a blue *A*. This anchoring effect would lead to overextension because the effect of compensation will mean that a stimulus that is above the borderline for one constituent may then fall below the borderline for the other constituent and still be borderline for the conjunction. Note that an “unbiased” independent decision model would predict 25% positive conjunctive categorization if each constituent had a 50% positive response rate.

An alternative account suggested by Chater et al. (1990) proposes that participants use an independent intersective decision process for conjunctions, but one that is based on inflated constituent category memberships. They argued that when faced with a conjunctive category decision, the participants would tend to relax the membership criterion on each constituent dimension and so would overextend membership of the conjunction, even though they were still deciding about each constituent independently and then applying an intersective rule to decide conjunctive membership. Chater et al. argued that because it is naturally harder to satisfy two criteria rather than just one, participants will be less strict in judging conjunctions—just as someone looking for a home will be more willing to compromise on any one dimension, the more different dimensions that have to be maximized.

In the context of the present experiment, one way of capturing this intuition is to suppose that there is a response bias for participants to try to give equal numbers of positive and negative responses to the conjunctive category. If participants are aware that only 25% of the stimuli are likely to satisfy

both the constituent membership constraints, they may be willing to stretch the constituent boundaries to maintain a higher number of conjunctive category members. Whereas 25% of stimuli were judged to belong in both constituents in Experiment 1, 37% of stimuli were judged to belong in the conjunction.

As an attempt to separate these two accounts, I designed the second experiment to introduce an experimental manipulation of response bias. In the standard condition of Experiment 2 I used the same stimulus distribution as in Experiment 1, which provided the opportunity to replicate the results of Experiment 1 with improved stimulus dimensions. In contrast, in the augmented condition of Experiment 2 I increased the frequency of expected conjunctive responses to approximately 50% by introducing in the final phase (i.e., Blocks 5 and 6) of the experiment additional filler stimuli that were good members of the conjunction. The manipulation is similar to that used by Hampton (1988, Experiment 3) to test a similar account of overextension for semantic categories. If Chater et al. are correct, and if participants are loosening their constituent category criteria in response to the frequency with which items can be found to place in the conjunction, then as they proceed through the conjunctive judgment task, participants in the augmented condition should show signs of retightening their criteria for category membership, and hence they should show less overextension than participants in the standard control group. In contrast, if participants are anchoring their similarity criterion for the conjunctive prototype to the center of the stimulus space, then they should be unaffected by the manipulation of relative stimulus frequency.

Experiment 2

Participants for this experiment were randomly assigned to one of two conditions. The first condition—the standard condition—was a partial replication of Experiment 1, with the same 29 stimuli presented for categorization for each constituent and for their conjunction. The second condition—the augmented condition—involved a change in the constitution of the stimulus set. The same 29 stimuli were still presented in all phases of the experiment, but in the final phase in which conjunctive categorizations were made, the stimulus set was augmented by the addition of 14 stimuli, all of which were clear members of the conjunction being tested (see Figure 1). If the response bias account of overextension is correct, then one would expect the relative preponderance of overextension over underextension responses to be less in the augmented condition because the expected frequency of a conjunctive categorization in this condition without any bias would already be 21.25 ($29/4 + 14$) out of 43, or 49%.

In addition to the introduction of the augmented condition in this experiment, a number of changes were made in an

⁶ Although neither of these accounts applies to the present data, it is of course still possible that they are responsible for some of the overextension found with semantic categories. The results of Experiment 1 suggest, however, that there may be more general accounts of overextension that could apply to both the present data and to the earlier findings.

attempt to improve the sensitivity and consistency of the materials. The colored letter shapes in Experiment 1 suffered from a color boundary that was too sharp to be sensitive to much overextension of the color dimension. For Experiment 2, pilot work was done to create a new color dimension that varied between red and orange, which was expected to show a more graduated boundary. The dimension of intelligence used for the cartoon faces in Experiment 1 was replaced in Experiment 2 by a dimension of age, with categories of adult versus child. As a result of pilot work, a new set of facial features was devised that permitted age and emotional expression (happy vs. sad) to be varied independently.

Method

Participants. Participants were 64 student volunteers at City University London who were paid for their participation.

Materials: Letter shapes. The 29 letter shapes used in Experiment 1 were adapted as follows. The shape dimension was varied exactly as in Experiment 1. Color of the shapes was varied in 11 steps along a dimension from red to orange. Pilot work established a graded series of colors, and the background against which the shapes were shown was made a dark grey (RGB = 10, 10, and 10, respectively) to improve the salience of the colors. For both angle and color, pilot studies were used to identify the category boundary, and stimuli were placed symmetrically around the boundary. The same angles were used as before for the letter shape dimension. For the color dimension, the palette ranged from RGB values of 43, 0, and 0, respectively, for a good red to 63, 26, and 5, respectively, for a good orange. In between these extremes, the red value increased in even steps of 2 from 43 to 63, and the green value jumped to 8 on Step 2 and then increased in steps of 2 from 8 to 26. To add luminance, the blue value was also incremented by one on alternate steps. The 29 stimuli were symmetrically distributed around the stimulus space according to the scheme shown in Figure 1. In addition, for each conjunction group, a set of 14 new stimuli was created for the augmented condition (shown by the letter *A* in Figure 1) by completely populating the remaining cells in Figure 1 for the corner of the diagram corresponding to that conjunction. Thus, for example, for the red *A* group, an additional 14 stimuli that were all red *As*, but varied along the first four steps of redness and *A*-ness, were created for inclusion in the conjunctive phase for the augmented condition.

Materials: Faces. The 29 faces were modified by using new combinations of features for eyes, smile wrinkles, and mouth. The apparent age of each face was manipulated independently by varying the size of eyes, the interocular distance, the type of nose, type of eyebrow, and size and shape of the head. Eleven steps of variation for each of the dimensions of age and emotional expression were achieved as a result of pilot studies. Compared with typical children, typical adults had narrower eyes set closer together, narrower eyebrows, sharper noses, and a vertically elongated head shape. Typical happy faces were distinguished by eyes looking ahead rather than down, by smile wrinkles beside the eyes rather than worry wrinkles, and by an upturned rather than a downturned mouth. Examples may be seen in Figure 4. As for the letter shape stimuli, the boundary between the two categories on each dimension was estimated in a pilot study, and the stimuli were designed to fall equally either side of the boundary. The same distribution of stimuli in the stimulus space was used as for the colored shapes (see Figure 1). Additional stimuli for the augmented condition were created in the same way as for the letter shape stimuli.

Procedure. The procedure was a modification of that used for Experiment 1. Changes in procedure were aimed at making the task simpler and clearer for participants. The typicality rating task was



Sad Child



Happy Child



Sad Adult



Happy Adult

Figure 4. Examples of the prototypical faces for the four conjunctions used in Experiments 2–4.

made easier by using a 5-point scale, and participants were no longer given the option allowed in Experiment 1 of giving a zero response meaning "Can't decide." Instruction screens appeared at the start of each block of trials. The color-rating task displayed squares of red and orange above the labels *typical red* and *typical orange*, and instructions were as follows:

This section of the experiment involves color judgments. The task is to decide whether different colored shapes look more like red, or more like orange. For each color, decide if the color is red or orange, and type an *R* or an *O* to indicate your choice. You will then be asked to rate how typical an example it is using a scale from 1 to 5. The squares above show clear examples of red and orange colors as they will appear. Ignore the shape of the displays—just judge the color.

As each stimulus appeared, the question "Red or Orange—R/O" appeared below it. Once a participant typed a response, then the question "How good an example is it—give a number" appeared together with a 5-point scale with responses ranging from 5 (*very*), (3) (*quite good*), to 1 (*poor*).

Instructions and procedure for the letter-rating task were very similar, displaying a typical letter *A* and a typical letter *H*, labeled as such, and the same instructions as indicated above, but referring to shapes rather than colors, and participants were instructed to ignore the color of the displays and just to judge the shape.

The faces were rated for age and for emotional expression. The initial instruction screen varied as a function of whether they had seen the faces in an earlier block or not. The procedure is illustrated by considering those participants who rated age first. On the first occasion that they saw the faces (in Block 1 or Block 2), participants were shown a typical adult and a typical child face (with category labels) and were given the following instructions:

The task is to decide whether line drawings of faces look like

adults or children. For each face, you must decide if you think it looks more like an adult or a child. First, however you will be shown the full set of faces so that you can get an idea of what they will look like.

The faces were then displayed on the screen one at a time at a fixed rate of 2 s per face. This modification was introduced to help participants to develop an idea of the relevant dimensions of variation in the stimuli, and the range of variation to be expected, before making any categorization decisions. Once the display had finished, participants were then given full instructions for the rating task as follows:

The task is now to decide whether each face looks like an adult or a child. For each face, you have to type an *A* for an adult or a *C* for a child, followed by a number to indicate how good an example it is. The decision may be difficult, but just give the best answer you can without spending too long on any one decision. Press any key when you are ready to start.

Those participants rating emotion first had the same procedure, except that typical happy and sad faces were displayed, and instructions directed participants to judge whether faces looked more happy or more sad. On the second occasion of judging the faces (in Block 3 or Block 4), only the second instruction screen was displayed, and the participants carried straight on with the rating task.

For the conjunction blocks of trials (Blocks 5 and 6), the task instructions for letter shapes were as follows for the red *A* conjunction group:

This section of the experiment involves color and shape judgments. The next task is to decide whether each of the letters you saw before looks like a red *A*. For each letter, you must decide if you think it fits this description. Type a *Y* for yes or an *N* for no, followed by a number to indicate how good an example it is. The decision may be difficult, but just give the best answer you can without spending too long on any one decision.

Instructions for the other conjunctions used the appropriate phrase in

place of red *A*. Instructions for the face conjunctions followed the same pattern. The complete procedure took about 30 min.

Design. The comparison of categorization responses to constituent and conjunction categories was within subjects and within stimuli. Participants were divided into standard and augmented conditions, with 32 participants taking part in each condition. As in Experiment 1, the participants in each condition were then divided into four groups of 8 participants each, each group allocated to one of the four possible conjunctions in each stimulus set. Each group was then further subdivided into four subgroups to provide counterbalancing of order of blocks. The standard condition groups judged the same 29 stimuli for each constituent and for the conjunction. The augmented groups judged the 29 standard stimuli for each constituent and then judged the augmented set of 29 + 14 stimuli for the conjunction so that the expected frequency of positive conjunctive categorizations was around 50%. Order of presentation of the constituent concepts in Blocks 1–4 and the conjunctions in Blocks 5 and 6 was fully counterbalanced, subject to the alternation of stimulus sets across blocks. As before, order within each block was randomized for each participant.

Results and Discussion

Consistency and overextension. Table 3 shows the pattern of responding within subjects and within stimuli. Only the standard set of 29 stimuli common to both groups was scored. Looking first at the effect of the manipulation of response frequency for the conjunction, the response bias hypothesis predicts that augmenting the stimulus sets for the conjunction would reduce the degree of overextension by reducing the probability of categorizing the standard 29 stimulus items in the conjunction. The results clearly failed to support this prediction. For the letter shapes, the probability of a positive conjunctive categorization for the standard items was 27% for the standard group and 26% for the augmented group—a

Table 3
Frequencies of Eight Possible Response Triples and Percentages of Inconsistency for Letter and Face Stimuli in Standard and Augmented Groups of Experiment 2

Stimulus	Response triple							
	+++	++-	+ - +	+ - -	- + +	- + -	- - +	- - -
	Underextension		Overextension of letter		Overextension of color		Overextension of both	
Colored letters ^a								
Standard								
Inconsistency	28%		19%		13%		3%	
Frequency	167	65	47	204	29	192	6	218
Augmented								
Inconsistency	30%		18%		12%		2%	
Frequency	166	72	44	194	28	198	5	221
	Underextension		Overextension of age		Overextension of emotion		Overextension of both	
Cartoon faces ^b								
Standard								
Inconsistency	26%		30%		20%		6%	
Frequency	178	62	70	160	44	173	16	225
Augmented								
Inconsistency	25%		41%		21%		16%	
Frequency	175	59	100	142	45	169	39	199

^aResponse triple refers in order to color, letter, and conjunction responses. ^bResponse triple refers in order to emotion, age, and conjunction responses.

negligible difference. For the face stimuli, the probability of positive conjunctive categorization was 33% for the standard group and 39% for the augmented group. Thus, classification of standard items in the conjunction actually increased in the augmented condition, quite contrary to the prediction. Seeing a larger number of good examples of the conjunction actually had the effect of making participants stretch the category boundary even wider. The absence of the predicted effect of this manipulation lends no support to the response bias account of overextension.

Table 3 also shows that the degree of overextension in letter shape stimuli was much less than that for the faces. In fact, when using the relative frequency of overextension and underextension as a guide, there was no significant net overextension for the letter stimuli. The overall positive response rate for the conjunction in each condition was 27% and 26%, respectively. The rate expected on the basis of the proportion of ++ constituent judgments, was in fact very similar—25% and 26%, respectively. Across subjects, there were 28 showing more overextension than underextension, and there were 27 showing the reverse effect. This balance was found equally in each group (13:14 and 15:13 for the standard and augmented groups, respectively). In contrast with Experiment 1, therefore, the letter shape stimuli showed no pattern of overextended conjunctions beyond that which could be attributed to noise in participants' categorization.

For the face stimuli, a different pattern emerged. Across subjects, there were 49 showing more overextension than underextension, and there were only 12 showing the reverse effect ($p < .001$, sign test). This proportion was similar in each group (23:8 and 26:4 for the standard and augmented groups, respectively). Unlike the letter stimuli, participants were clearly overextending the face conjunctions. The effect of the augmented condition was to increase the overextension of the age dimension (from 30% to 41%) and to increase the number of double overextensions (from 6% to 16%), although the number of underextensions was unaffected. Because these effects were in the opposite direction to those predicted by the response bias hypothesis, there was clearly no support for the idea that participants overextend because they wish to maintain a 50% frequency of yes responses when categorizing the conjunctive category.

In Experiment 1, the color dimension was less overextended than the letter shapes; the same was true in Experiment 2. Although much less sharply defined than the blue-green dimension in Experiment 1, the color dimension was still slightly sharper than the letter dimension ($S = 13%$ for colors vs. 18% for letters). The greater overextension of age compared with emotion is also consistent with the results of Experiment 1, in which emotional expression was also the less overextended dimension. However, there was no difference in the sharpness of the two dimensions ($S = 24%$ and 25%, respectively, for emotion and age) so that differences in fuzziness cannot explain the large difference in the rate of overextension between the two dimensions.

Experiment 3

In Experiment 2 I tested a response bias account of overextension and failed to find any evidence of sensitivity of overextension to the frequency of positive conjunctive stimuli

to be categorized. Experiment 3 was designed to test another possible account of overextension based on the difference in the kind of response required in Experiments 1 and 2 for the constituent and the conjunctive categorization responses. I noted earlier that although a yes-no decision was used in the conjunction categorization of Blocks 5 and 6, the earlier categorizations were presented as a choice between two poles of a dimension (red or orange and happy or sad).⁷ It is possible, therefore, that the apparent overextension may reflect a difference in criterion based on whether a contrast set is explicitly mentioned. If participants have a bias to saying yes to categorization questions in general, then they may be led to overextend the conjunctions (in which a yes-no decision was used) relative to the constituents that involved a choice between contrasting categories, which could not be affected by a yes bias. To control for this possibility, I used in Experiment 3 a yes-no decision throughout so that there was no confounding of type of response with the constituent-conjunction factor. If the difference in type of response is a contributory factor to overextension, then a reduced level of overextension, compared with Experiments 1 and 2,⁸ should be expected.

Method

Participants. Participants were 32 student volunteers at City University London who were paid for their participation.

Materials. The identical 29 letter shapes and 29 faces used in the standard condition of Experiment 2 were used in this experiment.

Procedure. The procedure was a modification of that used for the standard condition of Experiment 2. Instruction screens appeared at the start of each block of trials. As before, the color-rating task displayed squares of red and orange above the labels *typical red* and *typical orange* and the instructions. The instructions referred to the two poles (red and orange) but then followed the pattern used for conjunctions in Experiment 2, in which participants were asked to decide if each color was of a particular category (e.g., red) and to type a *Y* for yes or an *N* for no to indicate their choice. If the answer was yes, then they were asked to rate how typical an example the stimulus was by using a scale with responses ranging from 1 to 5. If the answer was no, then they were asked how close to the category the stimulus was on the same 1-5 scale. As each stimulus appeared, the question appeared below "Is this Red—Y/N?" Once a participant typed a response, then the question "How good (close) an example is it—give a number" appeared together with a 5-point scale with responses ranging from 5 (*very*), 3 (*quite good*), to 1 (*poor*). The choice of pole of each dimension for the participant to rate was determined by the conjunction group that the participant was allocated to. Thus, if the participant was going to be judging the conjunction of orange *A* in Block 5 or Block 6, the earlier tasks would ask about whether the colors were orange and whether the shapes were *As*. Instructions and procedure for the letter-rating task were the same, displaying a typical letter *A* and a typical letter *H*, labeled as such, and presenting instructions concerning shape rather than color. The faces were rated for age and for emotional expression as before. The procedure for faces was identical to Experiment 2, but with the same modifications as were made for the

⁷ I am indebted to Dedre Gentner for this suggestion.

⁸ This account does not of course explain the differences between the two sets of stimuli in Experiment 2. Nor does it apply to the results of experiments with semantic categories (Hampton, 1988) in which all categorizations were of a yes-no type, with no clearly defined contrast set for each constituent.

letter shapes. For the conjunction blocks of trials (Blocks 5 and 6), the task instructions and procedure were identical to Experiment 2.

Design. The comparison of categorization responses to constituent and conjunction categories was within subjects and within stimuli. The design, as well as counterbalancing, was identical to that used for the standard condition of Experiment 2.

Results and Discussion

Consistency and overextension. Two participants were found to have given all positive responses in the conjunction task to both sets of stimuli, and so were replaced. Table 4 shows the pattern of responding within subjects and within stimuli.

Table 4 shows that the degree of overextension in letter shape stimuli (18% for the letter dimension and 20% for the color dimension) was overall slightly greater than that found in the standard condition of Experiment 2 (19% and 13%, respectively). The amount of overextension increased and the amount of underextension decreased compared with that condition. Across participants, the difference between overextension and underextension for Experiment 3 was significant on a Wilcoxon test ($z = 2.58, p < .01$; Siegel, 1956) but not on a sign test (17 participants showing greater overextension, and 10 participants showing less, $p > .05$). For the face stimuli, the results were very similar to those for the standard condition of Experiment 2. Across participants, 25 showed greater overextension, and only 5 showed the reverse effect (significant on a sign test, $p < .001$). Most important, there was no indication that the degree of overextension observed in Experiments 1 and 2 was reduced in any way when the constituent judgments were made on the same yes-no basis as the conjunctive judgments. Overextension in the earlier experiments was clearly not the result of a bias to respond yes to categorization questions. In fact, when the probability of responding yes to each pole of each dimension was summed for each stimulus, the total was between .96 and .99, indicating a very slight bias against responding yes to constituent categories.

In Experiments 1 and 2 I found that the color dimension was less overextended than the letter shape dimension. In Experiment 3, the degree of overextension was much more equal, with 18% overextension of letter shape and 20% overextension of color. In terms of the sharpness of the category boundaries, letter shape was slightly less well defined than color ($S = 20\%$ and 16% , respectively). For the face stimuli, the greater

tendency to overextend age (50%) compared with emotion (20%), as seen in Experiment 2, was confirmed, while the boundaries were equally well defined ($S = 25\%$ and 22% for age and emotion, respectively).

In conclusion, the results of Experiments 2 and 3 have ruled out two possible accounts of the overextension phenomenon based on participant bias in responding. Neither the reduced frequency of positive categorizations for conjunctions (Experiment 2) nor the possible bias of participants to prefer yes to no responses for categorizations in general (Experiment 3) can account for overextension. Furthermore, the pattern of all three experiments confirms that overextension occurs for conjunctions of visually represented categories and that the more complex stimuli (the faces) show a larger and more consistent overextension effect than the structurally simple colored letters.

Experiment 4

In Experiment 4, I developed a further measure that could possibly help to differentiate between alternative accounts of overextension. This measure was *compensation*. According to the composite prototype model, participants decide about conjunctive membership by a similarity computation across the features of a stimulus. Assuming (as seems reasonable with these simple stimuli) that attribute inheritance by the conjunctive prototype is *compositional* (i.e., there is no inheritance failure, and there are no emergent attributes), then using similarity as a basis for categorization necessarily implies that a decrease in the degree of match of one stimulus dimension could be compensated for by a corresponding increase in the degree of match of another stimulus dimension, leaving the stimulus at the same level of conjunctive membership (this is a characteristic of integrative decision models; Ashby & Gott, 1988). For example, if one is deciding whether stimuli are red *H*s, then a stimulus that is a very typical shade of red need not be as good of an *H* to pass the membership criterion, as one which is a less typical shade of red. Degrees of typicality and membership in the two constituents can compensate for each other in arriving at the conjunctive categorization. In Experiment 4 I used a stimulus array designed to test whether this compensation occurs for each of the constituent dimensions in Experiment 2. The sampling of the stimulus space was changed

Table 4
Frequencies of Eight Possible Response Triples and Percentages of Inconsistency for Both Conjunctions in Experiment 3

Stimulus	Response triple							
	+++	++-	+-+	+--	-++	--+	---	---
	Underextension (18%)		Overextension of letter (18%)		Overextension of color (20%)		Overextension of both (4%)	
Colored letters ^a	191	43	40	179	43	169	10	253
	Underextension (27%)		Overextension of age (50%)		Overextension of emotion (20%)		Overextension of both (16%)	
Cartoon faces ^b	159	60	110	108	47	193	40	211

^aResponse triple refers in order to color, letter, and conjunction responses. ^bResponse triple refers in order to emotion, age, and conjunction responses.

Dimension B	1	C			S*	S*	S*					
	2		S*		S*	S*	S*			S*		
	3				S*	S*	S*				*	
	4				S*	S*	S*					
	5	S*	S*	S*	S*	S*	S*	S	S	S	S	
	6	S*	S*	S*	S*	S*	S*	S	S*	S	S	
	7	S*	S*	S*	S*	S*	S*	S	S	S	S	
	8				S	S	S				*	
	9				S	S*	S					
	10		S*		S	S	S			S*		
	11			*	S	S	S	*				
		1	2	3	4	5	6	7	8	9	10	11
		Dimension A										

Figure 5. Distribution of target stimuli for Experiment 4. In the first four blocks of trials, all stimuli marked with an *S* were used. In the conjunction trials (Blocks 5 and 6), the stimuli shown depended on which conjunction was being judged. The distribution shown by asterisks was that for the conjunction in the corner (1, 1), for which the prototype is indicated by the letter *C* (conjunction). (The *C* stimulus was not presented to participants.) The other three conjunctions had the corresponding distribution defined with respect to the other three corners of the diagram.

so that for each constituent category (e.g., red or happy) there would be a range of stimuli that would all be clear members of the category, but that would differ in their typicality.

To arrange this distribution, I redefined the steps along each constituent dimension, using new pilot data, so that the first four steps from each constituent were all likely to attract 95% or better consistent categorization in the constituent but would nonetheless differ in their typicality. The question of compensation then comes down to the question of whether the differences in typicality among stimuli that are all clearly in one of the constituent categories will affect the overall classification in the conjunctive category by compensating for a borderline membership in the other constituent. The critical comparison is therefore between particular sets of stimuli, all of which are clear members of the first constituent, and all of which are equally borderline for the second constituent, but some of which are judged to be more typical than others of the first constituent. If such stimuli are equally often included in

the conjunction regardless of their typicality in the first constituent, then this result would indicate that the two constituent concepts are being decided independently of each other, and membership in the conjunction is only positive when both constituent decisions are positive. On the other hand, if typical items are more often included in the conjunction than atypical items, then there would be further support for the similarity-based composite prototype model.⁹

⁹ Compensation is also consistent with a model in which the two constituent decisions are made first and then combined, but in which the two decisions are not independent—for example, the second decision may be biased by the outcome of the first. Such a model is difficult to distinguish from a composite prototype model without developing new experimental procedures. It is nonetheless valuable to be able to test whether the independence assumption is, in fact, false. An example of a sequential decision model with interdependence between constituent categorizations is given in the *Discussion* section.

Method

Participants. Thirty-two student volunteers at City University London were paid £2 for their participation.

Materials. The stimuli from Experiment 2 were adapted as follows. After further pilot studies, I redefined the physical feature dimensions corresponding to the 11 steps on each constituent dimension to extend further into each category. On the basis of the pilot studies, I expected that Steps 1-4 and Steps 8-11 would be clear members of the respective constituent categories, and so they would differ only in their typicality. The colors used were red and orange as in Experiment 2, but the selection of colors was extended toward each end of the scale. The angles used for creating the letter shapes did not need adapting as they already fit the required constraint. The domain of cartoon faces was likewise resampled to provide a wider range of clear constituent examples that differed in typicality.

Procedure. The same procedure was used as in Experiment 2.

Design. The design of the stimulus distribution is shown in Figure 5. For the constituent judgments in Blocks 1-4, the same stimulus distribution of 61 stimuli was used for all participant groups. Stimuli used are shown with an S in Figure 5 and were designed to sample the critical region of the stimulus space for testing the compensation hypothesis. For the conjunctions themselves, the stimulus distribution differed according to which conjunction was being judged. Figure 5 shows with asterisks those stimuli presented when the conjunction to be judged corresponded to the 1,1 combination of the two constituents (shown by C for conjunction in the figure). Thus, the stimuli were largely sampled from the two borderline regions for the conjunction in question. There were, in addition, six filler items (shown as single asterisks) added to even up the number of expected positive and negative conjunction responses. Four groups of participants each judged one of the conjunctions for each set of stimuli. The order of constituents in Blocks 1-4 and of conjunctions in Blocks 5-6 was balanced as in Experiment 2.

Results and Discussion

Consistency and overextension. An analysis of response triples was conducted for those stimuli in each participant group that were presented both for constituent and conjunction categorizations. The results are shown in Table 5. Both the letter and the face stimuli showed a net overextension effect in this experiment, although, as in earlier experiments, there was a greater amount of overextension for letter shape than for color categories and for age than for emotion categories. Across subjects, there was a significant degree of overextension for the face stimuli, with 26 participants showing more overextension than underextension, but only 3 showing

the reverse pattern ($p < .001$, sign test). For the letter shapes, however, there was no significant overextension (15 participants vs. 13 showing greater over- or underextension, respectively). The pattern of Experiment 2 was, therefore, replicated, with only the faces showing a significant overextension effect overall. The sharpness of the category boundaries was calculated for all groups across all 61 stimuli. For the letters, S was 6% for color and 20% for shape. For the faces, both age and emotion had an S value of 22%. Sharpness of the borderlines was, therefore, not a likely explanation of the differences in rates of overextension between dimensions.

Compensation. The experiment was designed to provide a test of the hypothesis that there may be compensation between the two constituent dimensions in categorizing stimuli in a conjunction. To test this hypothesis, I performed an initial analysis by using the objective (scale value) typicality based on the levels of the stimulus dimensions as defined by the experimenter. The critical stimuli for the test of compensation were those with a clear membership in one category (the *clear constituent*) and with marginal membership in the other (the *marginal constituent*). The stimulus dimensions were designed with the intention that the first four steps along each dimension would be clearly in each constituent category while differing in typicality. Analysis of the responses given to different constituent category judgments showed that stimuli lying between two and four steps from each end of the stimulus dimensions fulfilled a strict criterion of at least 97% of participants giving the same response. Analysis of compensation was based on these stimuli (as indicated by a footnote in Table 6). For example, there were two levels of the age dimension with 97% or better categorization as children; whereas for the color dimension, there were four levels that were clearly orange. There were, therefore, between 6 and 12 stimuli over which compensation of each dimension could be tested, corresponding to the two to four levels of the clear constituent, combined with three levels of the marginal constituent (Levels 5, 6, and 7—see Figure 5).

A compensation effect would be seen if for those stimuli that are all clear examples of (for instance) a letter H , their degree of typicality as an H influences categorization probability in the conjunction (be it a red H or an orange H). Thus, for a given level of the marginal constituent, compensation would be found if the probability of a positive conjunctive categorization

Table 5
Frequencies of Eight Possible Response Triples and Percentages of Inconsistency for Both Conjunctions in Experiment 4

Stimulus	Response triple							
	+++	++-	+-+	+- -	-++	--+	---	---
	Underextension (27%)		Overextension of letter (30%)		Overextension of color (11%)		Overextension of both (5%)	
Colored letters ^a	337	126	95	217	34	284	8	147
	Underextension (28%)		Overextension of age (42%)		Overextension of emotion (19%)		Overextension of both (14%)	
Cartoon faces ^b	285	110	143	195	67	279	23	145

^aResponse triple refers in order to color, letter, and conjunction responses. ^bResponse triple refers in order to emotion, age, and conjunction responses.

Table 6
Percentage of Categorization for One Constituent Category as a Function of Stimulus Dimension Scale in Experiment 4

Scale step	Letter shapes		Faces	
	p(Letter H)	p(Red)	p(Happy)	p(Adult)
1	99 ^a	99 ^a	99 ^a	97 ^a
2	99 ^a	100 ^a	99 ^a	98 ^a
3	99 ^a	100 ^a	92	94
4	91	32	86	92
5	66	12	61	78
6	47	6	51	43
7	25	4	41	52
8	8	2 ^a	12	30
9	3 ^a	1 ^a	2 ^a	6
10	1 ^a	1 ^a	1 ^a	1 ^a
11	0 ^a	1 ^a	0 ^a	1 ^a

^aIndicates stimuli used for the compensation analysis.

decreased from Level 1 to Level 3 of the clear stimulus dimension. Table 7 shows the percentage of positive conjunctive categorizations for each of the stimulus dimensions as a function of level for the clear constituent.

It may be seen that there was no evidence of compensation when color was clear cut but that some compensation existed when letter shape was the clear-cut constituent. For the face stimuli, emotion showed no compensation at all; whereas age (adult vs. child) showed the predicted decline in response from Level 1 to Level 2.

Given the small effect sizes, direct statistical analysis of these data was unlikely to provide strong evidence of an effect, and so a more sensitive way of analyzing the results was adopted. For some stimuli it may be problematic to assume that the objective levels of the stimulus dimension correspond exactly to differences in perceived typicality for each participant. Particularly for the face stimuli, there may be some participants who feel that Level 1 is actually less typical than Level 2. Others may see no difference in typicality between two shades of red or orange. A method of analysis was therefore developed that would take account of individual participants'

Table 7
Compensation Analysis Using Objective Typicality of Stimuli

Clear-cut category	Typicality on clear-cut dimension			
	Level 1	Level 2	Level 3	Level 4
Red	48	46	48	—
Orange	35	50	52	48
Letter H	60	46	48	—
Letter A	54	54	50	—
Happy	73	77	—	—
Sad	65	65	65	—
Adult	40	27	—	—
Child	50	46	—	—

Note. The table shows the percentage of positive conjunction responses for borderline cases on the marginal constituent dimension as a function of level of typicality on the clear-cut constituent dimension. Levels of the clear-cut constituent were only included where the probability of being judged to belong to the clear-cut constituent was at least 0.97. Level 1 was the highest typicality, and Level 4 was the lowest.

own ratings of the typicality of the stimuli on the clear dimension.

For each participant and for each of the eight categories, the participant's own typicality ratings of the critical 6 to 12 stimuli for the clear constituent were extracted from the data, and the stimuli were then divided into high- and low-typicality sets on the basis of a median split applied to that individual participant's typicality ratings for those stimuli. Thus the high-typicality set contained the stimuli which that participant rated as of higher typicality, regardless of their "objective" level on the stimulus dimension. The frequency of a positive categorization in the conjunction (by that participant) was then compared between those stimuli that were rated as more typical for the clear constituent by the participant (the high-typicality set) and those that were rated as less typical (the low-typicality set). If the participant placed more high-typicality than low-typicality stimuli in the conjunction, then the participant was counted as providing positive evidence for the compensation hypothesis. If the reverse pattern occurred, then the participant was counted as going against the compensation hypotheses. On the basis of this individual analysis, each participant was classified for each category as positive (conjunctive categorization favoring higher clear-constituent typicality items), negative (conjunctive categorization favoring lower clear-constituent typicality items), or neutral (the association could not be calculated, for example, all of the stimuli in question were given the same typicality rating or the association was neither positive nor negative). Table 8 shows the frequency of positive and negative participants for each category of the four stimulus dimensions.

It is apparent that when analyzed at the individual subject level, there was evidence for compensation for the letter dimension (9 participants positive and 1 negative, $p < .05$, sign test) and for the age dimension (15 positive and 6 negative, $p < .05$, sign test, one-tailed) but not for the color (8 vs. 9) or emotion (6 vs. 10) dimensions. These results follow the trend identified in Table 7. To be able to interpret these results as supporting the compensation hypothesis, it is important that the stimuli that were rated as more typical of the clear constituent by a participant were not also rated more typical or

Table 8
Net Compensation Effect

Clear-cut category	Positive	Negative
Letters		
Color		
Red	2	4
Orange	6	5
Letter		
H	7	0
A	2	1
Faces		
Age		
Child	9	3
Adult	6	3
Emotion		
Happy	1	3
Sad	5	7

Note. n for each row is 16, and for the total it is 32. Participants with neither positive nor negative associations have been excluded.

better members of the marginal constituent by that participant. As a check, the analysis was redone for those same stimuli rated individually as more and less typical for the clear constituent, but this time the analysis was directed at the relation of this factor to the typicality ratings and the probability of membership in the marginal constituent. The analysis showed a net negative association between typicality for the clear constituent and typicality or degree of membership for the marginal constituent, so the compensation observed could not have been an artefact of a positive association between constituents.

The experiment demonstrated that compensation occurred for only one of the dimensions in each stimulus set. When color was the clear constituent, there was no effect of typicality of color on conjunctive membership of stimuli that were marginal between *A* and *H*. However, when the letter shape categorization was clear, then the typicality of the letter shape did influence conjunctive categorization of stimuli in which color was marginal between red and orange. (This effect emerged in spite of a lower sensitivity in the case of the red–orange dimension, for which the marginal Stimulus Levels 5, 6, and 7 were in fact strongly biased toward the orange category—see Table 6.) For the face stimuli, a similar asymmetry was found. That is, when the emotion dimension was the clear constituent, there was no effect of how typical a happy or sad face was on the conjunctive categorization of stimuli that were marginal on the age dimension. However, when the age dimension was the clear one, the typicality of adult or child faces did affect the conjunctive categorization of stimuli that were marginal on the emotion dimension. This asymmetry could not be accounted for by differential sensitivity of the design to pick up compensation effects in the different dimensions. The two dimensions that showed compensation as the clear constituent were actually those with the smaller number of stimuli over which the analysis could be calculated (18 stimuli for letter shapes as opposed to 21 for colors and 12 for age as opposed to 15 for emotion). The compensation effects emerged in spite of a lower power to detect them in those constituents.¹⁰

General Discussion

The four experiments reported in this article confirmed that overextension effects in conjunctive categorization are not confined to semantic categories in which stimuli are verbal labels for subclasses of items but can also be obtained with concrete visual stimuli. For ambiguous colored letters, overextension was found in two of the four experiments. For the cartoon faces, overextension was found in all of the experiments. Dominance relations between constituent categories were also found, with color dimensions tending to show less overextension than letter shape (in four of the five conditions), and emotional expression of cartoon faces showing considerably less overextension than either apparent intelligence or apparent age. Between the two stimulus domains studied here, there was also a clear tendency for the faces to show a higher degree of overextension. This was true even when the absolute levels of underextension were the same (Experiments 2 and 4). For the purpose of providing an overall view, Table 9 summarizes the results of all four experiments.

Dominance between dimensions within each stimulus set was clearly present for the face stimuli, with the happy–sad emotion dimension showing less overextension than either intelligence (Experiment 1) or age (Experiments 2–4). This dominance was found regardless of variations in the fuzziness of the category boundaries for the two dimensions. For the colored letter shapes the pattern of dominance was less unequivocal. Table 9 shows that those conditions in which color tended to show less overextension than shape also tended to be those conditions in which the fuzziness of the two dimensions was most unequal. It could be, therefore, that the apparent dominance of the color dimension in Experiments 1 and 4 was the result of differences in how close to the category boundary the stimuli were distributed.

In Experiment 2 I tested an account of overextension based on the reduced expected frequency of conjunctive categorizations and found no evidence for any tightening of the conjunctive category boundary when the frequency of positive exemplars was increased to 50%. Thus, the overextension commonly observed in these conjunctive tasks was not due to any flexible strategic decision on the part of participants that would lead them to relax constituent membership criteria to find sufficient stimuli for the conjunction. If the Chater et al. (1990) account is correct, then it must be assumed that the tendency to relax constituent criteria when considering conjunctions is a relatively automatic reaction, which is insensitive to actual observed frequencies of stimuli of different types. To use Chater et al.'s own analogy, if one needs to find a home that meets a conjunction of several criteria rather than just a single one, then one may be willing to relax each individual criterion to increase one's chances of finding a suitable choice. However, if (as in this experiment) the market is flooded with good examples of houses meeting the conjoint criteria, then by the same argument, participants should be able to afford to become more choosy; this, they failed to do.

In Experiment 3 I tested a second artifactual account of overextension and showed that the preponderance of overextension responses was not explained by a general bias to respond yes to categorization questions.

In Experiment 4, I further distinguished between alternative accounts of conjunctive categorization, by finding evidence of compensation between constituent typicalities. When two stimuli were equally clear members of a constituent, but one was judged as more typical by an individual participant, then the more typical stimulus was also more likely to be judged by that participant to belong in the conjunction. This result rules out models in which participants might judge each constituent category independently and then combine the two decisions intersectively. The compensation effect was, however, not found in all stimulus dimensions but only in one dimension for each of the stimulus domains.

Assuming that this asymmetry in compensation was not owing to differences in the range of typicality of the critical stimuli, a correspondence was observed between those stimulus dimensions that showed overextension effects and those that provided compensation. Interestingly, it was not the

¹⁰ It is, however, possible that the variation in typicality across those dimensional levels may not have been equivalent.

Table 9
Summary of Results for Experiments 1-4

Experiment	Inconsistencies						
	Underextension	Overextension			Overextension > underextension ^a	Fuzziness	
		Shape	Color	Both		Shape	Color
Colored letters							
Experiment 1	13	36	10	6	+	16	9
Experiment 2 ^b						18	13
Standard	28	19	13	3	-		
Augmented	30	18	12	2	-		
Experiment 3	18	18	20	4	+	20	16
Experiment 4	27	30	11	5	-	20	6
Inconsistencies							
Experiment	Underextension	Overextension			Overextension > underextension	Fuzziness	
		Intelligence/ Age	Emotion	Both		Intelligence/ Age	Emotion
	Cartoon faces						
Experiment 1	20	55	16	8	+	30	13
Experiment 2 ^b						25	24
Standard	26	30	20	6	+		
Augmented	25	41	21	16	+		
Experiment 3	27	50	20	16	+	25	22
Experiment 4	28	42	19	14	+	22	22

^a+ indicates overextension was significantly greater than underextension, - indicates no significant difference. ^bFuzziness was calculated for the two conditions of Experiment 2 together.

correspondence that might have been expected. Letter shape and age were the "weaker," more labile dimensions when it came to overextension. Participants were much more willing to change their minds about classifying along these dimensions when constituent and conjunctive classifications are compared. Yet it was these same two dimensions that appeared able to influence conjunctive classification through variations in the typicality of stimuli that were otherwise equally clear members of the constituent category. In some sense they were therefore "stronger" dimensions in producing compensation effects.

Accounts of Dimensional Asymmetry

The experiments have thrown up asymmetries between the dimensions of each of the two stimulus sets, and as with any such unexpected result, accounts may be proposed that must, however, remain speculative and subject to confirmation by further experimentation with a wider range of stimulus materials and experimental procedures. Because of the potential theoretical interest of these asymmetries, three possible accounts are offered here, with the understanding that they remain as post hoc speculations.

To explain the asymmetry in dimensions within each stimulus set, one might look first to correlated differences between the dimensions in each set. One obvious difference between the color and emotion dimensions and the shape and age dimensions is simply the order in which they were expressed in the conjunction. Hampton (1988) found that when a relative clause conjunction such as "A which are B" is formed, the qualifier concept B is more influential in the conjunction than the head noun A. Although order within conjunctions is an

obvious confound in the present experiments, it is not particularly helpful as an explanation of the result. In particular, it is legitimate to ask why one order of the two concepts is so much more natural than the reverse. "Red *H*" is far more natural as a phrase than is an "H-shaped red object." Similarly, it is more natural to refer to a "happy child" than to a "child-aged happy person." The English language does not permit the experimental manipulation of order of concepts like the present ones, in which one is a noun and the other a modifier, without introducing a more serious confound of naturalness of expression. It may then be possible that concepts that are more naturally expressed as nouns (such as child or *H*) are more liable to overextension than are those that are commonly expressed as adjectives (such as happy or red). This effect could be related to the notion that the head noun acts as a given presupposition while the modifier adjective is the new information (Clark & Haviland, 1977).

It may be possible to manipulate order within a more neutral construction in which both concepts are expressed as qualifiers (e.g., a shape that is both red and an *H*, or a shape that is both an *H* and red) or to construct new materials in which both concepts are expressed as nouns or both as adjectives. It remains to be seen whether these manipulations of linguistic form would in fact eliminate asymmetries.

A second account based on left-right order of expression might be given on the following lines. Suppose that participants first consider the qualifier concept—the color or emotion dimension—and then reach one of three decisions: yes, no, or maybe. Suppose further that the typicality of a yes response is not carried forward, and so it cannot lead to compensation

effects on the second concept. If the answer is yes, then on considering the second dimension (shape or age), a participant may choose to relax their criterion of membership for the second constituent, thus overextending the conjunction on the second constituent. A no answer to the first category might lead to an automatic rejection of the stimulus from the conjunction without considering the second concept, so there would thus be no overextension of the first concept. Finally, if the answer to the first categorization were maybe, then the participant may be swayed by the typicality of a clear positive categorization for the second concept to let the maybe become a yes. A maybe together with a typical yes would be more likely to generate a positive conjunctive categorization than a maybe with an atypical YES.

A third account of these dimensional differences may relate to the familiarity of the different conceptual distinctions. Consider the colored letter stimuli. Color categories may be considered to be more well established in people's minds than are letter shapes. It is certainly the case that people will have frequently encountered the full range of variation of the colors seen; whereas the intermediate letter shapes are relatively novel. It is, therefore, more likely that people's color categories are more firmly fixed than their letter shape categories. This constancy was seen in Experiment 1 in which the blue-green dimension had a surprisingly high level of agreement across subjects, with only one level of the dimension providing any category fuzziness. Even when the color categories were changed to provide a wider range of fuzzy categorization at the borderline between red and orange, the dimension remained relatively immune to overextension. If color categories are more familiar to participants, then it could also be argued that the emotional expression of faces as happy or sad is also a highly familiar category—and one in which exact classification is probably more relevant to everyday concerns than for the categories of apparent intelligence or age.

According to this account, the more familiar categorizations of color and emotional expression are more immune to boundary shifts when placed in conjunctions and so are much less inclined to show overextension. The strong and stable category boundary, however, can also lead to an effect similar to that of categorical perception of phonemes (Harnad, 1987). Not only is the category boundary hard to shift, but the differences between stimuli on the same side of the boundary are less well marked. That is to say that if a shape were red, or a face happy, the strong category boundary for color and emotion leads to there being little discrimination between shades of red or ranges of happiness that vary in their typicality. Variations in typicality of concepts with strong boundaries do not compensate for marginality on other dimensions.

Conclusion

The main conclusion to be drawn from these experiments is that the phenomena of overextension and category dominance may also be found with categories of visual stimuli classified along familiar conceptual dimensions. Two alternative accounts of the phenomenon of overextension framed in terms of response biases were not found to be correct. The establish-

ment of the finding in Experiment 4 that typicality in a clear category can sometimes compensate for marginality in the second constituent is of prime importance in providing support for the composite prototype model and ruling out models in which participants make independent decisions about membership in each constituent. A number of interesting subsidiary findings emerged from the results—notably, that stimuli based on a family resemblance prototype concept gave rise to more reliable overextension effects than did a set based on unidimensional cognitive reference point prototypes and that within each stimulus domain there were intriguing correspondences between the dimensions that are more prone to overextension and those that are more liable to provide compensation effects. Some accounts of these subsidiary results were offered, which must remain speculative at present.

References

- Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 33–53.
- Chater, N., Lyon, K., & Myers, T. (1990). Why are conjunctive concepts overextended? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 497–508.
- Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–40). Norwood, NJ: Ablex.
- Hampson, S. E. (1990). Reconciling inconsistent information: Impressions of personality from combinations of traits. *European Journal of Personality*, *4*, 157–172.
- Hampton, J. A. (1979). Polymorphous concept in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*, 441–461.
- Hampton, J. A. (1986, November). *Conjunction, disjunction and negation of natural concepts*. Paper presented at the 27th annual meeting of the Psychonomic Society, New Orleans, LA.
- Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition*, *15*, 55–71.
- 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.
- Hampton, J. A. (1991). The combination of prototype concepts. In P. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 91–116). Hillsdale, NJ: Erlbaum.
- Hampton, J. A. (1993). Prototype models of concepts: Introduction. In I. van Mechelen, J. A. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67–95). London: Academic Press.
- Harnad, S. (Ed.). (1987). *Categorical perception*. Cambridge, England: Cambridge University Press.
- Huttenlocher, J., & Hedges, L. V. (1994). Combining graded categories: Membership and typicality. *Psychological Review*, *101*, 157–165.
- Jones, G. (1982). Stacks not fuzzy sets: An ordinal basis for prototype theory of concepts. *Cognition*, *12*, 281–290.
- Kunda, Z., Miller, D. T., & Clare, T. (1990). Combining social concepts: The role of causal reasoning. *Cognitive Science*, *14*, 551–578.
- Massaro, D. W. (1989). [Multiple book review of *Speech perception by ear and eye: A paradigm for psychological inquiry*]. *Behavioral and Brain Sciences*, *12*, 741–794.
- McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well-defined or fuzzy sets? *Memory & Cognition*, *6*, 462–472.

- Murphy, G. L. (1988). Comprehending complex concepts. *Cognitive Science*, 12, 529-562.
- Oden, G. C. (1977). Integration of fuzzy logical information. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 565-575.
- Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9, 35-58.
- Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). New York: Academic Press.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-232.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- Siegel, S. (1956). *Nonparametric statistics*. New York: McGraw-Hill.
- Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science*, 12, 485-527.
- Storms, G., De Boeck, P., van Mechelen, I., & Geeraerts, D. (1993). Dominance and non-commutativity effects on concept conjunctions: Extensional or intensional basis? *Memory & Cognition*, 21, 752-762.
- Storms, G., van Mechelen, I., & De Boeck, P. (1994). Structural analysis of the intension and extension of semantic concepts. *European Journal of Cognitive Psychology*, 6, 43-75.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.

Received April 21, 1994

Revision received April 7, 1995

Accepted April 10, 1995 ■

New Editor Appointed

The Publications and Communications Board of the American Psychological Association announces the appointment of Kevin R. Murphy, PhD, as editor of the *Journal of Applied Psychology* for a six-year term beginning in 1997.

As of March 1, 1996, submit manuscripts to Kevin R. Murphy, PhD, Department of Psychology, Colorado State University, Fort Collins, CO 80523-1876.