EMORY COGNITION PROJECT

Report #2
January, 1984

# Constructing Representations of Categories From Different Points of View

Lawrence W. Barsalou

and

Daniel R. Sewell

Emory University



## Abstract

Two experiments demonstrated that adopting different points of view when processing a category leads to different representations of the category. In Experiment 1, taking various international points of view (African, American, Chinese, French), domestic points of view (Businessman, Hippie, Housewife, Redneck), or one's own point of view resulted in highly divergent graded structures for the same category. Approximately equal shifts in graded structure occurred for common taxonomic and goal-derived categories. Experiment 2 showed that subjects can be extremely accurate in generating graded structures from points of view that are not their own. Undergraduate and graduate students generated graded structures from the faculty point of view that were identical to those generated by faculty taking their own point of view; in addition, graduate students accurately generated graded structures from the undergraduate point of view, although faculty were slightly inaccurate from this point of view. In both experiments, subjects adopting the same point of view always showed significant agreement in the graded structure they generated. A theoretical account of how points of view construct different representations of a category is presented, along with a technique for measuring whether two points of view generate different graded structures.

The ability to take another person's point of view plays a central role in many cognitive processes. Successful communication, for example, requires that a speaker take the hearer's point of view when constructing an utterance and that the hearer take the speaker's point of view when interpreting it. As Clark and Carlson (1981, 1982) suggest, utterances are constructed and interpreted in the context of a speaker's and hearer's beliefs about each other's knowledge, that is, in the context of their mutual knowledge (Clark & Marshall, 1981). More recently, Wilks and Bien (1983) have proposed a computational model of how people construct each other's point of view to establish mutual knowledge during language processing.

Taking a point of view can also be central to remembering, playing important roles in both encoding and retrieval. With respect to encoding, Owens, Dafoe, and Bower (1977) showed that the character whose point of view people adopt when reading a story determines the inferences they make about the characters and plot. Adopting a point of view can also be important to retrieval. After having subjects encode information from one point of view, Anderson and Prichert (1978) showed that adopting a different point of view at retrieval enabled subjects to recall information they had not been able to remember when recalling from the original point of view. Points of view serve as knowledge structures for elaborating upon presented materials and as retrieval devices capable of probing related information acquired at encoding. In addition, Nigro and Neisser (1983) have shown that people's memories vary in terms of whether they are viewed from the point of view of actually performing events (field memories) or from the point of view of an outside observer watching events (observer memories).

Taking a point of view can also play a central role in social decision making. Birnbaum and Stegner (1979) found that when judges assessed the



value of a used car, their estimates depended on whether they took the buyer's or seller's point of view. With regard to ethical decisions, McGeorge (1975) and Yussen (1976) provide evidence that people can adopt another person's moral point of view (cf. Bloom, 1976, 1977; Rest, 1979). For example, people have knowledge of how the ethics of a philosopher differ from the ethics of a police officer and can therefore make ethical decisions from the point of view of either.

In each of these processes--language comprehension, remembering, and social decision making--adopting a point of view appears to play the important role of constructing representations relevant to the respective task. In this paper, we focus on how taking a point of view may affect the representations for one particular kind of knowledge, namely, knowledge of categories. As we discuss shortly, complex cognitive tasks often require the ability to construct specialized category representations relevant to particular points of view.

An avalanche of recent work has shown that category organization takes the form of graded structure. Graded structure (also known as typicality and exemplar goodness) reflects how representative exemplars are of their category. The most highly representative exemplars are prototypical (e.g., *chairs* are prototypical of *furniture*); others are representative, but not prototypical (e.g., *lamp*); and some, whose category membership may be uncertain, are unrepresentative of their category (e.g., *refrigerator*). Graded structure within a category refers to the gradient of representativeness that exists over category members. Graded structure also extends beyond a category's boundary, some non-members being more representative of all non-members than others (e.g., *robin* is a better non-member of *furniture* than is *wall*).

So far, graded structure has been found in a wide variety of categories, and no natural categories without such structure have come to light (cf. Armstrong, Gleitman, & Gleitman, 1983).[1] Mervis and Rosch (1981), Smith and Medin (1981), and Medin and Smith (in press) review work that demonstrates the presence of graded structure in a wide variety of categories and that demonstrates the ability of graded structure to predict performance on fundamental categorization tasks. Better than any other variable, graded structure predicts the order in which a category's exemplars are acquired, the likelihood of generating particular exemplars during production tasks, and the time to verify exemplars as category members.

As discussed earlier, adopting a point of view appears to construct specialized representations during language comprehension, retention, and social decision making. What effect, however, does adopting a point of view have on a category's graded structure? Since categories often enter into comprehension, retention, and social decision making, their representation may well vary across points of view. For example, if a hippie uttered, "I cooked myself a great dinner last night," a listener might construct a representation of _things to cook for dinner_ from the hippie's point of view in which vegetarian dishes were typical and meat dishes were atypical. If the speaker were a redneck, however, the listener might construct a different representation of the category in which meat dishes were typical and vegetarian dishes were atypical. Notably, if the listener fails to construct a representation of the category appropriate to the speaker's point of view, less than optimal inferences may occur.

Nevertheless, it remains to be seen what effect point of view has on category structure, and the goal of the following experiments is to see what this effect is. In these experiments, subjects are asked to adopt a point of view prior to ranking category exemplars by typicality. For example, a subject might be asked to rank exemplars of _birds_ from the point of view of



the average Chinese citizen or from the point of view of the average American housewife. There are several possible outcomes.

One possibility, the _random shift hypothesis_, is that people will produce different graded structures from different points of view, but will do so in a haphazard fashion. Under such conditions, people may exhibit idiosyncratic conceptions about points of view or simply guess such that they shift graded structures, but show no agreement in doing so.

A second possible outcome, the _no shift hypothesis_, is that a category's graded structure remains invariant across all the points of view a person might adopt. According to this hypothesis, _robin_ should remain typical and _ostrich_ should remain atypical regardless of whether someone is processing _birds_ from the point of view of the average American citizen or the average Chinese citizen.

No shift in graded structure is predicted by the view that graded structure reflects the correlational structure of the environment. Rosch and Mervis (1975, p. 575), for example, argued that "... categories tend to become organized in such a way that they reflect the correlational structure of the environment in a manner which renders them maximally discriminable from each other." Since graded structure is the organization Rosch and Mervis are referring to, they are arguing that graded structure reflects the correlational structure of the environment. It follows from this position that each category should have only one graded structure. Assuming there is only one correlational structure in the physical world underlying a category, there should be only one graded structure reflecting it. And since the correlational structure of the environment remains invariant across the points of view someone could adopt, this theory predicts graded structure should remain invariant. Although processing assumptions could be added to this theory to account for shifts in graded structure, the theory in its current form is unable to do so.

Theories of semantic memory similarly assume that a given category has a single graded structure. In the Glass and Holyoak (1975) model, an exemplar's typicality reflects how well it is associated to its category concept. In the Smith, Shoben, and Rips (1974) model, an exemplar's typicality reflects how many properties it shares with the category concept. And in the Collins and Loftus (1975) model, typicality is a function of both associative strength and property overlap. Although these theories could be augmented to generate different graded structures from different points of view, none of them in its current form is designed to do so.

The third possible outcome, the _systematic shift hypothesis_, is that people will produce different graded structures for a category from different points of view and will show agreement in doing so. Many accounts are capable of generating systematic shifts. The one we will pursue, however, makes the following four assumptions.

First, adopting a point of view activates a stereotype for the culture, subculture, or individual whose point of view is being taken. Stereotypes provide information about individuals' values, habits, experience, beliefs, and so on that can be used to take their point of view.[2] Second, the representation of a category on a specific occasion contains only a small fraction of the knowledge comprising the category in a person's memory. Although a category's knowledge base may contain a tremendous amount of knowledge, only a small amount may typically be used in a specific situation. Third, the stereotype adopted for the current point of view determines which information in a category's knowledge base becomes incorporated into the category's representation on that occasion. That is, only information in a category's knowledge base relevant to the current point of view will be incorporated into the current representation. Fourth,



graded structure results from computing the similarity of currently relevant exemplars to the current representation of the category. Exemplars highly similar to this representation are typical, whereas less similar exemplars are less typical. Since the information in the category's representation changes with point of view, the standard against which exemplars are judged changes. This change in the standard causes changes in exemplars' similarity to the standard and consequently in how typical they are of the category.

As an example, consider processing the category of _food_ from the point of view of an anorexic. To do this, a person may activate a stereotype for anorexics, which may state that anorexics are compulsive dieters with the goal of being as thin as possible. Although the knowledge base for _food_ may contain a tremendous amount of knowledge, the anorexic stereotype may selectively activate information relevant to losing weight that is incorporated into the category's representation on this occasion. For example, the anorexic stereotype may selectively activate properties relevant to dieting, such as how many calories a food has, how filling it is, and so on. Once a representation for the category concept has been constructed, graded structure results from computing the similarity of currently relevant exemplars to this representation--as exemplars become increasingly low in calories and increasingly less filling, they become increasingly typical.

Notably, graded structure from the anorexic point of view may differ from graded structures constructed from other points of view. Drewnowski (1983) has shown that people's preferences for _foods_ depend on their body weight. Average weight people prefer foods that are _natural_ and _nutritious_, whereas overweight people prefer foods that are high in _flavor_. Consequently, when people adopt the point of view of the average weight person, typicality may reflect nutritiousness. But when adopting the point of view of overweight people, typicality may instead reflect flavor. By judging exemplars against different information selectively activated by a stereotype, a person could construct multiple graded structures for the same category. A more detailed framework incorporating these four assumptions will be presented in the general discussion.

An issue related to the systematic shift hypothesis is whether or not these shifts are valid. When taking a point of view, do subjects accurately shift to the graded structure true of the people whose point of view they are taking? When Americans adopt the Chinese point of view, for example, are their graded structures the same as those generated by Chinese people themselves? Accordingly, the systematic shift hypothesis has two subhypotheses--the valid systematic shift hypothesis and the invalid systematic shift hypothesis. Since accurately taking points of view is important to successful communication and interaction, systematic shifts may typically be valid. Nevertheless, people often seem to do poorly at taking points of view and report difficulty in doing so. Consequently, systematic shifts may typically be invalid.

Shift and Category Type

Whether or not people shift graded structure when adopting a point of view may depend on the type of category they are processing. Barsalou (1981, 1983, 1984) distinguishes between common taxonomic categories and goal-derived categories. Common taxonomic categories (e.g., fruit, birds, furniture) reflect the correlational structure of the environment, circumscribing sets of entities that share correlated properties. Because the members of a common taxonomic category are typically very similar to one another and very different from members of other categories, they stand out as a salient group of entities in the environment (cf. Rosch & Mervis, 1975;



Rosch et al., 1976).

In contrast, goal-derived categories violate the correlational structure of the environment and therefore are not naturally salient. Instead these categories become salient because perceiving them is necessary to achieving a goal. Consider things to take on a camping trip. Members of this category are very diverse, exhibiting low within category similarity (e.g., can opener, jacket, and shovel); in addition, category members bear high similarity to non-members of the category (e.g., toaster oven, suit, and post-hole digger, respectively). This set of objects only forms a salient category because it is necessary to successfully achieving the goal of going camping.

Because the salience of common taxonomic categories originates in the correlational structure of the environment, they may exhibit little or no shift with changes in point of view. A person might believe that all people, regardless of their cultural perspective, perceive the same structure in the environment and therefore structure these categories in the same way. In contrast, goal-derived categories may be much more likely to shift with point of view. Since these categories serve goals, and since goals vary widely with culture, people may believe that the structure of these categories should vary with point of view. Consequently, the structures of goal-derived categories may shift with point of view while the structures of common taxonomic categories remain fixed.

Alternatively, point of view may have the same effect on both category types. Regardless of whether there is no shift, random shift, or systematic shift, the same type of shift may occur equally for both common taxonomic and goal-derived categories. This result would suggest that the processing of both category types is fundamentally the same.

Testing the Shift Hypotheses

Subjects in the experiments to follow ranked the exemplars of various categories for typicality from one of several points of view. Two statistics computed from these rankings--subject reliability and shift--bear on the shift hypotheses. Each of these is briefly discussed in turn. A more complete specification of these measures is presented in the appendix.

Subject Reliability. Of interest is the extent to which subjects taking the same point of view agree in their graded structures for a given category. As defined in the appendix, such agreement will be assessed by a transformation of Kendall's coefficient of concordance. This statistic estimates how much subjects who rank typicality for a given category from the same point of view agree on the average. It estimates the average correlation between all possible pairs of subjects judging typicality from the same point of view. When subjects show little or no agreement, subject reliability approaches zero; as subjects become increasingly unanimous, subject reliability approaches one.

Notably, the magnitude of subject reliability is not affected by the number of subjects judging typicality. If subject reliability were computed for a sample of 18 subjects, the expected value would be the same as for a sample of 72 subjects (although the error of estimate would be larger for the smaller sample). Unfortunately, most previous reports of agreement for typicality have used measures that are biased by sample size. For example, Armstrong, Gleitman, and Gleitman (1983) and Rosch (1973, 1975) reported "agreement" scores of .90 and higher. Yet, the measure they used, split-half correlation, increases monotonically with sample size. By running enough subjects, agreement scores of .90 and higher can be obtained, even though the average agreement between pairs of subjects is small. For example, when the average correlation between pairs of subjects is only .20, a split-half correlation of .69 will be obtained when a total of 18 subjects



are run (i.e., 9 subjects in each "half"); when the number of subjects is quadrupled to 72, the split-half correlation increases to .90 (these values were obtained using Equations 5 and 6 in the appendix). When non-biased measures of agreement are used, the average agreement for typicality is generally found to range from .30 to .50 (e.g., Barsalou, 1983; Galambos & Rips, 1982).

According to the random shift hypothesis, subjects should show no subject reliability since they have idiosyncratic knowledge about points of view or are simply guessing. According to the no shift hypothesis, subjects should show significant subject reliability since their graded structures reflect the correlational structure of the environment, which they all perceive similarly. Subjects should also show significant subject reliability according to the systematic shift hypothesis, since they all shift their graded structures in a similar manner. (The predicted difference between the latter two hypotheses is for the measure of shift, described in the next section.)

Regarding category type, Barsalou (1983) found equal subject reliability for common taxonomic and goal-derived categories, so a similar result may occur here. With regard to point of view, subject reliability may drop when subjects take relatively unfamiliar points of view. For example, subjects may agree less when judging typicality from the point of view of the average Chinese citizen than from the point of view of the average American citizen.

Shift. This measure reflects the extent to which subjects taking the same point of view differ as a whole in comparison to subjects taking another point of view. More specifically, it reflects the extent to which the average graded structure across subjects taking one point of view differs from the average graded structure for subjects taking another. As discussed in the appendix, one shift score is computed for the graded structures generated by two points of view for one category. To the extent two points of view generate the same graded structure for the category, the respective shift score should not differ significantly from zero. To the extent two points of view generate different graded structures, shift should be significantly larger than zero. The possible range of values for this ratio scale measure will be presented in the context of individual experiments, but the largest possible values that could be observed here will be around three. Since this measure may be of interest in other research settings, its complete derivation is presented in the appendix.

According to the random shift hypothesis, shift scores should generally not exceed zero. To the extent subjects adopting two different points of view randomly construct graded structures, the average graded structures for the two groups should be about the same. That is, if there is no systematic difference between subjects taking two points of view, then there should be no difference in their average graded structures.

The no shift hypothesis also predicts that shift scores should generally not exceed zero. Since subjects always use correlational structure (according to Rosch & Mervis, 1975) or the same memory representation (according to current semantic memory models) to construct graded structure, changing points of view should not result in average graded structures that differ across points of view.

In contrast, the systematic shift hypothesis predicts that shift scores should generally exceed zero. To the extent different points of view systematically result in different versions of the same category, the average graded structures for different points of view should vary significantly.

The final predictions for shift concern category type. If correlational



structure plays an important role in the graded structure of common taxonomic categories, then significant shift may only occur for the goal-derived categories. But if the processing of both category types is fundamentally the same, then whatever shift occurs may occur equally for both.

The experiments to follow explore, first, whether subjects exhibit significant subject reliability when processing categories from various points of view, and second, whether the average graded structure of a category changes with the point of view from which it is processed. Experiment 1a observes graded structure when categories are processed from various international points of view, experiment 1b observes graded structure when categories are processed from various domestic points of view, and Experiment 1c observes graded structure when categories are processed from people's own points of view. Since the data from Experiment 1 strongly support the systematic shift hypothesis, Experiment 2 addresses whether such shifts are valid or invalid; that is, can one population of subjects accurately generate graded structures from another population's point of view. After samples of Emory undergraduates and faculty generated graded structures from their own points of view, additional samples of undergraduates and faculty were asked to generate graded structures from each other's point of view. Samples of graduate students also tried to generate graded structures from the points of view of undergraduates and faculty. Since the graded structures of the categories differed substantially between undergraduates and faculty, it was of interest whether a given population would accurately construct the graded structure of a different population, whether they would instead project their own graded structure, or whether they would generate an inaccurate graded structure different from their own. To the extent that adopting a point of view is necessary for successful performance on certain complex tasks, people should accurately construct graded structures for other points of view. To the extent that adopting other points of view is difficult, people should be inaccurate.

## EXPERIMENTS 1A, 1B, and 1C

Subjects in these experiments received eight exemplars from each of 40 categories and ranked each set of exemplars by typicality. In Experiment 1a, each subject took only one of four international points of view: American, French, Chinese, or African. More specifically, each subject was asked to rank the exemplars for a category as they thought the average member of their assigned point of view would rank them. To the extent people share knowledge about a point of view, they should generate similar graded structures when taking it. More specifically, such agreement should be reflected in significant subject reliability. More importantly, if subjects have different knowledge for these points of view, and if this knowledge affects graded structure, then different graded structures should emerge for the same categories under different points of view. We will assume that two points of view generate different graded structures for the same categories when their shift scores significantly exceed zero.

In Experiment 1b, each subject took one of four domestic points of view: Hippie, Businessman, Housewife, or Redneck. To the extent people share knowledge of these points of view, we should again observe significant subject reliability. To the extent these points of view generate different graded structures for the same categories, shift scores should exceed zero. We will also compare each domestic point of view to the American point of view from Experiment 1a to show how the shift measure can can be used to determine the relative similarity of different points of view (e.g., which domestic point of view is most similar to the average American point of view).



In Experiment 1c, each subject either took their own point of view or the American point of view. To the extent subjects in the Self condition take highly personalized points of view, they should show much less agreement than subjects taking a common point of view, in this case, the point of view of the avarage American. More specifically, subject reliability should be lower in the Self than in the American condition if subjects are able to take unique points of view that differ from the American point of view.

To explore the effects of point of view on category type, half of the categories in Experiments 1a, 1b, and 1c were common taxonomic categories, and half were goal-derived categories.

### Method

#### Materials

Except for the point of view that subjects were instructed to take, all materials for Experiments 1a, 1b, and 1c were the same. The 20 common taxonomic categories were selected from Battig and Montague (1969) and were: birds, clothing, colors, crimes, flowers, furniture, fruit, insects, kitchen utensils, mammals, musical instruments, occupations, precious stones, spices, sports, tools, trees, vegetables, vehicles, and weapons. The 20 goal-derived categories were ones whose graded structure we thought might change with point of view. They were: birthday presents, cold remedies, desserts, famous leaders, games, hobbies, household chores, important events in history, important goals in life, kinds of crises, kinds of music, places to donate money, places to vacation, qualities to instill in children, religions, things people worry about, things to buy at the grocery store, things to take from one's home during a fire, ways to meet someone, and weekend activities.

Forty-eight subjects were asked to generate the first four exemplars that came to mind for each category. Two groups of twenty-four subjects each received one of two random orders of the categories, which were not blocked by category type. The exemplars generated for each category were rank-ordered by the number of subjects out of 48 generating them. Eight exemplars were then selected for each category. One was the exemplar generated most often, another was an exemplar generated by only one subject, and the remaining six were, as much as possible, from equal intervals between these two. When several exemplars were tied in production frequency at a level from which an exemplar was to be chosen, exemplars were excluded that would not be familiar to most subjects or that were very similar to an exemplar already selected. Exemplars were not chosen on the basis of how much their typicality might change with point of view. Instead, a set of exemplars was selected that spanned production frequency and that was as representative of the category as possible.

Two versions of the final materials were constructed. Each version had a different random order of the 40 categories, which were not blocked by type, and each version contained a different random order of the eight exemplars for a given category. Ten categories were typed in two columns on each of four pages. For each category, the underlined category name appeared above its column of eight exemplars.

Design and Subjects

Point of view was a between subjects manipulation in all three experiments. In Experiment 1a, four different groups of twelve subjects each were asked to take the point of view of the average American, French, Chinese, or African citizen. In Experiment 1b, four different groups of twelve subjects each were asked to take the point of view of the average Businessman, Hippie, Housewife, or Redneck. In Experiment 1c, two groups of twenty-four subjects were asked to take their own point of view or the point



of view of the average American citizen. Twelve of this latter group of subjects were the twelve subjects for the American point of view in Experiment 1a. For each experiment, half the subjects taking a given point of view received one version of the materials and half received the other. Category type was manipualated within subjects.

Subjects were students participating for course credit from introductory psychology courses at the main Emory campus and the Oxford, Georgia campus of Emory University. Subjects from the two campuses were evenly distributed over the between subjects conditions of the design.

Procedure

Subjects received booklets with instructions describing how to rank a category's exemplars by typicality. In these instructions, subjects were also asked to maintain a particular point of view while performing typicality judgments. This was stressed throughout the instructions (e.g., "It is absolutely essential that you maintain the point of view of the average *X* while doing these rankings," where *X* was the point of view a subject was asked to take). In addition, "***Please remember to retain the proper point of view***" appeared at the top and middle of each page of the materials that contained categories. Two additional categories preceded the forty critical categories to provide subjects with practice at the task. Subjects were given as much time as necessary to complete the experiment.

Results: Experiment 1a

Subject Agreement

Subject reliability, as defined in the appendix, was computed for each of the 40 categories for each of the four points of view. These 160 subject reliabilities were transformed to Fisher $z$ scores and subjected to a point of view by category type by category ANOVA. The average subject reliabilities from this analysis are shown at the top of Table 1. Each is significantly greater than zero at the .001 level. When subjects take varous international points of view, they agree on the graded structures they generate.

Subjects exhibited equal agreement for common taxonomic and goal-derived categories, .330 and .356, respectively ($F<1$). This replicates Barsalou's (1983) finding that subjects agree equally well on the graded structures of common taxonomic and goal-derived categories.[3] Subject reliability varied significantly between points of view [$F(3,114)=10.05$, $p<.001$], with subjects showing more agreement for the American and Chinese points of view than for the French and African points of view. There was also a marginal interaction between category type and point of view [$F(3,114)=2.31$, $.10>p>.05$]. No obvious explanation for this pattern of differences suggests itself.

Shift

One shift score, as defined in the appendix, was computed for each pair of points of view for each category (i.e., 240 shift scores). The top of Table 2 shows the average group reliabilities, between group correlations, and shift scores for the common taxonomic and goal-derived categories in Experiment 1a. The correlations are clearly much less than the group reliabilities. Consequently, the average shift across all conditions, .612, significantly differs from zero [$t(239)=15.483$, $p<.001$]. Adopting different international points of view can clearly cause subjects to generate different graded structures for the same categories.

Surprisingly, shift was equally large for the common taxonomic and goal-derived categories [$t(238)=.747$, $p>.10$]. Changes in point of view caused equal change in the graded structures of both category types.

Given the size of the group reliabilities in this experiment, the shift scores can potentially range from 0 to around 2 on the average, with values around 1 indicating no correlation between the graded structures generated



by two points of view. Since the average shift obtained in Experiment 1a was .612, there is some but not a lot of similarity between the graded structures generated from these different points of view.

The average shift for the six pairs of international points of view are shown in Table 3 by category type. To test whether the average for a particular pair significantly exceeded zero for a particular category type, the variances of these averages were pooled, the resulting standard deviation being .583. On the basis of one-tailed $t$ tests with 19 degrees of freedom, average shift scores of .225 and .330 are significantly different from zero at the .05 and .01 levels, respectively. Except perhaps for the American-French and French-Chinese pairs for the common taxonomic categories, all pairs clearly generate different graded structures for the same categories. The most divergent points of view are the American and Chinese points of view, whose graded structures are completely unrelated (i.e., values of 1 on the shift scale for this data set indicate no correlation between graded structures). This lack of relation can not be attributed to lack of agreement among subjects within groups, since all groups have subject reliabilities significantly greater than zero (Table 1). Table 4 provides the average rankings of two categories' exemplars to illustrate the effect that the international points of view had on category structure.

## Results: Experiment 1b

Subject Agreement

The 160 subject reliabilities were transformed to Fisher $z$ scores and subjected to a point of view by category type by category ANOVA. The average subject reliabilities from this analysis are shown in Table 1. Each is significantly greater than zero at the .001 level, again indicating agreement among the graded structures generated for a given point of view.

Subjects exhibited equal agreement for common taxonomic and goal-derived categories, .414 and .401, respectively, again replicating Barsalou's (1983) finding that subjects agree equally well on the graded structures of common taxonomic and goal-derived categories ($F<1$). Subject reliability varied significantly between points of view [$F(3,114)=8.07$, $p<.001$], agreement for the Hippie point of view being lower than for the Businessman, Housewife, and Redneck points of view. There was also a marginal interaction between category type and point of view [$F(3,114)=2.32$, $.10>p>.05$]. No obvious explanation for this pattern of differences suggests itself.

Shift

Table 2 shows the average group reliabilities, between group correlations, and shift scores for the common taxonomic and goal-derived categories. As in Experiment 1a, the correlations are clearly much less than the group reliabilities. Consequently, the average shift across all conditions, .778, significantly differs from zero [$t(239)=19.450$, $p<.001$]. Adopting different domestic points of view can clearly cause subjects to generate different graded structures for the same categories. In contrast to Experiment 1a, the average shift for the common taxonomic categories was slightly larger than for the goal-derived categories [$t(238)=2.025$, $p<.025$]. Perhaps more important, however, is that different points of view cause substantial differences in graded structure for common taxonomic categories [$t(119)=11.369$, $p<.001$] as well as for goal-derived categories [$t(119)=16.955$, $p<.001$].

The average shift for the six pairs of domestic points of view are shown in Table 5. To test whether the average for a particular pair significantly exceeded zero, the variances of these averages were pooled, the resulting standard deviation being .583. On the basis of one-tailed $t$ tests with 19 degrees of freedom, average shift scores of .230 and .338 are significantly different from zero at the .05 and .01 levels, respectively. Except perhaps for the Businessman-Housewife pair, all pairs clearly generate different graded structures for the same categories. The most divergent points of view are Hippie-Businessman, Hippie-Housewife, and Housewife-Redneck, whose graded structures are close to being unrelated (as in Experiment 1a, shift scores can generally range from 0 to 2, with values around 1 indicating no correlation between graded structures). Again this absence of relation can not be attributed to lack of agreement within groups, since all have subject reliabilities significantly greater than zero (Table 1). Table 4 illustrates the effect domestic points of view had on the average rankings of two categories.



Comparisons Between the American and Domestic Points of View

Average shift scores were computed for the American point of view from Experiment 1a in conjunction with each of the domestic points of view in Experiment 1b. The average values across category types are shown in Table 6. To test whether the average for a particular pair significantly exceeded zero for a particular category type, the variances of these averages were pooled, the resulting standard deviation being .508. On the basis of one-tailed $t$ tests with 19 degrees of freedom, average shift scores of .197 and .289 are significantly greater than zero at the .05 and .01 levels, respectively. As can be seen, the American, Businessman, and Housewife points of view generated the same graded structures for the common taxonomic categories, but not for the goal-derived categories. In contrast, subjects see the Hippie and Redneck points of view as being quite different from the average American's point of view.

Results: Experiment 1c

Subject Agreement

The 80 subject reliabilities were transformed to Fisher $z$ scores and subjected to a point of view by category type by category ANOVA. The average subject reliabilities from this analysis are shown in Table 1. Each is significantly greater than zero at the .001 level.

If subjects are able to take their own point of view, then these individual points of view should not only differ from the American point of view but should also differ from each other. Consequently, subjects should show much less agreement among themselves when taking their own points of view than when taking the American point of view, for which they all are likely to have shared knowledge. Although subject agreement was significantly less for the Self than for the American point of view [$F(1,38)=5.59$, $p>.025$], subjects' individual points of view were much more similar than we had expected. Average agreement for subjects taking their own point of view was .393, which was close to the agreement for subjects taking the American point of view, .459. Again, there was no effect of category type [$F(1,38)=1.92$, $p>.10$].

Shift

Table 2 shows the average group reliabilities, between group correlations, and shift scores for the common taxonomic and goal-derived categories in Experiment 1c. If subjects in the Self condition basically all have the same point of view as the average American, then the average shift scores for the Self and American groups should not exceed zero--two groups taking the same point of view should show no difference. To test whether the average shift scores differed from zero, their variances were pooled, the resulting standard deviation being .420. Surprisingly, the Self and American points of view generated different graded structures for both the common taxonomic and goal-derived categories [$t(19)=2.936$, $p<.01$ and $t(19)=5.064$, $p<.001$, respectively]. Overall, common taxonomic and goal-derived categories did not differ [$t(38)=1.504$, $p>.10$].



One reason subjects in the Self condition may have showed unexpectedly high agreement among themselves, yet generated graded structures different from those for the American point of view, is that our subjects are from a fairly homogeneous subculture (i.e., Emory undergraduates). Given their homogeneity, they show agreement; and given that they are a subculture, their point of view differs slightly from their point of view of the average American. Table 4 illustrates the slight difference between these points of view in their average rankings for two categories.

## Discussion

In Experiments 1a and 1b, subjects taking each international and domestic point of view showed significant within-group agreement in the graded structures they generated. Such agreement indicates that subjects share common knowledge about these points of view and that this knowledge can be used to construct similar graded structures. In addition, agreement was equal for common taxonomic and goal-derived categories, replicating Barsalou's (1983) finding that graded structures are equally salient for both category types.

Although subjects taking the same point of view showed agreement, the graded structures generated by different points of view generally differed. In fact, some points of view generated graded structures that were completely uncorrelated with each other (e.g., the American and Chinese points of view; the Hippie and Businessman points of view). Given the large impact that points of view had on category organization, categories can clearly possess multiple graded structures. Categories do not appear to have a single graded structure that simply reflects the correlational structure of the environment. Instead, people appear capable of operating on the large knowledge base that comprises a category to construct category representations relevant to their current point of view.

Unexpectedly, different points of view reorganized common taxonomic and goal-derived categories equally in Experiments 1a and 1c. And although graded structures varied more for common taxonomic than for goal-derived categories in Experiment 1b, the reorganization of the common taxonomic categories was still substantial. This suggests that the correlational structure of the environment is not solely responsible for the graded structures of common taxonomic categories. If it were, the graded structure of these categories should have remained relatively stable across points of view. Changes in other kinds of information (e.g., function and frequency) may underly the reorganization of common taxonomic categories, as discussed later.

Also, subjects taking their own point of view unexpectedly showed nearly as much agreement in their graded structures as subjects taking the same point of view, in this case, the point of view of the average American. Subjects' individual points of view appear to be more similar than might be expected. In addition, the overall graded structures for these subjects significantly differed from those for the average American point of view. Subjects' membership in the subculture of Emory undergraduates may have caused their own points of view to differ slightly from the point of view of the average American.

Finally, these experiments show that our measurement technique is capable of measuring the extent to which points of view reorganize a category. For example, it is possible to determine which points of view generate unrelated graded structures (e.g., the Businessman and Hippie points of view) and to determine when two points of view generate the same graded structures (e.g., the Housewife and American points of view for common taxonomic). As we show in Experiment 2, this technique can also be used to determine whether subjects can take another point of view



accurately.

## EXPERIMENT 2

Subjects from two Emory populations, faculty and undergraduates, each took the point of view of the average member of their population while ranking category exemplars by typicality. Different subjects from these two populations then judged typicality from the point of view of the average member of the other population. In addition, subjects from another population, graduate students, judged typicality from the faculty or undergraduate points of view. To the extent subjects within a population agree when taking a point of view, there should be high subject reliability. The central question, however, is how well one population can take another population's point of view. The extent to which population A accurately takes the point of view of population B can be assessed by computing shift scores for A taking B's point of view and B taking their own point of view. To the extent A is inaccurate, shift scores should significantly exceed zero.

The categories selected for this experiment appeared to have graded structures that differed for the faculty and undergraduate points of view. Such divergence in graded structures is necessary if there is to be any potential for one population to be inaccurate while taking another population's point of view. Although being able to construct accurate graded structures from different points of view is probably central to many complex cognitive tasks, we suspected that our subjects would be far from accurate when taking a point of view that was not their own.

To confirm this intuition, we asked faculty, undergraduates, and graduate students to answer the following four questions: (1) how well can the average faculty member take the point of view of the average undergraduate; (2) how well can the average graduate student take the point of view of the average undergraduate; (3) how well can the average graduate student take the point of view of the average faculty member; (4) how well can the average undergraduate student take the point of view of the average faculty member. Thirteen subjects from each population answered these questions in one of two random orders. Subjects used a five point rating scale to answer each question, where, with regard to one population taking the point of view of another, 1 meant not at all, 2 meant a little, 3 meant fairly well, 4 meant a lot, and 5 meant perfectly.

Our intuitions were confirmed in that the average rating was quite low (i.e., 2.718), being less than the midpoint on the scale. Members of the Emory community clearly did not believe that these various populations could do a very good job of taking each other's point of view. The only effect in a population by question ANOVA was for question [$F(3,108)=5.84$, $p=.001$]: All subjects agreed that the best performance would be observed for undergraduates taking the faculty point of view (see Table 7).

## Method

### Materials

Twenty goal-derived categories were selected that appeared to have one graded structure characteristic of undergraduates and a different one characteristic of faculty. The categories were alcoholic beverages, athletic activities, birthday presents, cars, famous people admired, important goals in life, luxury items, movies, people to get advice from, personality characteristics, places to vacation, things done at school, things people worry about, things to do at parties, things to take on a picnic, types of books, types of food establishments, types of music, TV programs, and weekend activities.

Two initial lists of the 20 categories were constructed, one designed for faculty and the other for undergraduates. The lists differed in that



undergraduates received categories such as _types of music undergraduates listen to_, whereas faculty received categories such as _types of music faculty listen to_. Twenty-four subjects from each population were asked to generate the first four exemplars that came to mind for each category. Approximately half the subjects in each population received one random order of the list, and half received another.

For each population, the exemplars generated for each category were rank ordered by the number of subjects out of 24 generating them. Eight exemplars were then selected for each category, four from each population. The composition of each category's eight exemplars included: the exemplar generated most often by each population, an exemplar generated by only one subject for each population, and two exemplars that were equally spaced between the others for each population. These intermediate exemplars also met the criteria of being familiar to all subjects and not being highly similar to any other exemplar selected for the respective category.

Two versions of the final form were constructed. In each, the category names were changed back from the population specific versions to the more generic versions stated in the first paragraph of this section. Each version contained different random orders of the categories and different random orders of the exemplars for the same category. Ten categories were typed in two columns on each of two pages. For each category, the underlined name appeared above its column of eight exemplars.

### Design and Subjects

Of the subjects performing typicality judgments, 48 were Emory undergraduates participating for course credit, 48 were Emory graduate students, and 48 were Emory faculty. Within the sample of 48 subjects from each population, 24 took the point of view of the average Emory faculty member, and 24 took the point of view of the average Emory undergraduate.

Approximately half the subjects taking each point of view received one version of the materials, and half received the other.

There were two mass mailings of questionnaires to faculty and one mass mailing to graduate students. One mailing to faculty was to collect instances for the category norms and the other was to collect typicality judgments. Faculty and graduate students were randomly selected from university listings, and no individual was allowed to participate in more than one phase of the experiment. The response rate to these mailings was approximately 30 percent, which falls into the standard 25 to 40 percent range noted by Sudman and Bradburn (1982). When more questionnaires for a point of view were returned than were needed to fill the design, subjects were randomly excluded.[4]

Procedure

The procedure for Experiment 2 was the same as for Experiment 1, except that a cover letter was included with the questionnaires mailed to faculty and graduate students to insure participants that their data would remain anonymous.

## Results

Subject Agreement

The 120 subject reliabilities were transformed to Fisher $z$ scores and submitted to a population by point of view by category ANOVA. The average subject reliabilities from this analysis are shown in Table 8. Each is significantly greater than zero at the .001 level, indicating that subjects in each condition showed agreement in the graded structures they generated. There was a main effect of population [$F(2,38)=5.14$, $p<.01$], faculty showing more agreement than the other two populations. In addition, population and point of view interacted [$F(2,38)=12.83$, $p<.001$]. This resulted from faculty and graduate students showing more agreement for the faculty than



for the undergraduate point of view and the reverse being true for undergraduates. Faculty and undergraduates showed more agreement when they took their own point of view than when they took a different point of view.

Shift

Table 9 shows the average group reliabilities, between group correlations, and shift scores for pairs of groups taking different points of view (i.e., one group took the faculty point of view and the other took the undergraduate point of view) and for pairs of groups taking the same point of view (i.e., both groups took either the faculty or the undergraduate point of view). As can be seen, groups taking different points of view generated very different graded structures for the same categories, the average shift for these pairs being significantly greater than zero [$t(179)=27.824$, $p<.001$]. Given the size of the group reliabilities in this experiment, the shift score can potentially range from 0 to around 3 on the average, with values around 1.5 indicating no correlation between the graded structures generated by two points of view. Since the average shift for groups taking different points of view was 1.270 there was very little relation between the graded structures generated by two groups taking different points of view.

Quite surprisingly, subjects taking the same point of view generated the same graded structures, the average shift for these pairs (.043) not differing significantly from zero [$t(119)=1.153$, $p>.10$]. Groups taking the same point of view appeared able to do so with remarkable accuracy. In addition, groups taking the same point of view generated graded structures that differed significantly less than did the graded structures generated by groups taking different points of view [$t(298)= 20.815$, $p<.001$].

The average shift for the fifteen pairs of groups are shown in Table 10. To test whether the average for a particular pair significantly exceeded zero, the variances of the averages were pooled, the resulting standard deviation being .541. On the basis of one-tailed $t$ tests with 19 degrees of freedom, average shift scores of .209 and .307 are significantly different from zero at the .05 and .01 levels, respectively.

Each pair of groups taking different points of view clearly generated graded structures for the same categories that differed significantly. Four of these pairs are of particular interest. First, the average difference score for faculty and undergraduates each taking their own point of view (i.e., $F_SFp$ versus $U_SUp$ in Table 10) represents how much the points of view truly differ from each other. Given that the two points of view differ substantially, there was much room for error when faculty and undergraduates took each other's point of view. The other three pairs of interest are those in which the same population takes the two different points of view (i.e., $F_SFp$ versus $F_SUp$, $G_SFp$ versus $G_SUp$, and $U_SFp$ versus $U_SUp$). In all cases, a population perceives as large a difference between the two points of view as actually exists. As shown next, these perceived differences reflect the real difference represented by the $F_SFp$-$U_SUp$ comparison.

Turning to pairs in which two groups took the same point of view, only one pair generated graded structures that differed significantly. When faculty tried to take the undergraduate point of view ($F_SUp$ versus $U_SUp$), they were significantly inaccurate, but the degree of error was not very large (i.e., the average shift score was only .356). For the other five pairs in which groups took the same points of view, there was perfect agreement: undergraduates accurately took the faculty point of view ($U_SFp$ versus $F_SFp$); graduate students accurately took the faculty point of view ($G_SFp$ versus $F_SFp$) and the undergraduate point of view ($G_SUp$ versus $U_SUp$); faculty and graduate students generated the same graded structures when they took the undergraduate point of view ($F_SUp$ versus $G_SUp$); and undergraduate



and graduate students generated the same graded structures when they took the faculty point of view ($U_SFp$ and $G_SFp$ ). Notably, when two groups take a third point of view, they can do so accurately.

Discussion

Since subjects showed significant agreement when generating graded structure, they clearly share common knowledge about the faculty and undergraduate points of view. In addition, the graded structures generated by faculty taking their own point of view and by undergraduates taking their own point of view were very different, supporting our intuition that the graded structures for these categories differed. More importantly, this indicates that there was much potential for error when a population took a point of view other than their own.

Surprisingly, subjects were extremely accurate in taking other points of view. All groups taking a different point of view did so perfectly, except for faculty taking the undergraduate point of view, who were not all that inaccurate. Not only are these results contrary to our original expectations, they are contrary to those of the subjects in the poll we discussed earlier. Even though we selected categories whose graded structures differed for the two points of view, the populations we observed were, except for one group, perfect at taking each other's point of view.

Two other findings deserve mention. First, subjects within a given population perceived the same difference in graded structures between two points of view as actually existed. Graduate students, for example, generated graded structures for the faculty and undergraduate points of view that were as different as those generated by faculty and undergraduates themselves. Given that graduate students were perfectly accurate at taking these two points of view, the difference they perceived between them was accurate, as well as being of the correct magnitude. Second, two populations can adopt a point of view that belongs to a third population and generate identical graded structures that are accurate. For example, graduates and undergraduates adopting the faculty point of view generated accurate and identical graded structures.

## GENERAL DISCUSSION

Regardless of the point of view adopted, subjects taking the same point of view always exhibited significant agreement in their graded structures. Although the graded structures for some points of view showed more agreement than for others, agreement in most cases fell within the .30 and .50 range characteristic of typicality judgments. Agreement did not necessarily decrease as points of view became more exotic. Agreement for the Redneck point of view was higher than the American point of view for both common taxonomic and goal-derived categories; and the Chinese point of view showed more agreement than the American point of view for goal-derived categories. Subjects share mutual knowledge that allows them to construct similar graded structures when taking points of view.

Although subjects showed agreement when taking the same point of view, they generated different graded structures when taking different points of view. Subjects adopting various international and domestic points of view generated highly divergent graded structures for the same categories. In some cases, the graded structures generated by two points of view were unrelated on the average across categories (e.g., the American and Chinese points of view; the Hippie and Businessman points of view).

As these findings demonstrate, people are quite facile at manipulating category organization, and a given category clearly does not have a single graded structure. Multiple graded structures raise problems for the view that graded structure reflects the correlational structure of the environment. Assuming the environment has only a single correlational



structure, this position predicts that a category should have only one typicality gradient, namely the one reflecting the environment's correlational structure. Another problem for this position is that graded structure varied as much for common taxonomic categories as it did for goal-derived categories in Experiments 1a and 1c and almost as much in Experiment 1b. If the structure of common taxonomic categories simply reflects the structure of the environment, their graded structures should not show such variability.

People can also be quite accurate in generating graded structures from an adopted point of view. Even though members of the Emory community believed that faculty, graduate students, and undergraduates would do a poor job of taking each other's point of view, these populations performed surprisingly well. Except for faculty taking the undergraduate point of view, subjects were able to reconstruct each other's graded structures perfectly, and the faculty were not off by much. Subjects exhibited high accuracy even though the graded structures generated by faculty and undergraduates taking their own points of view were nearly unrelated. This ability to accurately construct points of view may play a central role in language comprehension and other forms of social interaction.

## ACCOUNTING FOR MULTIPLE GRADED STRUCTURES

In an independent project conducted somewhat simultaneously with this one, Roth and Shoben (1983) also demonstrated that categories have multiple graded structures. Instead of manipulating the point of view from which a category was processed, however, Roth and Shoben varied a category's linguistic context. The category of <u>animals</u>, for example, had different graded structures when it appeared in "Stacy volunteered to milk the <u>animal</u> whenever she visited the farm" and "Fran pleaded with her father to let her ride the <u>animal</u>." Whereas <u>cow</u> and <u>goat</u> are typical in the first context, horse and mule are typical in the second. Not only did Roth and Shoben find that subjects' judgments of typicality changed with context, they found that comprehension and category verification times changed as well. In general, an exemplar's typicality with respect to a context sentence predicted performance, whereas an exemplar's context-independent typicality did not.

Roth and Shoben compared two hypotheses of how linguistic context could alter a category's graded structure: the refocusing hypothesis and the restructuring hypothesis. In presenting these, Roth and Shoben depict a category as a multi-dimensional space in which the similarity of one exemplar to another is a monotonically decreasing function of the distance between them. According to the refocusing hypothesis, a linguistic context determines the point in a category space that represents the category's prototype. The category's graded structure simply results from exemplars in this space varying in their closeness to the prototypical point. Close exemplars are typical, and distant ones are atypical. Multiple graded structures occur because different contexts focus on different prototypes, thereby changing exemplars' closeness to the current prototype. In Experiment 2, however, Roth and Shoben observed graded structures in linguistic contexts that failed to support this hypothesis. Their alternative explanation, the restructuring hypothesis, was more consistent with their findings. Similar to the refocusing hypothesis, the restructuring hypothesis assumes that different linguistic contexts focus on different prototypes; but in addition, this hypothesis assumes that distances in the space between exemplars change. Roth and Shoben suggest that the configuration of exemplars in a space changes because "constraints imposed by the context" alter the distances between exemplars (p. 352).

We next present a specific account of how points of view may impose constraints on a category's structure. Similar to Roth and Shoben, we



assume that points of view restructure the space representing a category. In this sense, point of view and linguistic context can both be seen as affecting graded structure in the same way.

As summarized earlier, our account first assumes that when people adopt a point of view, they activate a stereotype for the culture, subculture, or individual whose point of view is being taken. During the processing of a category, this stereotype selectively activates information from the category's knowledge base to represent the category as a whole (i.e., the category concept) and to represent exemplars. Finally, a process that compares the similarity of these category and exemplar concepts generates the category's graded structure from that point of view (cf. Barsalou, 1983; McCloskey & Glucksberg, 1979; Smith, Shoben, & Rips, 1974). Since different stereotypes cause different information to enter into these comparisons, the similarity of exemplars to one another and to the category concept changes, thereby restructuring the category.

Knowledge of Points of View

One simple way to adopt a point of view is to activate a stereotype for the person or persons whose point of view is being taken. Although stereotypes are often inaccurate, much research has shown that they contain information people use to perceive and interact with one another (e.g., Hamilton, 1979; Taylor, 1980; Taylor & Crocker, 1981). Since stereotypes contain information about individuals' values, attitudes, habits, occupations, social style, and so on, they also provide the kind of information necessary to adopting points of view. Once a stereotype has been adopted for a point of view, it may serve as a context to selectively activate related information in categories that determines their graded structure.

Knowledge of Categories

As suggested earlier, the representation of a category on a specific occasion is probably only a subset of all the knowledge comprising the category. A category's knowledge base may contain knowledge relevant to many points of view, and knowledge relevant to a particular point of view may only become active when the category is processed from that point of view. Barsalou (1981, 1984) found that three kinds of information in a category's knowledge base--objective, strategic, and frequency information--can determine its graded structure. Generalizing from these findings, a category may exhibit different graded structures from different points of view because the objective, strategic, and frequency information active for each point of view differ. To consider more specifically how this might work, consider Barsalou's (1984) model of typicality judgments. According to this model, the typicality T of exemplar $e$ in category $c$ from point of view $p$ can be accounted for as follows:

$$T(e,c,p) = G(v_1\ obj,\ v_2\ strat,\ v_3\ freq) \quad (1).$$

The function G primarily depends on *obj*, *strat*, and *freq*, which are the outputs of three other functions, O, S, and F defined later. *obj* represents the contribution of objective information, *strat* represents the contribution of strategic information, and *freq* represents the contribution of frequency information. Each kind of information has a weight--$v_1$, $v_2$, or $v_3$--which specifies the importance of that information for typicality in category $c$:

*Objective Information*. Standard accounts of prototype theory (Smith and Medin, 1981) state that the prototype of a category represents central tendencies of the category's exemplars. This central tendency information can include average values on dimensions that structure the category (e.g., *size* for *birds*) and properties or correlations of properties that typically occur for category members (e.g., the property of *wheels* for *cars*; the correlated properties of *has feathers* and *lays eggs* for *birds*). Our use of objective properties here refers to this type of central tendency information about a category's exemplars.



As noted by Barsalou (1983, 1984), computing the similarity between an exemplar and a category concept with respect to objective information is equivalent to computing the exemplar's family resemblance (cf. Rosch & Mervis, 1975). Although family resemblance is the primary determinant of graded structure in common taxonomic categories, it plays no role in the structure of goal-derived categories (Barsalou, 1984).

In the model of typicality judgments stated in Equation 1, the value of *obj* represents the contribution of objective information to typicality, where *obj* is obtained from the function O as follows:

For all objective properties active for point of view $p$,

$$obj = O[w_i Sim(o_{ei}, o_{ci})] \quad (2).$$

More specifically, this contribution depends on the similarity of each objective property in the category concept, $o_{ci}$, to its related property in the exemplar, $o_{ei}$, where similarity is some decreasing function of the difference between the two properties. For example, if the average height of *dogs* were two feet, a six inch dog would have a less similar value on the height dimension than a one foot dog. The importance of each particular similarity depends on its weight, $w_i$, and the overall contribution of objective information to typicality is some function, O, of the weighted similarities across all objective properties. Since this function is defined over all objective properties active for the current point of view, distinctive as well as shared properties (cf. Tversky, 1977) enter the computation. That is, when one concept has an objective property not shared by the other, the resulting similarity comparison returns a value of zero, which lowers the overall value of the function.

Finally, the function O could be multiplicative if the individual comparisons are combined interactively, or it could be additive if the comparisons are independent (cf. Medin & Schwanenflugel, 1981). Similar flexibility exists for the functions G and S and enables the model to account for different strategies subjects may adopt on different occasions (cf. Smith & Medin, 1981).

*Strategic Information*. This kind of information specifies properties of a category's exemplars that would be maximally expedient to achieving a goal the category was constructed to serve. Regarding the category of *things to eat on a diet*, for example, *zero calories* is a property of exemplars that would be maximally expedient to achieving the goal of *losing weight*. Similarly, the property of *zero weight* is maximally expedient for *things to take backpacking* in achieving the goal of *minimizing fatigue*. Notably, a strategic property is generally *not* the central tendency of a category's exemplars. *Zero calories* is certainly not the central tendency of *things to eat on a diet*, and *zero weight* is certainly not the central tendency of *things to take backpacking*. Instead, strategic information represents ideal properties exemplars should have to maximally achieve a goal that the category serves. Barsalou (1984) found that strategic information played an important role in the graded structures of both common taxonomic and goal-derived categories.

In the model of typicality judgments stated in Equation 1, the value of *strat* represents the contribution of strategic information to typicality, where *strat* is obtained from the function S as follows:

For all strategic properties active for point of view *p*,

$$strat = S[w_j Sim(s_{e_j}, s_{c_j})] \quad (3).$$

More specifically, this contribution depends on the similarity of each strategic property in the category concept, $s_{c_j}$, to its related property in



the exemplar concept, $s_{e_j}$. If the strategic property of *zero calories* exists in the category concept for *things to eat on a diet*, for example, then *apple* will have a higher similarity on this dimension than will *eclair*. Again, the importance of each particular similarity depends on its weight, $w_j$, and the overall contribution of strategic information is some function of the weighted similarities across all strategic properties. As for O, both shared and distinctive properties are in the domain of the function.

*Frequency of Instantiation*. Barsalou (1984) also found that subjects' estimates of how often exemplars occur as category members predicted typicality in both common taxonomic and goal-derived categories. In the category of *birthday presents*, for example, frequent exemplars such as *clothing* and *records* were more typical than less frequent exemplars such as *tools* and *paintings*. In the model stated in Equation 1, the value of *freq* represents the contribution of frequency information to typicality, where *freq* is obtained from the function F as follows:

$$freq = F(e,c,p) \quad (4).$$

More specifically, this contribution depends on an exemplar *e*'s perceived frequency of instantiation in category *c* from point of view *p*. As discussed later, these values are indexed by the stereotype responsible for the current point of view.

*The Selective Activation of a Category's Knowledge Base by a Point of View*

As stated in the functions for objective, strategic, and frequency information, a particular typicality judgment is defined with respect to a particular context, namely, point of view *p*. This results in different objective properties, strategic properties, and frequency information being incorporated into category and exemplar concepts when a category is processed from different points of view (although context-independent properties may be incorporated into a concept for all points of view;

Barsalou, 1982).

<u>Constructing Category Concepts</u>. The stereotype active for the current point of view may determine the strategic and objective information that becomes incorporated into a category concept. Regarding strategic information, values and attitudes in the stereotype may migrate in some form from the stereotype into the category concept. For example, when processing the category of <u>things to take from one's home during a fire</u> from a nurse's point of view, the property of <u>values maintaining bodily functions</u> may migrate in some form from the nurse stereotype into the category concept, where it becomes a strategic property of the category (i.e., good examples of the category will have the property of <u>useful for maintaining bodily functions</u>). In contrast, when processing this category from the businessman point of view, the property of <u>values things worth a lot of money</u> may migrate from the businessman stereotype into the category concept in some form (i.e., good examples will have the property <u>worth a lot of money</u>).

Since the nurse and businessman points of view cause different strategic properties to be incorporated into the category concept for <u>things to take from one's home during a fire</u>, its similarity to exemplars changes. This in turn causes exemplars' typicality to change. For example, exemplars that are <u>useful for maintaining bodily functions</u> will be similar to the category concept from the nurse point of view and therefore be typical. In contrast, exemplars that are <u>worth a lot of money</u> will be similar to the category concept from the businessman point of view and therefore be typical. Notably, exemplars similar to one category concept may not be similar to the other, thereby causing a shift in graded structure.

Turning to objective information, stereotypes may contain central tendency information about categories that reflects the point of view of the respective individuals. Such knowledge may migrate in some form from



stereotypes to category concepts during processing. For example, Americans believe that Africans are exposed to larger and wilder animals than are Americans. Consequently, when Americans adopt the African point of view, the objective values for size and wildness incorporated from the African stereotype into the category concept for <u>animals</u> may be higher than when Americans take their own point of view. Larger and wilder animals would therefore be more typical of animals from the African point of view. To the extent an exemplar has properties that are highly similar to objective properties currently in the category concept, it is typical of the category.

<u>Constructing Concepts for Exemplars</u>. The representation of an exemplar on a particular occasion may depend on both the current category concept and on the stereotype for the current point of view. Properties of the exemplar that are related to objective and strategic properties in the current category concept may become active in the concept for the exemplar. Activating related information for an exemplar enables comparing it to the category concept on all aspects relevant in the current context. For example, if the anorexic point of view caused <u>low calories</u> to be incorporated into the category concept of <u>food</u>, this property could activate the number of calories for each exemplar subsequently considered. Consequently, the representations of particular exemplars would possess a property relevant to the category's use on that occasion.

In addition, the current stereotype may access each exemplar's frequency of instantiation from the respective point of view. In some cases, this frequency information may reflect the actual number of exposures a person has had to an exemplar from that point of view. For example, someone who has eaten at restaurants with his or her spouse on numerous occasions may have frequency information from the spouse's point of view for the category of <u>things ordered at restaurants</u>. More often, however, frequency information is probably obtained indirectly. People may have had little if any direct experience of a point of view, thus precluding any basis for directly encoding frequency information. Nevertheless, people appear to have extensive frequency information for these points of view, much of which is probably obtained through hearsay. For example, most Americans have never visited Africa and so have had little chance to encode the frequency of various animals from the African point of view. Yet most Americans have been exposed to such frequency information through various media (e.g., books, magazines, films, and television). Consequently, information such as lions occur frequently in Africa and rarely in America and penguins occur frequently in Antarctica and rarely in Africa become stored for exemplars.

Stereotypes may act as retrieval cues to access frequency information associated with various points of view. Elephants, for example, may be associated with separate frequency information for Africa, Asia, Europe, and so on. Adopting a point of view for which such frequency information exists may simply activate how frequently that exemplar occurs for that point of view. When such information does not exist, frequency information for the most similar point of view having such information may be used instead.

The Construction of Graded Structure.

Assuming that typicality is computed as stated in Equation 1, categories may have different graded structures from different points of view for two basic reasons. First, the standard of comparison (i.e., the category concept) may have different strategic and objective properties for different points of view. Exemplars having properties similar to the strategic and objective properties for one point of view may differ substantially from the strategic and objective properties for another. Consequently, exemplars may be typical of a category from one point of view and atypical from another. The second reason graded structure may change with point of view concerns frequency of



instantiation. Since exemplars that are frequent instantiations from one point of view may be infrequent from another, they may be typical of the category from the first point of view and atypical from the second.

This account produces the fundamental restructuring of a category discussed by Roth and Shoben (1983). Because the similarity comparison process that constructs graded structure operates on different information for different points of view, the similarity of exemplars to the category concept and to each other changes with point of view. This account also suggests why populations in the Emory community were able to accurately construct graded structures from each other's point of view. Such accuracy may simply reflect the fact that the various populations of the Emory community share the same stereotypes about each other. Undergraduates, for example, may have the same stereotype of faculty that faculty have of themselves. By having mutual stereotypes, different populations may selectively access information from categories in the same way when taking the same point of view, thus resulting in the same graded structure. Although these populations apparently have an extensive amount of mutual knowledge, it follows from this account that populations whose stereotypes were less similar would exhibit less accuracy when adopting each other's point of view.

## Conclusion

Given that a category can have many graded structures, what can we conclude about graded structure in general? One possibility is that graded structure represents the organization of a category's knowledge base in long term memory. In terms of search models, graded structure may reflect how strongly exemplars are associated to one another and to their category concept. In terms of comparison models, graded structure may reflect how many properties exemplars share with their category concept. If graded structure represents a category's organization, however, then how do we interpret the multiple graded structures of a category? Does each graded structure partially represent the category's overall organization in memory; and if so, how can we combine all of a category's graded structures to determine its overall organization? Even more problematic, graded structures may not reflect certain kinds of information stored in a category's knowledge base. For example, Barsalou (1981) found that individual subjects' typicality rankings correlated only .27 with the order in which they generated exemplars from goal-derived categories. Similarly, Barsalou and Sewell (1984) observed correlations of only .27 between group-derived typicality rankings and the order in which individual subjects generated exemplars from common taxonomic categories. Consequently, the production of exemplars from a category appears to involve information that is not captured by its graded structure. It would not be surprising if other conceptual tasks also established and used information not indexed by graded structure.

Perhaps a more fruitful way to view graded structure is as an index of the concepts people construct in working memory. Assuming that graded structure results from comparing exemplar concepts to category concepts, careful analysis of typical and atypical exemplars may reveal the information represented in a category concept. Properties of typical exemplars are likely to be represented in their category concept, whereas properties of atypical exemplars are not. Most importantly, observation of graded structures across points of view can reveal the specific versions of category concepts that become constructed for each point of view.

Although a category may be represented by many different concepts across many points of view, a dominant category concept may become entrenched in long term memory if the category is processed from the same point of view on



many occasions. Consequently, people having different dominant points of view may establish different dominant concepts of a category, although these dominant concepts may coexist with numerous other less dominant concepts established while taking less frequent points of view. For example, frequently perceiving <u>cars</u> from the point of view of a car thief might cause properties relevant to breaking into, starting, and illegally selling cars to become central to the category concept. In contrast, frequently perceiving <u>cars</u> from the point of view of a car buyer might cause properties relevant to mileage, comfort, safety, and style to become central. For any dominant point of view, the respective category concept should become entrenched in memory, assuming that frequent processings of a representation in working memory increasingly establish it in long term memory. These dominant concepts may be the ones that guide processing in neutral contexts, when no point of view or context is specified.

To the extent entrenched knowledge guides the processing of new events (cf. Goodman, 1955; Sternberg, 1982), representations established by a dominant point of view may control encoding. A car thief may see the weak points of a car's security system even in the absence of an intention to steal it, and the recent buyer of a car may continue assessing the quality of cars in the absence of needing a new one. But given people's ability to adopt points of view and construct novel concepts for categories, a new point of view may become dominant and, in time, lead to new dominant concepts. By acquiring a new point of view and frequently using it to construct new category concepts, the entrenched knowledge that typically guides processing may change. The ability to see the world from new points of view makes us extremely adaptive organisms and is probably central to our creative abilities.

## References

Anderson, R.C., & Pichert, J.W. (1978). Recall of previously unrecallable information following a shift in perspective. Journal of Verbal Learning and Verbal Behavior, 17, 1-12.

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

Barsalou, L.W. (1981). Determinants of graded structure in categories. Unpublished doctoral dissertation, Stanford University.

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

Barsalou, L.W. (1983). Ad hoc categories. Memory & Cognition, 11, 211-227.

Barsalou, L.W. (1984). Determinants of graded structure in categories. Under review.

Barsalou, L.W., & Sewell, D.R. (1984). Contrasting the organization of scripts and categories: Evidence for associations between actions in scripts. In preparation.

Battig, W.F., & Montague, W.E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology, 80, (Monograph Supplement 3, Pt. 2).

Birnbaum, M.H., & Stegner, S.E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. Journal of Personality and Social Psychology, 37, 48-74.

Bloom, R.B. (1976, May). Morally speaking, who are todays teachers? Phi Delta Kappan, 624-625.

Bloom, R.B. (1977). Resistance to faking of the defining issues test of moral development. Unpublished manuscript, College of William and Mary, Williamsburg, VA.



Clark, H.H., & Carlson, T.B. (1981). Context for comprehension. In J. Long & A. Baddeley (Eds.), Attention and performance IX (pp. 313-330). Hillsdale, NJ: Erlbaum.

Clark, H.H., & Carlson, T.B. (1982). Speech acts and hearers' beliefs. In N.V. Smith (Ed.), Mutual knowledge (pp. 1-59). New York: Academic Press.

Clark, H.H., & Marshall, C.R. (1981). Definite reference and mutual knowledge. In A.K. Joshi, I. Sag, & B. Webber (Eds.), Elements of discourse understanding (pp. 10-63). Cambridge, Eng.: Cambridge University Press.

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

Drewnowski, A. (1983). Cognitive structure in obesity and dieting. In M.R.C. Greenwood (Ed.), Obesity: Contemporary issues in clinical nutrition (pp. 87-101). New York: Churchill Livingstone.

Galambos, J.A., & Rips, L.J. (1982). Memory for routines. Journal of Verbal Learning and Verbal Behavior, 21, 260-281.

Glass, A.L., & Holyoak, K.J. (1975). Alternative conceptions of semantic memory. Cognition, 3, 313-339.

Goodman, N. (1955). Fact, fiction, and forecast. Cambridge, MA: Harvard University Press.

Guilford, J.P., & Fruchter, B. (1973). Fundamental statistics in psychology and education. New York: McGraw-Hill.

Hamilton, D.L. (1979). A cognitive-attributional analysis of stereotyping. In L. Berkowitz (Ed.), Advances in experimental social psychology, Volume 12. New York: Academic Press.

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

McGeorge, C. (1975). Susceptibility to faking on the defining issues test of moral development. Developmental Psychology, 11, 108.

Medin, D.L., & Schwanenflugel, P.J. (1981). Linear separability in classification learning. Journal of Experimental Psychology: Human Learning and Memory, 7, 355-368.

Medin, D.L., & Smith, E.E. (1981). Strategies and classification learning. Journal of Experimental Psychology: Human Learning and Memory, 7, 241-253.

Medin, D.L., & Smith, E.E. (1984). Concepts and concept formation. Annual Review of Psychology, in press.

Mervis, C.B., & Rosch, E. (1981). Categorization of natural objects. Annual Review of Psychology, 32, 89-115.

Nigro, G., & Neisser, U. (1983). Point of view in personal memories. Cognitive Psychology, 15, 467-482.

Owens, J., Dafoe, J., & Bower, G.H. (1977). Taking a point of view: Identification and attributional processes in story comprehension and memory. Talk given at the American Psychological Association Convention, San Francisco.

Reed, S.K. (1972). Pattern recognition and categorization. Cognitive Psychology, 3, 382-407.

Reitman, J.S., & Bower, G.H. (1973). Storage and later recognition of exemplars of concepts. Cognitive Psychology, 4, 194-206.

Rest, J.R. (1979). Development in judging moral issues (pp. 214-219). Minneapolis: Minnesota Press.

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

Rosch, E. (1975). Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 104, 192-233.



Rosch, E., & Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. Cognitive Psychology, 8, 382-439.

Roth, E.M., & Shoben, E.J. (1983). The effect of context on the structure of categories. Cognitive Psychology, 15, 346-378.

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

Smith, E.E., Shoben, E.J., & Rips, L.J. (1974). Structure and processes in semantic memory: A featural model for semantic decisions. Psychological Review, 81, 214-241.

Sternberg, R.J. (1982). Natural, unnatural, and supernatural concepts. Cognitive Psychology, 14, 451-488.

Sudman, S., & Bradburn, N.M. (1982). Asking questions. San Francisco: Jossey-Bass.

Taylor, S.E. (1980). A categorization approach to stereotyping. In D.L. Hamilton (Ed.), Cognitive processes in stereotyping and intergroup behavior. Hillsdale, NJ: Erlbaum.

Taylor, S.E., & Crocker, J. (1981). Schematic bases of social information processing. In E.T. Higgins, C.P. Hermann, & M.P. Zanna (Eds.), Social cognition: Cognitive structures and processes underlying person memory and social judgment. Hillsdale, NJ: Erlbaum.

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

Wilks, Y., & Bien, J. (1983). Beliefs, points of view, and multiple environments. Cognitive Science, 7, 95-119.

Yussen, S.R. (1976). Moral reasoning from the perspective of others. Child Development, 47, 551-555.

## Footnotes

This work was supported by a grant from the Emory University Research Committee and by grant IST-8308984 from the National Science Foundation to the first author. We are grateful to George Furnas, Douglas Medin, and Ulric Neisser for helpful comments. Requests for reprints should be addressed to Lawrence W. Barsalou, Department of Psychology, Emory University, Atlanta, GA 30322.

[1]Certain formal categories arguably are without graded structure. For example, all instances of the number *two* may be equally typical. Generally, graded structure can only occur in categories whose members differ. As Barsalou (1981, 1984) shows, a wide variety of differences among members may underly their ordering by typicality.

[2]Although stereotypes such as *hippie* and *redneck* may lead to accurate inferences on some occasions, we hasten to point out that they may lead to inaccurate inferences on others and that they generally offer a narrow and prejudiced view of the world.

[3]Actually, Barsalou (1983) found equal agreement between common taxonomic categories and ad hoc categories, the latter categories being goal-derived categories that are poorly established in memory.

[4]Given the rate of response to our mailings was less than unity, data for the faculty and graduate populations are subject to sampling biases. However, we have been unable to find any systematic differences between these data and those for the undergraduates, where standard sampling from a subject pool was in effect. Consequently, it appears that sampling biases may have had only minimal effects on our results, if any at all.

[5]Group reliability also estimates the correlation between *sums* of rankings across judges as well as between *means* of rankings across judges.

[6]Both subject and group reliability can be computed for ratings as well



as for rankings.

[7]Negative values of $s$ can and do occur due to measurement error.

[8]Actually, Fisher $z$ scores can become infinitely large or small, but values less than |3| occur for correlations less than |.995|.

## Appendix

### Subject Reliability

Reliability can take two forms, what we will refer to as *subject reliability* and *group reliability*. Imagine that a group of $m$ judges, randomly sampled from some population, rank orders $n$ items on some dimension (e.g., $m$ undergraduates rank $n$ animals by size). Subject reliability is basically how much any two judges agree in their rankings on the average (i.e., the average correlation across all possible pairs of judges).

Imagine that a second group of $m$ judges, randomly sampled from the same population, ranks the same $n$ items on the same dimension. Group reliability estimates how much the first group's *mean* rankings of the items correlate with the mean rankings of the items by the second group of $m$ judges.[5] Whereas subject reliability assesses agreement among judges within a sample, group reliability assesses the stability of a group's mean rankings. Guilford and Fruchter (1973) discuss various measures of subject and group reliability.[6]

To measure subject reliability in our experiments, we will first compute Kendall's coefficient of concordance, $W$, for the $m$ judges by $n$ items matrix of rankings for each category. As described by Guilford and Fruchter (1973), a transformation of $W$ provides a measure of subject reliability, $S$, where

$$S = \frac{mW - 1}{m - 1} \quad (5).$$

This measure estimates the average correlation between all pairs of subjects judging the same category from the same point of view.

### Shift

The concept of group reliability is central to our technique for measuring shift. Although group reliability can be interpreted as the



correlation between the mean rankings for two samples, it can be estimated from a single sample. In this case, group reliability estimates how well the sample's mean rankings would correlate with those of an imaginary sample of the same size from the same population. A transformation of subject reliability, $S$, provides a measure of group reliability, $G$, where

$$G = \frac{mS}{1 + (m-1)S} \quad (6).$$

As will be seen in a moment, hypothesis testing in our experiments requires that we *estimate* the correlation between the means of two actual samples from the same population, $x$ and $x'$, without actually computing the correlation. When the two samples have been randomly sampled from the same population, this correlation can be estimated by the geometric mean of the samples' group reliabilities, $\sqrt{G_x G_{x'}}$, where $G_x$ is the group reliability of one sample, and $G_{x'}$ is the group reliability of the other. This equivalence follows from the relation between the correlation of two variables and their group reliabilities (Guilford & Fruchter, 1973):

$$r_{xy} = R_{xy}\sqrt{G_x G_y} \quad (7)$$

where $r_{xy}$ is the obtained correlation between $x$ and $y$, $R_{xy}$ is the true correlation, and $G_x$ and $G_y$ are the group reliabilities. Basically, this equation states that an obtained correlation between two variables is less than their true correlation to the extent the means entering the correlation are unstable, that is, to the extent their group reliabilities are low.

When the means of two samples from the same population are being correlated, the equation becomes

$$r_{xx'} = R_{xx'}\sqrt{G_x G_{x'}} \quad (8),$$

where $x$ represents the means from one sample, and $x'$ represents the means from the other. Since the two samples differ only in being different random samples from the same population, $R_{xx'}$ is assumed to be 1. That is, the true correlation between two groups should be 1 if the two groups have been randomly sampled from the same population. Consequently, the correlation between the mean rankings of two samples is equivalent to the geometric mean of their group reliabilities:

$$r_{xx'} = \sqrt{G_x G_{x'}} \quad (9).$$

The relation between a correlation and its reliabilities stated in Equation 9 can be used to determine whether the graded structures generated by two points of view differ. Assume that $m$ subjects rank order $n$ exemplars of a category from the point of view of the average American citizen, and that another $m$ subjects rank the exemplars from the point of view of the average Chinese citizen. Let the null hypothesis be that the point of view manipulation has no effect, namely, subjects in both groups generate the same graded structure (i.e., the no shift hypothesis). If subjects were randomly assigned to the two points of view, it follows that the relationship in Equation 9 should be obtained. That is, if two identical groups judge the same items, the correlation of their mean rankings should equal the geometric mean of their reliabilities. If the point of view manipulation has an effect, however, then the two groups should generate overall rankings that differ, and their correlation should be less than the geometric mean of their reliabilities. The null and alternative hypotheses can be summarized as follows:

$$(H0) \quad \sqrt{G_x G_y} - r_{xy} = 0$$

$$(H1) \quad \sqrt{G_x G_y} - r_{xy} > 0.$$

H0 states that if point of view has no effect on how judges order items, then the correlation of sample means should equal the geometric mean of the group reliabilities. H1 states that if point of view has an effect, then the correlation should be less than the group reliabilities. As the size of



this difference increases, the graded structures produced from the two points of view increasingly differ.

Given the nature of the scale on which correlations (including reliabilities) are measured, it is impossible to compare different values of $\sqrt{G_x G_y} - r_{xy}$. When $\sqrt{G_x G_y} - r_{xy}$ is .90 - .70, for example, the difference of .20 is larger than the difference of .20 when $\sqrt{G_x G_y} - r_{xy}$ is .50 - .30. Consequently, it is necessary to transform $\sqrt{G_x G_y}$ and $r_{xy}$ to Fisher $z$ scores to obtain comparable differences. The following transformed difference score for shift, $s$, allows meaningful comparisons of the graded structures generated from two points of view:

$$s = F(\sqrt{G_x G_y}) - F(r_{xy}) \quad (10),$$

where $F(w)$ refers to the value of $w$ obtained from Fisher's $r$ to $z$ transformation. After computing values of $s$ for two points of view across a set of categories, the average value of $s$ can be tested with a $t$ test to determine whether it significantly exceeds 0. If it does, we will reject H0, and if it does not, we will accept H0.

As can be seen from Equation 10, $s$ generally ranges from 0 to $2F(\sqrt{G_x G_y})$. $s$ equals 0 when $r_{xy} = \sqrt{G_x G_y}$, namely when two points of view generate identical graded structures.[7] $s$ equals $F(\sqrt{G_x G_y})$ when $r_{xy} = 0$, namely when two points of view generate unrelated graded structures. And $s$ equals $2F(\sqrt{G_x G_y})$ when $r_{xy} = -\sqrt{G_x G_y}$, namely when two points-of-view generate perfectly inverse graded structures. Since Fisher $z$ scores generally range from -3 to +3, $s$ generally ranges from 0 to 6 (i.e., $2F(\sqrt{G_x G_y}) = 2\sqrt{3^2}$).[8] However, since the group reliabilities obtained in the experiments to follow average around .90 at the highest, $s$ will range from 0 to 3 (i.e., $2F(\sqrt{.90^2}) = 2.942$). The relevant ranges of $s$ will be discussed in the context of individual experiments.

Table 1

Average Subject Reliabilities

for Experiments 1a, 1b, and 1c

| | | | | |
|---|---|---|---|---|
| | International Points-of-View | | | |
| | American | French | Chinese | African |
| Common Taxonomic Categories | .446 | .226 | .337 | .300 |
| Goal-Derived Categories | .408 | .304 | .444 | .262 |
| | Domestic Points-of-View | | | |
| | Hippie | Businessman | Housewife | Redneck |
| Common Taxonomic Categories | .248 | .482 | .394 | .511 |
| Goal-Derived Categories | .330 | .354 | .438 | .447 |
| | Self and American* Points-of-View | | | |
| | Self | American | | |
| Common Taxonomic Categories | .426 | .494 | | |
| Goal-Derived Categories | .360 | .423 | | |

* Twenty-four subjects are included in this American point-of-view, whereas only half of these subjects are included in the same point-of-view in the International experiment.



Table 2

Average Group Reliabilities, Correlations, and Shift

for Experiments 1a, 1b, and 1c

| | Reliability | Correlation | Shift |
|---|---|---|---|
| International Points-of-View | | | |
| Common Taxonomic Categories | .744 | .352 | .582 |
| Goal-Derived Categories | .783 | .403 | .641 |
| Domestic Points-of-View | | | |
| Common Taxonomic Categories | .771 | .376 | .697 |
| Goal-Derived Categories | .832 | .323 | .859 |
| Self and American Points-of View | | | |
| Common Taxonomic Categories | .926 | .850 | .276 |
| Goal-Derived Categories | .917 | .780 | .476 |

Table 3

Average Shift for Point-of-View Pairs

for the International Points-of-Views

| | French | Chinese | African |
|---|---|---|---|
| **Common Taxonomic Categories** | | | |
| American | .287 | .970 | .665 |
| French | | .327 | .649 |
| Chinese | | | .594 |
| **Goal-Derived Categories** | | | |
| American | .600 | 1.072 | .720 |
| French | | .557 | .464 |
| Chinese | | | .433 |

Note: Shift scores greater than .225 and .330 are significantly different from zero at the .05 and .01 levels, respectively. Differences between average shift scores of .373 and .499 are significant at the .05 and .01 levels, respectively.



Table 4

Average Exemplar Rankings for

Two Categories from Experiments 1a, 1b, and 1c

| | International Points-of-View | | | | Domestic Points-of-View | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Exemplar | Amer | Fren | Chi | Afri | Hip | Busmn | Hsewf | Rdnek | Self |
| **Fruit** | | | | | | | | | |
| apple | 1.7 | 3.2 | 4.4 | 4.6 | 3.9 | 1.4 | 1.8 | 2.1 | 2.6 |
| banana | 3.6 | 3.8 | 4.6 | 2.0 | 3.5 | 3.0 | 3.2 | 3.6 | 3.8 |
| peach | 3.3 | 4.4 | 5.1 | 6.5 | 4.6 | 4.8 | 3.9 | 3.5 | 4.1 |
| pineapple | 5.3 | 4.7 | 3.8 | 4.7 | 4.2 | 6.2 | 5.9 | 6.3 | 4.2 |
| grapefruit | 4.8 | 5.0 | 4.8 | 5.7 | 6.4 | 3.7 | 4.5 | 5.2 | 4.8 |
| watermelon | 4.8 | 5.6 | 6.5 | 4.1 | 4.2 | 4.8 | 5.1 | 2.5 | 4.6 |
| cantaloupe | 5.3 | 3.2 | 4.3 | 5.4 | 5.7 | 4.9 | 4.5 | 4.8 | 4.4 |
| mango | 7.2 | 5.8 | 2.6 | 3.1 | 3.5 | 7.2 | 7.2 | 8.0 | 7.5 |
| **Hobbies** | | | | | | | | | |
| collecting stamps | 5.1 | 4.9 | 4.2 | 6.5 | 4.8 | 4.6 | 5.8 | 7.0 | 5.1 |
| sports | 1.8 | 4.8 | 3.7 | 2.8 | 5.1 | 1.2 | 5.8 | 2.5 | 3.4 |
| photography | 4.6 | 2.9 | 3.0 | 6.2 | 3.7 | 3.5 | 3.7 | 5.8 | 3.2 |
| horseback riding | 5.6 | 4.2 | 7.0 | 4.9 | 3.2 | 5.7 | 5.8 | 3.5 | 4.8 |
| dancing | 6.1 | 4.8 | 4.9 | 3.3 | 5.8 | 5.2 | 3.2 | 5.1 | 4.7 |
| sewing | 5.2 | 5.2 | 4.3 | 4.6 | 4.9 | 7.5 | 2.1 | 4.8 | 6.4 |
| playing the piano | 4.8 | 3.8 | 4.7 | 5.7 | 5.5 | 5.3 | 2.9 | 6.1 | 4.2 |
| fishing | 2.8 | 5.5 | 4.2 | 2.0 | 3.1 | 3.0 | 6.9 | 1.2 | 4.2 |

Table 5

Average Shift for Point-of-View Pairs

for the Domestic Points-of-View

| | Common Taxonomic Categories | | |
|---|---|---|---|
| | Businessman | Housewife | Redneck |
| Hippie | 1.030 | .787 | .545 |
| Businessman | | .281 | .762 |
| Housewife | | | .775 |
| | **Goal-Derived Categories** | | |
| | Businessman | Housewife | Redneck |
| Hippie | .982 | 1.048 | .828 |
| Businessman | | .563 | .780 |
| Housewife | | | .953 |

Note: Shift scores greater than .230 and .338 are significantly different from zero at the .05 and .01 levels, respectively. Differences between average shift scores of .381 and .510 are significant at the .05 and .01 levels, respectively.



Table 6

Average Shift for

the American Versus the Domestic Points-of-View

| | Hippie | Businessman | Housewife | Redneck |
|---|---|---|---|---|
| Common Taxonomic Categories | .769 | .155 | -.084 | .517 |
| Goal-Derived Categories | .943 | .444 | .324 | .653 |

Note: Shift scores greater than .197 and .289 are significantly different from zero at the .05 and .01 levels, respectively. Differences between average shift scores of .326 and .437 are significant at the .05 and .01 levels, respectively.

Table 7

Average Estimated Ability of One

Population to Take the Point-of-View of Another

| Subjects | F take U | G take U | G take F | U take F |
|---|---|---|---|---|
| Faculty | 2.462 | 2.308 | 2.615 | 2.769 |
| Graduates | 2.692 | 2.692 | 2.308 | 3.538 |
| Undergraduates | 2.769 | 2.769 | 2.538 | 3.154 |

*Note*. *X* take *Y* is how well population *X* can take the point-of-view of of population *Y*, where F = faculty, G = graduates, and U = undergraduates. The scale is 1 = not at all, 2 = a little, 3 = fairly well, 4 = a lot, 5 = perfectly.



Table 8

Average Subject Reliabilities

for Experiment 2

| | Subject Population | | |
|---|---|---|---|
| Point-of-View | Faculty | Graduate | Undergraduate |
| Faculty | .446 | .394 | .326 |
| Undergraduate | .408 | .337 | .443 |

Table 9

Average Group Reliabilities, Correlations, and Shift

for Experiment 2

| | Reliability | Correlation | Shift |
|---|---|---|---|
| Different Points of View | .901 | .248 | 1.287 |
| Same Points-of-View | .902 | .881 | .043 |



Table 10

Average Shift for

Point-of-View Pairs for Experiment 2

| | $F_sU_p$ | $G_sF_p$ | $G_sU_p$ | $U_sF_p$ | $U_sU_p$ |
|---|---|---|---|---|---|
| $F_sF_p$ | 1.292 | -.069 | 1.223 | .107 | 1.384 |
| $F_sU_p$ | | 1.307 | -.206 | 1.229 | .356 |
| $G_sF_p$ | | | 1.239 | -.110 | 1.328 |
| $G_sU_p$ | | | | 1.196 | -.002 |
| $U_sF_p$ | | | | | 1.232 |

Note. F = faculty, G = graduates, U = undergraduates, s = subjects, p = point-of-view. Shift scores greater than .209 and .307 are significantly different from zero at the .05 and .01 levels, respectively. Differences between average shift scores of .347 and .464 are significant at the .05 and .01 levels, respectively.