Journal of Personality and Social Psychology
1992, Vol. 63, No. 4, 553-567

0022-3514/92/$3.00

Role of Meaningful Subgroups in Explaining Differences in Perceived Variability for In-Groups and Out-Groups

Bernadette Park, Carey S. Ryan, and Charles M. Judd
University of Colorado at Boulder

Five aspects of the complexity of the knowledge representation of business and engineering majors were examined to see whether these differed by group membership and whether these differences were related to differences in perceived variability. Significantly more subgroups were generated when describing the in-group than the out-group; this difference predicted the relative tendency to see the in-group as more variable, and when controlled for statistically, out-group homogeneity effects were eliminated. Familiarity, redundancy, number of attributes used to describe the group, and the deviance of the subgroups from the larger group generally showed differences for in-group and out-group but did not show consistent evidence of mediation. In a 2nd study, Ss who were asked to sort group members into meaningful subgroups perceived greater variability relative to those who did not perform the sorting task.

Researchers in the area of stereotyping have been concerned with the tendency to view members of an out-group as less diverse and more stereotypic than members of the group see themselves, that is, the out-group homogeneity effect. The effect appears to be quite robust. It has been demonstrated with groups that are relatively unfamiliar with one another (Linville, Fischer, & Salovey, 1989), with highly familiar groups (Park & Judd, 1990; Park & Rothbart, 1982), and even in a minimal group situation (Judd & Park, 1988; Mullen & Hu, 1989). A number of tentative hypotheses have been advanced to explain the out-group homogeneity effect, but none have received strong empirical support (see Park, Judd, & Ryan, 1991, for an overview). In this article, we investigate one particular mechanism hypothesized to mediate differences in perceived variability for in-groups and out-groups. Specifically, we examine differences in the organization of knowledge about groups into subgroups or subtypes that are at a subordinate level of generality relative to knowledge about the superordinate category.

The impetus for the present studies comes from an interesting, but unexpected, finding from Experiment 2 in Park and Judd (1990). In that study, subjects were asked to think aloud as they completed various measures that indexed perceived variability for both the in-group and the out-group. Our interest was in the types of information subjects would use from their knowledge base in completing these judgments. We were particularly interested in whether subjects differentially used the self and specific known instances of the group when reasoning about perceived variability. Although the self was more likely to be used when completing judgments for the in-group, specific known instances were more likely to be talked about when judging the out-group. Most interesting, subjects appeared to use subgroups or subtypes when thinking about the in-group and not when thinking about the out-group. The protocols suggested that knowledge about the in-group was organized into meaningful subgroups that differed from one another in important ways. When judging the degree of variability within the in-group, subjects reasoned from differences among these subgroups. Such organization was not present in the protocols for the out-group. Here, subjects seemed instead to use the few instances they personally knew to reason about the group. We wondered, therefore, whether differences in perceived variability for in-groups versus out-groups might derive from having a highly organized knowledge structure about the in-group that enables systematic comparisons among subgroups in a manner that is not possible for out-groups. Although suggestive, the data from this experiment were far from conclusive. The study did not directly measure subjects' knowledge of subgroups for in-groups versus out-groups. In addition, it was difficult to assess the mediation of out-group homogeneity by differences in the mention of subgroups because, as the result of the collection of protocols, a relatively small number of subjects provided data. We attempt to pursue this line of inquiry in the present experiments.

Particularly relevant to our subgrouping hypothesis is a series of studies conducted by Park and Rothbart (1982), which were designed to demonstrate out-group homogeneity with men and women. In the fourth study reported, the authors investigated a mechanism through which out-group homogeneity might be maintained. They found that both men and women better remembered subordinate or differentiating information (i.e., occupation) for in-group relative to out-group members. The authors argued this was because the superordinate category information was relatively uninformative to in-group members, whereas it was meaningful to the out-group. Thus, for a male subject, knowing that the protagonist was male was not very

This research was partially funded by Grant BNS-8819372 from the National Science Foundation to Bernadette Park, Grant BNS-8806757 from the National Science Foundation to Charles M. Judd, and Grant RO1MH45049-01A2 from the National Institute of Mental Health to Bernadette Park and Charles M. Judd.

Correspondence concerning this article should be addressed to Bernadette Park, Campus Box 345, Psychology, University of Colorado, Boulder, Colorado 80309.

useful, and instead, it was important to know the individual's occupation. In contrast, knowing that the protagonist was a female in and of itself carried meaning so that the subordinate information was less essential. In general then, subordinate information should be encoded, remembered, and used more often for in-group than for out-group members. Park and Rothbart argued that the tendency to think about the in-group in terms of subordinate attributes, and the out-group in superordinate attributes, should only perpetuate differences in perceived variability.

In the second study reported by Park and Rothbart (1982), subjects were asked about their perceptions of the prevalence of gender stereotypic and counterstereotypic attitude statements among subgroups of men and women who fit or did not fit the superordinate stereotype (e.g., male economics majors vs. male dance majors). The mean estimates suggested that judgments of the out-group depended primarily on the superordinate stereotype, such that stereotypic attributes (as defined by the superordinate category) were judged as more prevalent than counterstereotypic attributes, regardless of whether the subgroup was consistent or inconsistent with the superordinate stereotype. Judgments of the in-group, in contrast, were based primarily on the attributes of the subgroup, such that consistent subgroups were judged to have more stereotypic than counterstereotypic attributes, but for inconsistent subgroups, the reverse was true (Rothbart & John, 1985). Thus, for example, when judging male economics majors and male dance majors, female subjects rated both groups as having more of the stereotypically masculine items than the stereotypically feminine items, although the slope of this difference was less steep for the (nonstereotypic) male dance majors. Female subjects thus saw both groups as fitting the male stereotype. In contrast, male subjects judged male economics majors to have more stereotypically masculine items than stereotypically feminine items, but for male dance majors, the reverse was true such that feminine items were seen as more prevalent. In a sense, then, the stereotypicality of the major (i.e., the subgroup) determined judgments for the in-group, whereas the stereotypicality of the gender (i.e., the superordinate group) determined judgments for the out-group. (see Rothbart & John, 1985, for a more complete discussion of these data).

Other researchers have certainly discussed the issue of subtypes or subgroups within the stereotyping literature. For example, Brewer, Dull, and Lui (1981), Brewer and Lui (1984), and Taylor (1981) have all argued that superordinate stereotypes are further organized along the lines of subtypes within the general category. Weber and Crocker (1983) argued that members who disconfirm the stereotype are subtyped and set apart, and this enables the superordinate stereotype to remain relatively unchanged. Only Brewer and Lui (1984), however, have examined differences in subtyping as a function of group membership. In their research, elderly subjects tended to differentiate more among a set of elderly stimuli relative to young subjects (suggesting greater subtyping for the in-group), but this was true only for those subgroups that elderly subjects considered themselves to be members of. One problem with this study is that both target groups were not included in the design (i.e., the primary comparisons were for the elderly target group only, not for the young group), so that it is impossible to separate in-group–out-group effects from subject group main effects (i.e., the tendency for the elderly to see everyone in a more differentiated manner). None of the subtyping research examines the relation between subtyping and perceived variability of the groups. Interestingly, both Taylor (1981) and Weber and Crocker (1983) focused on the ability of the subtyping process to protect superordinate stereotypes from change. In the present analysis, we examine subgrouping as a means for producing greater perceived variability in the larger group stereotype, an effect apparently opposite that suggested by Taylor (1981) and Weber and Crocker (1983).

Linville (1982; Linville & Jones, 1980) has argued that individuals have more complex representations for in-groups than for out-groups. However, *complexity,* as used in Linville's work, means something quite different from our conceptualization of organization by subgroups. Specifically, complexity is assessed in Linville's work by the number of orthogonal dimensions along which a group is evaluated. The fewer the number of dimensions, the less complex and the more extreme will be an evaluative judgment of the group. Thus, Linville argued that in-groups are evaluated on a greater number of orthogonal dimensions, and as a result, evaluative ratings of the group are less extreme than in the case of out-groups.

Our interest is not in individual dimensions, but rather in clusters of individuals, with each cluster evaluated on a number of dimensions. The number of orthogonal dimensions need not be related to the number of subgroups or to similarities and differences among the subgroups. Moreover, we are interested in how subgroup representation relates to perceptions of variability of group members, not simply to the extremity of evaluation of the members. It is possible, however, that redundancy among the subgroups (as defined by Judd & Lusk, 1984) is negatively related to perceived variability. To the extent that all of the subgroups described by a subject are similar and different in predictable ways, this may lead to less perceived variability than if the subgroups each have unique characteristics that are not relevant to the other subgroups. To explore this possibility, we assess redundancy in the perception of subgroups in the first experiment in a manner analogous to Judd and Lusk (1984), but our focus on organization of knowledge into subgroups is clearly different from Linville's (1982; Linville & Jones, 1980) conception of complexity.

The data from Park and Rothbart (1982, Experiments 2 and 4), Park and Judd (1990, Experiment 2), Brewer et al. (1981), and Brewer and Lui (1984) are all suggestive that subjects organize in-group information into meaningful subgroups and that these play an important role in reasoning about the characteristics of the group. However, these data do not directly assess differences in the tendency to organize information by subgroups for in-groups and out-groups, nor do they establish a direct link between knowledge of subgroups and the perception of variability.

In the following two studies we explore these issues. Experiment 1 uses a correlational design to examine the role of subgroups in perceptions of the larger group. Specifically, we examine whether the number of subgroups that one can generate and describe differs for in-groups versus out-groups. Furthermore, we test whether such differences are related to differences in perceived variability for in-groups and out-groups. We include

in this study a number of additional variables that have been hypothesized to mediate judgments of variability. Specifically, we examine the role of familiarity with the group (Linville et al., 1989), the similarity between perceptions of self and the in-group (Park & Judd, 1990), and the variability in specific retrieved instances in mediating judgments of variability for the in-group and out-group (Park & Judd, 1990). We also assess the extent to which the subgroups are perceived in a redundant manner (Linville, 1982; as defined by Judd & Lusk, 1984), and the relation of this to perceived variability, in addition to several other measures of the complexity or degree of differentiation of the stereotype. Experiment 2 uses an experimental manipulation of the presence of subgrouping and again assesses the affect of this on perceived variability.

Experiment 1

Method

Subjects

Thirty business majors and 30 engineering majors who were in their junior or senior year agreed to participate in a study about "students' perceptions of various majors on campus." Subjects were paid $10 for their participation.

Procedure

Each subject completed a questionnaire that contained various judgments about the group of business majors and the group of engineering majors who were in their junior or senior year at the University of Colorado. All subjects thus rated both the in-group and the out-group. Previous studies (Judd, Ryan, & Park, 1991) indicate that subjects in both majors have stereotypes about both groups. Order of group ratings was counterbalanced. The order of judgment tasks was as follows:

Open-ended descriptions. For each group, subjects first provided open-ended descriptions of their impression of what the group as a whole was like. Specifically, they were asked to write 8 to 10 sentences that described the group so that in 5 years they could reread the description and know what their impression was at the current time. They were then asked to list as many "sorts or types" of business (engineering) majors that came to mind when they thought about the larger group and, for each one, to "write a few sentences about what this subgroup is like and what makes them different from other business (engineering) majors." Next subjects identified one characteristic ("a word or phrase") that most aptly described members of each subgroup and that differentiated them from the rest of the group. Using the characteristics that she or he had generated, the subject then rated the extent to which each subgroup had each of the characteristics on a 7-point scale on which a 1 indicated *not at all* and a 7 indicated *definitely* (Judd & Lusk, 1984; Lusk & Judd, 1988).

Perceived group variability. Subjects performed three tasks to assess perceived dispersion and perceived stereotypicality (Park & Judd, 1990). First, subjects completed the dot task in which they used three sizes of stick-on dots to indicate the relative numbers of group members who fall at each point along a dimension. The dimension was labeled with endpoints (e.g., *analytical–not analytical*) and broken into eight boxes. Subjects were to place a dot in or leave blank each box to indicate the relative number of individuals at that point on the distribution. Next, subjects completed the percentage estimation task in which they estimated the percentage of group members that have a given trait or that would endorse a given attitude statement (Park & Rothbart, 1982). Finally, subjects completed the mean and range estimation task in which they rated where a group fell on average as well as where the group's most extreme members would fall on a scale labeled only at the endpoints (Jones, Wood, & Quattrone, 1981). These ratings were subsequently measured on a 21-point scale (0–20). These tasks have been described in detail by Park and Judd (1990) and by Judd et al. (1991).

For each task, subjects judged both groups with respect to the same set of four attitude statements and four traits. The attitude items included two that were considered stereotypic of business majors and counterstereotypic of engineering majors ("I often act on my gut feelings rather than rationally analyzing a situation" and "I enjoy meeting new people and having a large number of acquaintances") and two that were considered stereotypic of engineering majors and counterstereotypic of business majors ("I prefer a quiet evening of reading to a loud party" and "One of my favorite past-times is solving brain teasers such as Rubic's Cube"). Similarly, the four traits included two that were stereotypic of business majors and counterstereotypic of engineering majors (extroverted and impulsive) and two that were the reverse (reserved and analytical). The stereotypicality of these four attitude statements and four traits has been established in our earlier work (Judd et al., 1991; Park & Judd, 1990, Experiment 2).

Familiarity. Next subjects completed four tasks to assess their familiarity with group members. Subjects provided two global familiarity ratings in which they indicated how many group members they knew and how well they knew them. Ratings were made on 7-point scales in which a 1 indicated *don't know many/don't know well* and a 7 indicated *know a lot of them/know very well.*

Subjects were also asked to list the names of as many as 10 group members whom they knew personally. They then rated how well they knew each of the individuals they had listed on a 5-point scale in which a 1 meant they *knew the person very well* and a 5 meant they *did not know the person very well.*

Ratings of group members. Next subjects completed an individual mean and range estimation task, this time judging the first three group members they had listed in the task described in the preceding paragraph. Subjects indicated, for example, how extroverted each group member was, using the same type of 21-point scale used in the group mean and range estimation task. We then explained that people typically vary in their behavior from one situation to another and asked subjects to make two slashes on the same scale to indicate the most and the least extroverted this group member ever was. Subjects judged each group member with respect to the same four traits on which they had judged the group as a whole: extroverted, analytical, impulsive, and reserved (and not on the attitude statement items).

Self-ratings. Finally, after rating both groups, subjects rated themselves on the same set of eight dimensions on which they had rated the two groups. For each of the four trait and four attitude items, subjects were asked to make two judgments. First, they indicated simply whether the trait described them and, similarly, whether they agreed or disagreed with the attitude statement. Second, they indicated the extent to which each trait described them and the extent to which they agreed or disagreed with each attitude statement. These ratings were made on a 7-point scale in which a 1 indicated *definitely does not describe me/strongly disagree* and a 7 indicated *definitely does describe me/strongly agree.*

Results

In all of the following analyses, subject group was included as a between-subjects variable and target group as a within-subject variable. The test of the Subject Group × Target Group interaction reflected differences in the ratings of in-groups versus out-groups. We focus on this interaction where present. All such reported differences existed regardless of target or subject

group (i.e., the main effects for these variables were not reliable) unless otherwise noted. All reported effects were reliable with alpha set at .05 unless otherwise noted.

Measures of Knowledge Organization

The questionnaire included a variety of measures designed to assess the organization of subjects' knowledge about the two groups that we believed might mediate differences in perceived variability. Accordingly, we first tested for in-group–out-group differences in these various measures. The means corresponding to each of the measures discussed next appear in Table 1.

Number of subgroups. To see whether the number of subgroups subjects generated differed for the in-group and out-group, we simply counted the number of subgroups described in the subgroup-listing task of the open-ended description portion of the questionnaire. Consistent with Study 2 of Park and Judd (1990), subjects listed significantly more subgroups for the in-group (M = 4.77) than for the out-group (M = 3.65), $F(1, 58) = 29.38$.

Familiarity. A factor analysis was performed separately by target group on the four measures designed to assess familiarity. Again these measures were a global rating of the number of group members known, a global rating of how well these group members were known, number (up to 10) of known acquaintances listed, and average rating of how well these 10 acquaintances were known. The means are presented for each variable in Table 1. All four measures loaded highly on one factor (eigenvalue for the first unrotated factor = 2.53 for the business target group, 63% of the variance accounted for, and 2.42 for the engineering target group, 60% of the variance accounted for). Accordingly, a factor score combining the four familiarity ratings was computed separately for the in-group and the out-group for each subject. Clearly, subjects were more familiar with the in-group than with the out-group, $F(1, 58) = 122.10$.

Table 1
Differences in Measures of Knowledge Organization as a Function of Group Membership

Measure	In-group	Out-group
No. of subgroups	4.77	3.65
Familiarity[a]		
No. known	6.45	3.51
How well know	6.10	4.46
No. listed	9.55	5.55
How well know listed	2.25	2.55
Redundancy		
Subgroup	0.52	0.63
Attribute	0.50	0.57
No. of attributes listed		
Group	6.71	5.93
Subgroups	3.15	2.97
Subgroup deviance	3.11	2.95

[a] No. known: 1 = *don't know many* and 7 = *know a lot of them;* how well know: 1 = *don't know well* and 7 = *know very well;* no. listed: in free-listing task of known group members, maximum = 10; how well know listed: average ratings of listed individuals, 1 = *know person very well* and 5 = *do not know person very well.* (This item loaded negatively on the familiarity factor.)

Redundancy. These measures were intended to assess whether the redundancy in subjects' knowledge of subgroups, rather than the number of subgroups per se, predicts perceived variability. Perhaps it is only when the subgroups are each described on different and independent dimensions that the perception of greater variability ensues. Recall that for each subgroup listed, the subject generated an attribute that characterized the group and that distinguished the subgroup from the rest of the group. For example, one subject listed the following attributes for the four engineering subgroups she generated: self-centered, logical, prideful, and outdoors. Each subgroup was then rated on each of these subject-generated attributes. Thus, an Attribute × Subgroup matrix of ratings for each target group was generated by each subject. Using these ratings, redundancy was assessed in two ways. First, we calculated the average absolute correlation between all possible pairs of subgroups across attribute dimensions (Lusk & Judd, 1988). To the extent that the attributes are combined in a predictable way for each subgroup, ratings of a given subgroup will be highly correlated (i.e., redundant) with ratings of a second subgroup. If instead, knowing about one subgroup reveals little about a second subgroup, the two are nonredundant, and each contributes a unique bit of information about the larger group. Thus, the subgroup redundancy correlation assessed the extent to which the subgroups were seen in a similar fashion.

Second, we computed the attribute redundancy by taking the average absolute value of all possible correlations between attributes, across subgroups. To the extent that knowing where a subgroup stands on one attribute tells you where they stand on a second, the knowledge associated with the subgroups is relatively redundant. If instead, knowledge of one attribute is uninformative with respect to a second, then both attributes are needed to capture differences among the subgroups.

For each subject, both a subgroup redundancy correlation and an attribute redundancy correlation were calculated for the in-group and the out-group. These two measures of redundancy were highly correlated (r = .79 for the in-group and r = .64 for the out-group), so that for each subject they were combined to form a single measure of redundancy in subgroup knowledge for each group. Fisher's r-to-z transformation was applied to these correlations. Although the mean value of the correlations suggests greater redundancy for the out-group than for the in-group, this difference was not reliable, $F(1, 58) = 2.30$, $p = .14$. The relevant reverse-transformed mean correlations appear in Table 1.

Number of attributes. The open-ended descriptions were coded to obtain several additional measures of subjects' knowledge about the groups. First, the total number of different attributes or characteristics used in the description of the group as a whole was determined. Second, the number of different attributes used to describe each subgroup, averaged across subgroups, was calculated. Two coders scored all 60 subjects' questionnaires. Reliability between the two judges was .82 for the number of attributes in the global descriptions and .89 for the average number of attributes in the subgroup descriptions. The ratings of the two coders were averaged to provide one judgment. The number of attributes for the group as a whole correlated .54 for the in-group and .20 for the out-group, with the average number of attributes for the subgroups. The two mea-

sures were averaged and used as an index of the number of attributes in the knowledge structure. A greater number of attributes was generated for the in-group than for the out-group, $F(1, 58) = 6.25$. The main effect of subject group was also reliable in this analysis, such that business subjects generated more attributes overall than engineering subjects, $F(1, 58) = 5.66$.

Subgroup deviance. The two coders also scored the extent to which each subgroup was described in a manner that was deviant from the description of the group as a whole. Using a 5-point scale, the judges rated whether the description of a given subgroup was essentially identical to that of the entire group (subgroup deviance = 1) or completely different (i.e., used different dimensions or described as having the counterstereotypic pole of a dimension; subgroup deviance = 5). The reliability for the judgment was .61. This measure was designed to test in yet a different manner whether discrepancies between the subgroups and the larger group were necessary to produce greater perceived variability, as opposed to simply the presence of subgroup knowledge. In-group subgroups were judged to be marginally more deviant on the whole than out-group subgroups, $F(1, 58) = 2.82$, $p = .10$.

Table 2 presents the intercorrelations among the various measures of knowledge organization for in-group and out-group. We subtracted target group means (e.g., the average familiarity rating of business majors across both business and engineering subjects) from each variable before calculating the correlations so that the in-group–out-group correlations would not be influenced by differences in these.

Measures of Perceived Variability and Out-Group Homogeneity

Park and Judd (1990) identified two types of variability measures: those that assess the perceived dispersion of the group, such as a standard deviation or range, and those that assess the extent to which the group is seen to fit the group stereotype, such as the extremity of mean ratings on stereotypic and counterstereotypic items. We included measures of both forms in the questionnaire.

Perceived dispersion. We computed two measures of perceived dispersion. First, the dot task resulted in an estimated frequency distribution for the group on each item. We assigned the largest circles a value of three, the middle circles a value of two, the smallest circles a value of one, and blanks a value of zero and calculated a standard deviation for each distribution. These were averaged across the eight items to form one estimate of the standard deviation from the dot task. Contrary to previous research using these same groups (Judd et al., 1991; Park & Judd, 1990, Experiment 2), these estimates were essentially identical for the in-group ($M = 2.015$) and the out-group ($M = 2.016$), $F < 1$. The second measure of perceived dispersion came from the range estimation task. Specifically, we calculated the difference between the rating of the most and least extreme member of the group on each item. These range estimates were averaged across the eight items. As the means in Table 3 show, the in-group was judged to have a reliably greater range on average across the eight items relative to the out-group (i.e., out-group homogeneity), $F(1, 58) = 4.15$.

Perceived stereotypicality. We computed three measures of perceived stereotypicality. First, we examined subjects' estimates of the percentage of group members who possessed each trait, or would endorse each attitude statement, on average across the four stereotypic items and across the four counterstereotypic items. As the means in Table 3 suggest, stereotypic items were judged to be more prevalent by the out-group than by the in-group, whereas counterstereotypic items were judged to be less prevalent by the out-group, reflecting perceptions of greater stereotypicality by the out-group (i.e., out-group homogeneity), $F(1, 58) = 5.42$. The target group main effect was also reliable in this analysis, such that business majors were judged in a more stereotypic manner than engineering majors, $F(1, 58) = 4.24$.

Similarly, we examined the mean ratings from the mean estimation task on the four stereotypic and four counterstereotypic items. Table 3 again shows the predicted out-group homogeneity effect, such that ratings by the out-group were on average higher for the stereotypic items than those by the in-group and lower for the counterstereotypic items than those by the in-group, $F(1, 57) = 7.11$ (note 1 subject had missing data for this measure). Finally, examining the mean ratings from the dot task (i.e., the calculated mean of the frequency distribution)

Table 2
Intercorrelations Among Measures of Knowledge Organization for In-Group and Out-Group

Group	Familiarity	Redundancy	Attributes	Deviance
In-group				
No. of subgroups	.154	−.313**	.089	.147
Familiarity		.107	−.020	.006
Redundancy			−.088	−.256*
No. of attributes				.243*
Subgroup deviance				—
Out-group				
No. of subgroups	.189	−.295**	−.098	.103
Familiarity		−.026	.098	−.179
Redundancy			−.019	−.076
No. of attributes				−.157
Subgroup deviance				—

* $p < .10$. ** $p < .05$.

Table 3
Out-Group Homogeneity Effects

Measure	In-group	Out-group
Dispersion		
Range	14.39	13.60
Stereotypicality		
Percentage estimates		
S	62.40	65.33
CS	35.31	29.53
Dot task		
S	12.72	13.21
CS	7.68	6.53
Mean estimate task		
S	5.06	5.19
CS	4.16	3.84

Note. S = stereotypic items; CS = counterstereotypic items.

revealed that ratings by the out-group again showed greater stereotyping. That is, the mean judgments by the out-group were higher on stereotypic items relative to the in-groups' judgments, and they were lower on counterstereotypic items, $F(1, 58) = 8.47$.

On the whole then, there is clear evidence of out-group homogeneity in these data. The three measures of perceived stereotypicality all showed the predicted effect. One of the perceived dispersion measures showed the effect (range) and one did not (standard deviation from the dot task). Although quite consistent, the magnitude of the out-group homogeneity effect in the present study was smaller than that in previous studies using the same groups (i.e., Judd et al., 1991; Park & Judd, 1990). An interesting speculation concerning this difference is that the open-ended descriptions used in this study may have caused subjects to appreciate the diversity in the out-group. In the present study, subjects were explicitly asked and encouraged to think not only about the entire group but also about subgroups and differences and similarities among the subgroups. This procedure may have focused subjects on variability within the groups, reducing the magnitude of the out-group homogeneity effect. At any rate, the effect, though weakened, was clearly present in these data, and our next set of analyses focused on the relation between measures of knowledge organization and the out-group homogeneity effect. The standard deviation from the dot task was excluded because it was impossible to test for mediation of out-group homogeneity given it was not obtained on this measure.

Mediation of Out-Group Homogeneity

To explore mediation of the out-group homogeneity effect by the measures of knowledge organization, we first computed the correlation between each of the measures of perceived variability and each of the measures of knowledge organization for the in-group and the out-group. These correlations appear in Table 4. Note that these are the average within subject group correlations (i.e., for the in-group, the average of the correlation for engineering majors rating engineers and the correlation for business majors rating business majors), to eliminate spurious results from differences in the ratings of the different subject groups.

All of the measures of knowledge organization except for redundancy were scored in such a way that higher numbers should lead to greater perceived variability (e.g., a greater number of subgroups, greater familiarity, and so forth). For redundancy, a smaller amount of redundancy should be related to greater perceived variability. For the variability measures, larger ranges correspond to greater perceived variability. Thus, the correlations with range should all be positive except that for redundancy, which should be negative. All three stereotypicality measures reflect the magnitude of the extremity or stereotypicality in subjects' ratings. The scores are computed by subtracting estimates on the counterstereotypic items from those on the stereotypic items, such that larger scores reflect greater perceived stereotypicality. Thus, the correlations with the stereotypicality measures should all be negative except those for redundancy, which should be positive.

We also computed the correlation between the magnitude of out-group homogeneity demonstrated by a given subject and each of the knowledge organization measures. To compute the measure of out-group homogeneity, we first subtracted the mean target group rating on each variability measure from the appropriate target group (this controls for target group differences). We then subtracted variability estimates for the out-group from those for the in-group. A similar procedure was followed for the knowledge organization measures. These correlations also appear in Table 4, and again these are average within subject group correlations.

For the range, a positive correlation is expected for all variables except redundancy. Such a correlation indicates that, for example, the larger the difference in how dispersed the in-group is perceived relative to the out-group (i.e., a large out-group homogeneity effect), the larger the difference in number of subgroups generated for the in-group relative to the out-group. This correlation should be negative for redundancy. For the stereotypicality measures, negative correlations are again expected for all variables except redundancy. A negative correlation indicates, for example, that the more negative the difference between stereotypic perceptions of the in-group relative to the out-group (i.e., the out-group is perceived as much more stereotypic than the in-group, or a large out-group homogeneity effect), the greater the number of subgroups generated for the in-group relative to the out-group. For redundancy, the stereotypicality measures should be positively correlated with greater levels of redundancy.

The simple correlations for the in-group show very few relations to the knowledge organization measures. For the out-group, redundancy clearly predicts perceived variability. When the subgroups generated were relatively nonredundant, the out-group was perceived to have a large range. The more redundant the subgroups, the more stereotypic the out-group was perceived to be. Rated subgroup deviance showed a similar pattern (the more deviant, the greater the range, etc.) albeit less strongly.

We were most interested in the correlations with out-group homogeneity because the primary purpose of this work was to determine whether subgroup representation mediates the out-group homogeneity effect. Here, differences in the number of

Table 4
Correlations of Measures of Knowledge Organization With Perceived Variability

Measure	No. of subgroups	Familiarity	Redundancy	No. of attributes	Subgroup deviance
In-group					
Range	.17	.21	−.07	.15	.14
Percentage estimates	−.19	.19	.10	.24*	.03
M estimates	−.17	.19	.06	.18	−.01
M from Dot task	−.05	.21	−.02	.09	−.07
Out-group					
Range	.45**	.20	−.42**	−.03	.22*
Percentage estimates	.00	−.04	.29*	.13	−.24*
M estimates	.04	.01	.27*	.05	−.19
M from Dot task	−.13	−.33**	.32**	.11	−.13
Out-group homogeneity					
Range	.30**	−.03	−.24	.10	.31**
Percentage estimates	−.45**	.12	.21	.25*	−.23*
M estimates	−.42**	.09	.14	.26**	−.18
M from Dot task	−.45**	−.12	.34**	.11	−.28**

* $p < .10$. ** $p < .05$.

subgroups generated for the in-group relative to the out-group clearly predicted the magnitude of the out-group homogeneity effect. As the number of subgroups generated for the in-group grew large in comparison to that for the out-group, the in-group was seen as more dispersed than the out-group and less stereotypic than the out-group. Subgroup deviance showed a consistent and relatively strong similar pattern. Redundancy also showed a consistent similar pattern (i.e., the more redundant the in-group was perceived relative to the out-group, the more stereotypic the in-group was seen relative to the out-group, or reverse out-group homogeneity), although it was much weaker. Familiarity clearly showed no relation whatsoever to the level of perceived variability for each group or to the magnitude of differences in perceived variability. In summary, knowledge about the in-group differed from that for the out-group in many ways. Specifically, subjects reported greater familiarity with the in-group, generated a greater number of subgroups for the in-group, generated a greater number of attributes for the in-group, and described the subgroups on average as more deviant from the superordinate stereotype for the in-group. Of these variables, the difference in the number of subgroups was most highly and consistently correlated with differences in perceived variability for the in-group and the out-group. Subgroup deviance and redundancy also predicted differences in perceived variability to some extent.[1]

To argue that an effect is mediated by some variable, it is necessary to demonstrate both that the hypothesized mediator is correlated with the effect and that when differences on the mediation variable are controlled, the effect of interest is no longer reliable (Baron & Kenny, 1986; Judd & Kenny, 1981). The results of tests of the first of these two conditions were presented in the preceding paragraphs and in Table 4. The results of tests of the second condition are presented in Table 5, which includes the *F* value associated with the test of out-group homogeneity for each of the four measures of variability on which we initially obtained the effect, first with no mediators controlled and then with each of the five hypothesized mediators of interest controlled, one at a time. If, as we have suggested, the greater tendency to organize information into subgroups for the in-group than for the out-group (indexed here by the number of subgroups listed) leads to greater perceived variability for the in-group relative to the out-group, then when the difference in the number of subgroups is controlled for, out-group homogeneity should disappear.

When the difference in the number of subgroups was included, the test of out-group homogeneity was no longer reliable for any of the four measures of perceived variability. Thus, the difference in the number of subgroups was significantly correlated with differences in perceived variability on all four measures of variability, and when the difference in the number of subgroups was controlled, the out-group homogeneity effect was no longer present on any of the four measures. None of the other mediators came close to this level of consistency in meeting both conditions specified by Judd and Kenny (1981). Although differences in subgroup deviance showed strong correlations with differences in perceived variability, when this variable was controlled for out-group homogeneity was still present for three of the four measures.

Redundancy also showed a weak pattern of mediation. In only one case was it reliably correlated with differences in perceived variability (stereotypicality from the dot task), but controlling for this difference did not eliminate out-group homogeneity. Differences in familiarity were not correlated with any of the measures of perceived variability. Controlling for differences in familiarity did eliminate out-group homogeneity on

[1] Although it is not possible to look at mediation of out-group homogeneity on the standard deviations from the dot task (because out-group homogeneity was not present here), it is possible to look just at the correlation between in-group and out-group differences in the standard deviation from the dot task and the set of hypothesized mediators. We examined these correlations and found that none of the mediators were significantly related to differences on the standard deviations from the dot task.

Table 5
F Values for Out-Group Homogeneity Effects Controlling for Measures of Knowledge Organization

Covariate	Variability measure			
	Range	Percentage estimates	*M* estimates	*M* from Dot task
None	4.15**	5.42**	7.11**	8.47**
No. of subgroups	0.12	0.01	0.12	0.19
Familiarity	0.07	3.31*	3.03*	0.09
Redundancy	0.98	1.45	2.98*	3.28*
No. of attributes	2.83*	3.19**	10.30**	9.10**
Subgroup deviance	2.39	3.70*	5.53**	6.05**

* $p < .10$. ** $p < .05$.

two of the four measures, but because the first condition of mediation (viz., correlation with the dependent variable) was not met, familiarity was not a strong candidate for mediation even for these two dependent variables. Note that where reliable *F* values remain in Table 5, they are always in the direction of out-group homogeneity (as opposed to a reversal of the effect). Also, we examined the homogeneity of regression assumption and found that this assumption was met in all of the analyses depicted in Table 5.

Because familiarity has most frequently been suggested as the cause of differences in perceived variability, we performed the following competitive test between familiarity and number of subgroups as mediators of out-group homogeneity. Recall that in the analyses presented in Table 5, each hypothesized mediator was included by itself, with no other mediators in the equation. In the competitive analyses, we looked at differences for the in-group and the out-group on each of the four measures of perceived variability when simultaneously controlling for differences in the number of subgroups and familiarity. In all four of the analyses, out-group homogeneity was eliminated when these two hypothesized mediators were included, largest $F(1, 47) = .44$. Importantly, number of subgroups was reliably related to perceived variability in all four analyses, all Fs$(1, 47) > 4.26$. Familiarity, in contrast, was not related to perceived variability in any of the analyses, all Fs$(1, 47) < .82$, $p = .40$. The respective proportion of variance accounted for by familiarity and number of subgroups was 0% and 8% for the range, 2% and 14% for stereotypicality from the dot task, 1% and 15% for stereotypicality from the rated-means task, and 2% and 19% for stereotypicality from the percentage estimate task. Thus, differences in the number of subgroups both significantly predicted differences in perceived variability, and when controlled, out-group homogeneity was no longer reliable on any of the four measures. Although other variables showed some evidence of relation to perceived variability, none showed the level of consistency demonstrated by number of subgroups.

We should note that as Baron and Kenny (1986) and Judd and Kenny (1981) discussed, the ability of a variable to mediate an effect is dependent on the degree to which that variable is reliably measured. It may be that the superiority shown by number of subgroups was due in part to that variable having been measured with a higher degree of reliability. The only hypothesized mediator for which we have an estimate of reliability is familiarity. Cronbach's alpha for the measures used to assess familiarity was .80. We cannot conclusively say what role reliability may have played in our conclusions, but at least for familiarity, the reliability was respectable. It seems unlikely that the reliability of the number of subgroups would be much higher.

Judgments of the Self and Known Instances

Finally, because of our previous work (e.g., Park & Judd, 1990), we explored the relation among judgments of the self, judgments of known instances, and perceived variability. Specifically, for judgments of known instances, we calculated both a standard deviation and a range from the ratings of the four in-group and the four out-group members. Consistent with Park and Judd (1990), although the difference in perceived variability for the in-group and out-group instances was highly correlated with differences in the perceived range of the in-group and out-group, in general, when these variables were controlled for, the out-group homogeneity effect remained. Similarly, a measure of the judged stereotypicality of the instances was calculated, and again the in-group–out-group difference was highly correlated with in-group–out-group differences in perceived stereotypicality, but the latter difference remained highly reliable when controlling for the former. For the self, measures of discrepancy between self-ratings and ratings of both the in-group and the out-group were calculated. Larger self-discrepancies were clearly correlated with greater perceived variability for the in-group, whereas this was not the case for the out-group. Consistent with the results of Park and Judd (1990), judgments of specific instances and the self appear to parallel judgments of the group as a whole, but there is no evidence that these differences mediate the out-group homogeneity effect.

Discussion

On the whole, the results suggest that the strongest mediator assessed in the present study of out-group homogeneity is the difference in the simple number of subgroups generated by the subject for the in-group and the out-group. Subjects listed a significantly greater number of subgroups for the in-group than

for the out-group. The difference in the number of subgroups generated for the in-group relative to the out-group significantly predicted the magnitude of the out-group homogeneity effect for all four measures on which out-group homogeneity was obtained. Moreover, when the difference in the number of subgroups was controlled for, out-group homogeneity disappeared on all four measures.

Other hypothesized mediators fared less well. Differences in both the rated deviance of the subgroups and the redundancy in the subgroups showed mixed relations to the magnitude of out-group homogeneity and similarly mixed results in terms of accounting for the effect. These three measures, the number of subgroups, the deviance of the subgroups from the larger group, and the redundancy in the subgroups certainly could be related to some extent. Number of subgroups may be a less refined measure of the amount of redundancy or similarity among the subgroups, and average deviance of each subgroup from the larger group should be somewhat related to degree of redundancy among the subgroups. At least in these data, however, the less refined measure of number of subgroups appears to provide a better account of differences in perceived variability. Differences in familiarity clearly were not able to predict differences in perceived variability, either for each group alone or when looking at differences in perceived variability for the in-group and the out-group.

These data suggest a rather different conceptualization of the process by which perceptions of variability within a group are determined from those offered in the existing literature. Linville (1982; Linville et al., 1989) has argued that familiarity plays a central role in perceived variability. Judd and Park (1988; Park & Judd, 1990; Park et al., 1991) have argued that encoding or retrieval biases involving the self and known instances determine perceived variability. The present data, however, suggest that subjects' knowledge of specific subgroups within the larger group play the key role in perceptions of both how dispersed the group is and the extent to which the group is seen as fitting the group stereotype. Nevertheless, these data are correlational, and it is impossible to draw strong causal conclusions about the influence of subgroup knowledge on perceived variability. Accordingly, we designed a second study in which we experimentally manipulated the presence or absence of subgroups to test whether this influenced the level of perceived variability within the group.

Experiment 2

In this experiment, we wanted either to induce subjects to develop subgroups within a larger group or to discourage subjects from doing this. We reasoned that a design in which the groups of subjects learned identical information, and varied only in whether subgrouping was present, would provide the strongest test of the hypothesis that the presence of subgroups results in increased perceived variability and therefore accounts for out-group homogeneity. Accordingly, we presented subjects with fictitious information about members of a group. The individuals were allegedly volunteers in the Big Brother program. The alleged group consisted of a total of 100 members, but subjects learned about only 16 of these.

Subjects learned five facts about each individual, and each fact was related to a different dimension. On two of the five dimensions, all 16 members were described at the same end of the continuum (i.e., all 16 performed a helpful behavior, and all were politically liberal). On the other three dimensions, three individuals were described with facts that were inconsistent with the rest of the group. Thus, although the group in general was socially competent, academically competent, and had a relaxed and somewhat permissive upbringing, a different subset of individuals was inconsistent on each of these dimensions, and each subset could thus be seen as a subgroup within the larger group.

Subjects were either asked to simply read and learn about the 16 volunteers in the Big Brother program or they were asked to sort the 16 individuals as they read about them, grouping together those who were similar in some way and different from the individuals in the other subgroups. Our hypothesis was that subjects who performed the sorting task would come to see the group as more variable than those who did not.

Method

Subjects

Eighty undergraduates at the University of Colorado served as subjects in exchange for credit toward a course requirement. Forty subjects were randomly assigned to each experimental condition (sort vs. nonsort). Subjects participated in groups of 6 or fewer. The experimental manipulation required subjects in the sorting condition to engage in a visibly different task (physically sorting index cards) from that of subjects in the nonsort condition (who were simply asked to read and think about the information on the index cards). We therefore tested subjects in the two conditions in separate sessions.

Stimulus Materials

Descriptions were constructed to characterize 16 individuals who ostensibly belonged to a larger group of volunteers in the Big Brother program. Each description consisted of five facts intended to characterize group members with respect to five dimensions: helpfulness, political orientation, social competence, academic competence, and parental style of upbringing (relaxed vs. strict). To give cohesion to the group, all 16 group members were characterized as both helpful and liberal, and behaviors from these two dimensions were always given first in the description. The order of presentation of behaviors from the remaining three dimensions was counterbalanced across individuals within a subgroup. With respect to the other three dimensions, 13 of the 16 facts indicated social competence, academic competence, and a relaxed family upbringing. Three of the 16 facts indicated social incompetence, academic incompetence, and a strict upbringing. The facts on these three dimensions were organized in such a way that the 16 descriptions of the individuals could be perceived as forming five types or subgroups of the larger group of volunteers.

Pretest results suggested that the most effective way to induce perceived subgroups was to vary the valence of the facts on a dimension (i.e., academically competent vs. incompetent), and to group the 13 consistent facts such that individuals within a subgroup confirmed the group stereotype in a similar way—in a way that was different from the other subgroups. That is, there are different ways in which social competence might manifest itself, and each subgroup was paired with one of these. For example, one subgroup was characterized as consisting of social leaders who organized activities and outings, a second as socially popular individuals whom others liked to be with, a third as socially

successful with women, a fourth as socially wild and rowdy, and a fifth as socially incompetent and shy.

Four individuals formed the first subgroup. These individuals were characterized as consistent with the larger group of volunteers. That is, like the larger group of volunteers, these individuals were all academically and socially competent and from relaxed family backgrounds. More specifically, these individuals were all social leaders and academic leaders and had parents who supported and confided in them. One member of this subgroup was described as follows:

Chuck Foster
Visited the elderly in a nearby home at Christmas
Volunteered with the Head Start program for poor children
Was voted captain of his intramural football team
In high school, was able to tell his parents anything that was on his mind
Has made the Dean's honor roll every semester in college

The second subgroup consisted of 3 individuals who also confirmed the group stereotype on all three dimensions, but in a manner very different from Subgroup 1. These individuals were all socially popular (e.g., was voted big man on campus by his friends), academic achievers in the professions (e.g., has been accepted into 5 top M.B.A. programs), and had parents who were economically permissive (e.g., parents gave him a VISA gold card to use as he needed in high school).

The three remaining subgroups each consisted of 3 individuals who disconfirmed the larger group stereotype on one of three trait dimensions. Thus, the members of Subgroup 3 were all from strict families (e.g., was not allowed to date in high school except for school functions, such as dances), were socially successful with women (e.g., has a steady girlfriend who is quite attractive), and were academic helpers (e.g., is currently an academic peer counselor). The members of Subgroup 4 were all socially awkward and shy (e.g., sat in a corner alone at the Christmas party), earned high grades in math and science courses (e.g., earned straight *A*s last term in five engineering courses), and had parents who were uninvolved with their children (e.g., his parents traveled a great deal so he was often home alone while growing up). The members of Subgroup 5 were all academically irresponsible (e.g., arranged to take a final exam late and then failed to show up for it), socially wild (e.g., went out drinking with friends and ended up dancing on the tables), and came from very permissive households (e.g., had no curfew during high school and often stayed out all night). The name and five facts for each stimulus person were typed on a 5 × 8 in. (12.7 cm × 20.32 cm) index card.

Procedure

Subjects in both the sort and the nonsort conditions were told that they would be given information about 16 college men who were randomly selected from a larger group of 100 volunteers of the Big Brother program and that, after reading this information, they would be asked to give their impressions of the group as a whole. We felt it was important first to establish an overall group stereotype so that subjects would have some basis for forming subgroups, that is, for determining whether there were types of group members who deviated from the overall group. Thus, before presenting information about the 16 group members, we first asked subjects to read four "summary evaluations" of the group that had ostensibly been prepared by social service personnel who had interviewed the volunteers. For example, one of the summaries stated that "this group of volunteers tends to be politically active . . . successful both socially and academically. These kids appear to be from good family backgrounds, which should aid them in their task of providing supportive role models." The other three summary descriptions provided similar information, worded in different ways.

Subjects were then asked to write a short description of their initial impression of the group based on just the summary evaluations, again to help set the initial stereotype. Next subjects were told that they would be given a set of 16 index cards, each of which listed five behaviors provided by a specific group member as part of the in-depth application process to become a volunteer. These behaviors were to be considered representative of what that person was like. Subjects in the sort condition were asked, as they read about the 16 individuals, to sort them into as many groups as they wished, such that all the people in one group were similar to one another in some way and different from the people in the other groups. Subjects in the nonsort condition were told to read through the cards carefully and use the information to further develop their impression of what the group was like as a whole.

Subjects in both conditions were instructed to use the full 10 min provided to perform their task and to develop their impression of the group. The index cards were ordered such that the first 5 cards consisted of 1 member from each of the 5 subgroups so that subjects could see the full range of individuals very early in the presentation of information. The remaining 11 cards were presented in a random order. Each subject received an identical set of 16 cards. Because subjects in both conditions received identical information about the group and its members, the actual variability of the group members was the same in both conditions. The only difference between the two conditions was the identification of subgroups by subjects in the sort condition. Thus, any differences in perceived variability must be a result of this.

Following the presentation of the 16 group members, subjects completed a variety of pencil-and-paper measures. In the sort condition, subjects were asked to record the subgroups they had formed by writing down for each subgroup a summary name or title and listing the individuals in the group. At this point, the experimenter collected all of the cards. Subjects were then asked to refer to the subgroup titles they had identified and for each one, write a half page description of what the subgroup was like and how it was different from the other subgroups. Subjects in the nonsort condition were asked simply to write a description based on their impression of the group in general, as if describing the group to a close friend.

All subjects then completed four tasks to assess their perceptions of group variability with respect to the three critical trait dimensions (strict vs. relaxed upbringing, socially competent vs. socially incompetent, and academically competent vs. academically incompetent) and one global measure of perceived variability. Subjects were instructed to base their responses on their impression of the total group of 100 volunteers using what they had learned about the 16 group members. First, subjects rated the global similarity of group members to one another, and the similarity of group members to one another on each of the four dimensions, using a 7-point scale ranging from *extremely dissimilar* (1) to *extremely similar* (7). Subjects then estimated the percentage of the group they believed would agree with or endorse each of 12 attitude statements, 4 of which were relevant to each of the three critical trait dimensions. Within these 4, 2 were consistent with the group stereotype (e.g., socially competent) and 2 were inconsistent with the group stereotype (e.g., socially incompetent). Thus, a total of 6 statements were stereotypic of the group (e.g., I pride myself on my academic achievements), and 6 were counterstereotypic (e.g., I find myself struggling to perform adequately in college). Next, subjects completed the dot task and the mean and range estimation task (both described in Experiment 1) for each of the three dimensions.

We were concerned about differences that the sorting manipulation might produce in depth of processing of the presented information. To examine this, subjects were next asked to complete a recall task in which they were to list, in any order, as many bits of information as they could remember about the 16 individuals. Subjects were told that they did not need to identify which of the 16 individuals had performed a particular behavior, and recall did not need to be verbatim, but rather they were to write any information they could recall.

Results

Perceived Variability

The mean values for all of the perceived variability measures are presented for the sort and nonsort conditions in Table 6. We first looked at differences in the global similarity ratings and determined that nonsort subjects saw group members as reliably more similar to one another ($M = 4.63$) than sort subjects ($M = 4.00$), $F(1, 78) = 5.15$ (higher numbers indicate greater perceived similarity). For the remaining four measures of perceived variability, we averaged across trait dimension before performing the analyses because we had no a priori reason to expect differences in the effect of the manipulation for the various trait dimensions. As is clear from Table 6, the pattern of differences in variability judgments for the sort and nonsort conditions was very consistent across the trait dimensions. For the similarity judgments on specific trait dimensions, subjects in the nonsort condition again perceived group members as more similar ($M = 4.72$) than sort subjects ($M = 4.00$), $F(1, 78) = 9.65$. A standard deviation was calculated from the dot task (as in Experiment 1). The average standard deviation, across trait dimensions, tended to be larger in the sort ($M = 2.11$) than the nonsort condition ($M = 1.98$), although this difference was only marginally reliable, $F(1, 78) = 3.07$, $p = .08$.

Pilot testing led us to be skeptical that differences would emerge on the range task as a function of the sorting manipulation. Recall that in this task, subjects indicate where the least and most extreme member of the group falls on the attribute dimensions. In this particular experiment, pretest subjects tended to use specific individuals they could remember to complete the range task, and given that these were the same for sort and nonsort subjects, we worried that differences as a function of the sorting manipulation might not be present. As we suspected, the difference between the two conditions on the range task was not reliable, although it was in the direction consistent with the remaining measures of perceived variability, $F(1, 77) = 2.24$, $p = .14$ (sort $M = 18.90$ and nonsort $M = 17.97$).

To analyze the percentage estimate task, judgments for the six statements written to be stereotypic of the group were averaged together to form an estimate of the prevalence of stereotypic attributes. Similarly, judgments for the six statements written to be counterstereotypic of the group were also averaged into a judgment of the prevalence of counterstereotypic attributes. We compared the magnitude of the estimates for stereotypic versus counterstereotypic items as a function of the sort manipulation. The interaction between these two variables was reliable, $F(1, 78) = 7.06$, such that subjects in the nonsort group estimated a higher prevalence of stereotypic attributes relative to sort subjects ($M = 62.63$ vs. 52.87) and a lower prevalence of counterstereotypic attributes ($M = 26.90$ vs. 27.97). That is, nonsort subjects judged the group to be more stereotypic than subjects in the sort condition.

Table 6
Differences in Perceived Variability for Sort and Nonsort Conditions (Experiment 2)

Measure	Sort	Nonsort
Global similarity	4.00	4.63
Social similarity	4.18	4.70
Academic similarity	4.63	5.08
Upbringing similarity	3.18	4.38
Social *SD*	2.06	1.92
Academic *SD*	2.08	1.92
Upbringing *SD*	2.18	2.11
Social competence		
S	60.20	67.58
CS	24.03	21.54
Academic competence		
S	53.89	64.00
CS	23.53	20.28
Upbringing		
S	44.51	56.31
CS	36.36	38.95

Note. S = stereotypic items; CS = counterstereotypic items.

Table 7
Recall and Clustering (Experiment 2)

Measure	Sort	Nonsort
Recall		
Consistent	12.15	14.53
Inconsistent	3.88	4.98
Other	4.65	6.05
Total	20.68	25.60
ARC		
Subgroup	0.211	0.104
Trait dimension	0.352	0.321
Person	0.033	0.008

Note. ARC = adjusted ratio of clustering.

Recall and Clustering of Presented Facts

Subjects in the sort condition clearly perceived the group as more variable than subjects in the nonsort condition. To test whether this may have been due simply to better memory for facts about the group rather than to the formation of subgroups per se, we examined the number of items recalled by subjects in the two conditions. Subject recall was coded for gist. That is, recall did not need to be verbatim, but it was essential that the coder be able to identify which of the presented items was being recalled to score the item as correct. For example, for the item "was voted Captain of his intramural football team," a response such as "was head of his football team" would be scored as correct. "Played football" would not be correct. We analyzed the mean number of items recalled as a function of the sorting manipulation. As is clear from the means presented in Table 7, subjects in the nonsort condition actually recalled more total facts ($M = 25.60$) than subjects in the sort conditions ($M = 20.68$), $F(1, 78) = 6.03$.

To further explore the question of whether subjects in the two conditions had differential memory for the behaviors, we broke the total recall scores into recall for three specific item types. Namely, we calculated the total number of items recalled that were consistent with the overall group behavior. This included the 13 socially competent, academically competent, and re-

laxed upbringing items, for a total of 39 possible. We similarly calculated the total number of inconsistent items recalled, that is, the three socially incompetent, academically incompetent, and strict behaviors, for a total of 9 possible items. Finally, the total number of helpful and liberal behaviors was tallied (32 total possible). We wondered whether the overall difference in recall for the two conditions was true for all item types or only for some. In particular we worried that sort subjects might actually have better recall for the inconsistent items, and this may have produced judgments of greater variability in this condition. Accordingly, we examined the sort–nonsort difference separately for all three item types (the relevant means appear in Table 7). Nonsort subjects recalled significantly more inconsistent, $F(1, 78) = 6.81$, and consistent items, $F(1, 78) = 4.18$, and the difference for the helpful/liberal items was marginally reliable, $F(1, 78) = 3.29$, $p = .07$.

Thus, the tendency to see greater variability among group members in the sort condition was clearly not due simply to paying better attention to the information. Nonsort subjects recalled a greater number of the facts presented, and this was true of the inconsistent as well as the consistent facts. This better recall may have occurred because subjects in the sort condition were primarily occupied with comparing the similarities and dissimilarities among targets, whereas nonsort subjects spent their time studying the specific items of information. In the sort condition information may have been encoded as certain "types" of items, with a loss of the particulars of any given item. Clearly, it appears that it was the process of mentally organizing individuals into coherent subgroups that led to the perception of greater heterogeneity in the sort condition rather than simply increasing attention to this information.

Finally, we were interested in the extent to which subjects appeared to use the 5 subgroups we had a priori identified in the stimulus materials to organize presented information. (Interestingly, the average number of categories sort subjects used during the sorting task was 4.90, very close to the number of a priori defined subgroups.) As an index of this, we computed clustering scores from the recall data using the adjusted-ratio-of-clustering (ARC) score (Roenker, Thompson, & Brown, 1971). ARC scores reflect the extent to which the order of output in a given subject's recall protocol reflects a particular organization. We computed three ARC scores: one using the 5 subgroups as the basis for organization, one using the five trait dimensions, and one using the individuals or persons in the stimulus set. For example, in the subgroup-clustering computation, any time two facts were sequentially recalled that were performed by individuals in the same subgroup (including the same individual), this was counted as a repetition.

For the trait dimension computation, any sequentially recalled facts from the same trait dimension (e.g., two helpful facts) were counted as a repetition. For the person computation, any two sequentially recalled facts that were performed by the same person were counted as a repetition. Note that subjects were not required to identify the target associated with each of the recalled facts. The clustering analyses were based on which target was actually paired with a given fact. ARC scores vary from +1.00, reflecting perfect organization by the experimenter defined dimension, to 0, reflecting chance levels of organization in terms of this particular dimension, to −1.00, reflecting a strong tendency to organize by some dimension other than that defined by the experimenter.

The average ARC scores using each of the three organizing schemes discussed earlier are presented by condition in Table 7. The test of whether reliable clustering was present along each organizing scheme across condition indicated that both the average subgroup and trait dimension ARC scores were significantly different from zero, $F(1, 77) = 32.41$ and $F(1, 77) = 83.34$, respectively, whereas this was not true for the person ARC scores ($F < 1$). Thus, subjects appeared to use both the subgroups and trait dimensions to organize recall output. In addition, sort subjects tended to use the subgroup organization scheme to a greater extent than nonsort subjects, $F(1, 77) = 3.71$, $p < .06$. No such difference was present for the trait dimension ARC scores ($F < 1$) or for the person ARC scores ($F < 1$).

The recall data are informative in two respects. First, they suggest that the sort manipulation resulted in different interconnections among the presented information for sort than nonsort subjects. Although both sort and nonsort subjects organized recall by trait dimension, subjects in the sort condition were more likely to recall together items performed by individuals within the same a priori defined subgroups, compared with subjects in the nonsort condition. The manipulation did not produce simply greater clustering by target person but, rather, clustering by individuals within a subgroup. Thus, the manipulation effectively altered the organization by which information about the group was recalled. At the same time, the manipulation did not simply produce greater attention or greater memory for the information. Subjects in the nonsort condition actually demonstrated better recall for the presented information than subjects in the sort condition.

Discussion

In this study, we presented subjects with identical information about members of a group. One group of subjects simply read about the individual group members, whereas the remaining subjects actively identified similarities and differences among subsets of the larger group. The effect of forcing subjects to recognize meaningful subgroups within the larger group resulted in perceptions of greater variability on four of the five variability measures used (albeit marginally on the dot standard deviation measure). This finding is all the more remarkable in that nonsort subjects actually recalled a greater number of facts about group members, both those consistent and inconsistent with the larger group stereotype. Both sort and nonsort subjects demonstrated organization of output at recall by the 5 subgroups and by the trait dimensions. Organization by subgroup tended to be stronger, however, in the sort condition. On the whole, these data present strong evidence for the importance of organizing social category information into subgroups for increased levels of perceived variability.

General Discussion

The primary goal of this research was to investigate differences in the organization of knowledge for in-groups and outgroups and to assess the importance of such differences in pro-

ducing the out-group homogeneity effect. In Experiment 1 we explored five aspects of knowledge representation: number of subgroups spontaneously generated when asked to describe such subgroups, average rated deviance of descriptions of these subgroups from the description of the larger group, degree of redundancy in perceptions of the subgroups (calculated both as the average correlation across attribute ratings between all possible pairs of subgroups and as the average correlation across subgroup ratings between all possible pairs of attributes), total number of attributes used to describe the group as a whole and the subgroups, and familiarity with the group. Four of these measures showed reliable in-group–out-group differences (rated deviance only marginally, so $p = .10$). The difference for the redundancy measure was not reliable, although it was in the expected direction.

Subsequent analyses of the role of each of these variables as potential mediators of out-group homogeneity suggested that the number of subgroups resulted in the most consistent and convincing mediational pattern. Specifically, differences in the number of subgroups generated for the in-group versus the out-group were significantly correlated with in-group–out-group differences in perceived variability on all four measures for which out-group homogeneity was obtained. Moreover, controlling for differences in the number of subgroups eliminated the out-group homogeneity effect, again on all four measures for which the effect was initially reliable. The remaining measures of knowledge organization showed much weaker patterns of mediation, and none came close to the level of consistency demonstrated by the number of subgroups variable.

Of particular interest were the results for familiarity. Intuitively, it seems that increased familiarity should result in increased perceived variability (Linville et al., 1989). Jones et al. (1981) found no evidence for this hypothesis, and their results are soundly confirmed in the present research. Familiarity did not predict perceived variability for the in-group or out-group, nor did differences in perceived familiarity predict differences in perceived variability. Taylor (1981) suggested that increases in familiarity should lead to increased subtyping. In our data, this hypothesis was not supported. The correlation between familiarity and number of subgroups was low and nonsignificant for both the in-group and the out-group.

The hypothesis that organizing information into meaningful subgroups produces greater perceived variability was further investigated in Experiment 2. All subjects in this experiment were presented with identical information about 16 individuals of a hypothetical group. However, the task of sorting information about the group members into meaningful subgroups resulted in perceptions of greater variability among the group members than when the sorting task was not performed. The order of recall of these subjects further suggested that the a priori defined subgroups were used in organizing information about the group. Importantly, the effect was not due simply to greater attention in the sort condition. Subjects in the nonsort condition actually remembered more of the information about the group members, and this was true for all types of information presented, including group-inconsistent items.

We should note that in Experiment 1, number of subgroups was related to perceived variability only when making relative comparisons between the in-group and the out-group. That is, the simple correlations between number of subgroups and perceived variability within the in-group and out-group were not reliable. It was the difference in the number of subgroups for in-group and out-group that predicted difference in the perceived variability of in-groups and out-groups. In Experiment 2, we argued that the presence of subgroups led to greater perceived variability, a claim somewhat at odds with the results of Experiment 1. Although we cannot say for certain, we suspect that the discrepancy in the two studies derives from the greater role of individual differences in Experiment 1. Specifically, in Experiment 1 there was a great deal of variability in the number of subgroups generated. Some subjects listed a large number, some a small number, and the number of subgroups generated for the in-group was highly correlated with the number for the out-group ($r = .50$). Factors such as motivation, mood, and wordiness all probably influenced the number of subgroups listed by a given subject for both the in-group and the out-group in this free elicitation task. When we examine differences in the number of subgroups and their relation to the difference in perceived variability, this essentially utilizes a within-subjects rather than a between-subjects design in which variance due to these individual differences can be controlled. In Experiment 2 the subgrouping task was rigorously controlled by the manipulation and the construction of the stimulus materials. We were careful to embed five clear subgroups in the materials, and the average number of groups formed by subjects was 4.70. In addition, subjects either had no subgroups present or close to these five subgroups. Consequently, the degree of within-subject variability in the number of subgroups formed in Experiment 2 was greatly minimized as compared with Experiment 1. Our theoretical argument clearly suggests that the development of subgroups should be related to perceived variability (not just in a relative sense). However, given the nature of this dependent variable (viz., free elicitation), individual differences need to be taken into account.

Why is it that knowing about a greater number of subgroups leads to greater perceived variability? We suspect that knowledge about subgroups plays a central role at the time subjects are asked to judge group variability. In our earlier research, we developed the working hypothesis that subjects keep essentially a running estimate of the level of perceived variability within a group, along the lines suggested by Fried and Holyoak (1984). A number of events have led us away from that hypothesis. Specifically, the consistent failure to find evidence for the mediation of out-group homogeneity by retrieved instances, the self, or both have led us to doubt that these might be mechanisms by which out-group homogeneity is produced (see Park et al., 1991, for a more complete discussion of this hypothesis). In addition, a series of more process-oriented studies in our laboratory has led us to the current working hypothesis that subjects organize knowledge about a group into essentially a frequency distribution by keeping track of the relative numbers of individuals who fall at various points on a given dimension (Kraus, Ryan, Judd, Hastie, & Park, in press).

These data, in conjunction with the present results, suggest that when asked to judge, for example, how similar the members of a group are to one another, subjects think about various subgroups and their similarity to one another when knowledge about such subgroups is available. In Experiment 2,

for example, when asked how similar the 100 group members are likely to be on the dimension of social competence, subjects in the nonsort condition extrapolated from the presented information for the 16 group members. Thirteen of the 16 members were clearly socially competent, so that perceptually these subjects saw a fair degree of similarity among the group members on this attribute dimension. Sort subjects thought not only about how socially competent the 16 individuals were but also about the ways in which they were competent because this was made salient by the subgrouping task. Although these subjects also received information that the majority of the group was socially competent, they did not necessarily think about the group members in this way. Instead, these subjects thought about the subgroups and the ways in which they differ from one another on the dimension.

Although 4 of the 5 subgroups were socially competent, they were so in very different ways. One group was popular with women, one consisted of social organizers and leaders, one consisted of "party animals," and so on. So for these subjects, the overall similarity among group members appears less than for the nonsort subjects because the presence of the subgroups accentuates the ways in which the individuals are differentially socially competent. Similarly, we suggest that with natural groups, when subgroups are present, subjects use the subgroups to reason about variability among the group on a given dimension and a given judgment task. Thinking about meaningful subgroups leads to an accentuation of the perceived variability among group members relative to simply considering the aggregate of the individuals. Thus, it is the organization of the information that is crucial to determining perceived variability. Two subjects may be equally familiar or knowledgeable about a group, but if one organizes the information into subgroups and one does not (as in Experiment 2), then the former will see greater variability than the latter.

What are the conditions responsible for producing organization into subgroups? First, one can artificially induce subgrouping as we did in Experiment 2. It is interesting to reflect on the process by which we developed the subgroups for this study. Through a series of pilot investigations, it became clear that subgroups needed to be defined in a particular way. It was not sufficient to simply cut the dimension at numerous points and ask subjects to sort in this way. Thus, in one pilot we simply varied where on the dimension the subgroup was (low, moderately low, moderately high, or high) and combined dimensions in a haphazard manner. This led simply to perceptions of chaos. Instead, it seemed important to make the subgroups coherent or meaningful so that they did not necessarily fall at different points on the attribute dimension, but rather they manifested the trait in distinctly different ways. At a surface level, this corresponds to the way in which we use subgroups with natural groups. The subgroup descriptions from Experiment 1 were informative in this regard. The subgroups of engineering majors were by and large always hard working and bright but in very different ways. Some were motivated only by money, some by parental expectations, and some by larger environmental goals. There was clearly a coherence to the ways in which the subgroups differed from one another. In sum, one way to produce subgrouping is through artificial manipulation, but here it seems essential that the subgroups differ in meaningful ways.

A second factor involved in the formation of subgroups (highlighted here) is subjects' membership in the group. When one belongs to a group, subtle but meaningful distinctions become important and, we suggest, play a central role in the organization of knowledge about the group. Again note that this is due to more than simply becoming familiar with the group. Organization into subgroups derives from a motivation to understand the nuances in a group. Such circumstances are less likely to be present in the case of out-groups (see Park & Rothbart, 1982, Experiments 2 and 4). One can certainly imagine other factors besides membership in the group that might also lead to subgroup formation. In particular, if there is a differential power status to an in-group and out-group, it is entirely possible that the low-status group will organize information about both in-group and out-group into meaningful subgroups because this is functionally useful, whereas the high-status group will form subgroups only for the in-group. Others have similarly suggested this knowledge differential as a function of status differences, but to our knowledge, no one has explored the hypothesis using knowledge organization into meaningful subgroups.

Finally, we raise an apparent contradiction in the role that subgroups might play in stereotype accuracy and revision. We have conceptualized the formation of subgroups in such a way as to suggest that each substereotype provides a somewhat better fit to the members of the subgroup than the group stereotype does. *Stereotypes* are generalizations, and necessarily overgeneralizations, about members of a group. The extent to which the stereotype is an overgeneralization (i.e., it fails to reflect variation among group members) should be minimized at the subgroup relative to superordinate group level. Moreover, we have argued that increased numbers of subgroups lead to an appreciation of greater heterogeneity among group members [and that this is more likely to occur for groups to which one belongs.] All of this suggests that the tendency to organize information into subgroups is beneficial and might help ameliorate potential negative consequences resulting from the stereotyping process.

At the same time, research by Weber and Crocker (1983), and the theoretical analysis of Taylor (1981) and Rothbart and John (1985), all suggest that forming subgroups or subtypes may provide a mechanism by which disconfirming information about the group is functionally isolated and prevented from influencing perceptions of the group as a whole. Brewer and Miller (1984; Miller, Brewer, & Edwards, 1985) and Hewstone and Brown (1986; Hewstone, 1989) have similarly worried about the relative merits of adopting an intergroup versus interpersonal perspective on stereotype change.

It seems plausible that subgrouping processes might have both sorts of effects. As the differences among members of a group become important and salient, the formation of subgroups should provide a better, more highly resolved characterization of group members. At the same time, if subgroups are only used as a means for "re-fencing" discrepant group members (Allport, 1958), then greater stereotype inaccuracies may result. The former case is likely to apply to in-groups and

groups where there is a vested interest in accuracy. Interestingly, in this case the basic level (Rosch, 1981; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), or the level at which the group is thought about, is most likely the subgroup level, so that it is important for the subgroups to provide a good fit to category members. Here, the subgrouping process may be beneficial, resulting in a more highly differentiated and less overgeneralized knowledge structure. When instead there is some motivation to maintain the stereotype, the basic level of the category is likely at the superordinate level, and this is the manner in which information about the group is typically processed. In this case, the subgrouping process may serve only to isolate potentially disconfirming information from the larger group stereotype. We really do not know what the mode of mental operation is when there is no vested interest in maintaining the stereotype but, at the same time, no active pressure to refine it (as in the case of passive racism). Clearly, the topic is full of interesting directions for future research, and the general issue seems of central importance to the field of stereotyping.

References

Allport, G. W. (1958). *The nature of prejudice.* Garden City, NJ: Doubleday.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Brewer, M. B., Dull, V., & Lui, L. (1981). Perceptions of the elderly: Stereotypes as prototypes. *Journal of Personality and Social Psychology, 41,* 656–670.

Brewer, M. B., & Lui, L. (1984). Categorization of the elderly by the elderly: Effects of perceiver's category membership. *Personality and Social Psychology Bulletin, 10,* 585–595.

Brewer, M. B., & Miller, N. (1984). Beyond the contact hypothesis: Theoretical perspectives on desegregation. In N. Miller & M. Brewer (Eds.), *Groups in contact: The psychology of desegregation* (pp. 281–302). San Diego, CA: Academic Press.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 234–257.

Hewstone, M. (1989). Changing stereotypes with disconfirming information. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Strobe (Eds.), *Stereotypes and prejudice: Changing conceptions* (pp. 207–223). New York: Springer-Verlag.

Hewstone, M., & Brown, R. (1986). Contact is not enough: An intergroup perspective on the "contact hypothesis." In M. Hewstone & R. Brown (Eds.), *Contact and conflict in intergroup encounters* (pp. 1–44). Oxford, England: Basil Blackwell.

Jones, E. E., Wood, G. C., & Quattrone, G. A. (1981). Perceived variability of personal characteristics in in-groups and out-groups: The role of knowledge and evaluation. *Personality and Social Psychology Bulletin, 7,* 523–528.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5,* 602–619.

Judd, C. M., & Lusk, C. M. (1984). Knowledge structures and evaluative judgments: Effects of structural variables on judgment extremity. *Journal of Personality and Social Psychology, 46,* 1193–1207.

Judd, C. M., & Park, B. (1988). Outgroup homogeneity: Judgments of variability at the individual and group levels. *Journal of Personality and Social Psychology, 54,* 778–788.

Judd, C. M., Ryan, C. S., Park, B. (1991). Accuracy in the judgments of in-group and out-group variability. *Journal of Personality and Social Psychology, 61,* 366–379.

Kraus, S., Ryan, C. S., Judd, C. M., Hastie, R., & Park, B. (in press). Use of mental frequency distributions to represent variability among members of social categories. *Social Cognition.*

Linville, P. W. (1982). The complexity–extremity effect and age-based stereotyping. *Journal of Personality and Social Psychology, 42,* 193–211.

Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology, 57,* 165–188.

Linville, P. W., & Jones, E. E. (1980). Polarized appraisals of out-group members. *Journal of Personality and Social Psychology, 38,* 689–703.

Lusk, C. M., & Judd, C. M. (1988). Political expertise and the structural mediators of candidate evaluations. *Journal of Experimental Social Psychology, 24,* 105–126.

Miller, N., Brewer, M. B., & Edwards, K. (1985). Cooperative interaction in desegregated settings: A laboratory analogue. *Journal of Social Issues, 41,* 63–79.

Mullen, B., & Hu, L. (1989). Perceptions of ingroup and outgroup variability: A meta-analytic integration. *Basic and Applied Social Psychology, 10,* 233–252.

Park, B., & Judd, C. M. (1990). Measures and models of perceived group variability. *Journal of Personality and Social Psychology, 59,* 173–191.

Park, B., Judd, C. M., & Ryan, C. S. (1991). Social categorization and the representation of variability information. In M. Hewstone & W. Stroebe (Eds.), *European review of social psychology,* (Vol. 2, pp. 211–245). New York: Wiley.

Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology, 42,* 1051–1068.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin, 76,* 45–48.

Rosch, E. (1981). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 28–48). Hillsdale, NJ: Erlbaum.

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

Rothbart, M., & John, O. P. (1985). Social categorization and behavioral episodes: A cognitive analysis of the effects of intergroup contact. *Journal of Social Issues, 41,* 81–104.

Taylor, S. E. (1981). A categorization approach to stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 83–114). Hillsdale, NJ: Erlbaum.

Weber, R., & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality and Social Psychology, 45,* 961–977.

Received September 5, 1991
Revision received March 4, 1992
Accepted March 21, 1992 ■