

Of beasties and butterflies: Evidence for the stability and domain-specificity of individual differences in categorization

Benjamin D. Garber and Mark L. Miller, *The Pennsylvania State University*

Abstract

Previous research has failed to demonstrate unambiguously the relative stability across time and consistency across tasks of individual differences in categorization, or categorization breadth. The present study of categorizing behavior in college students assesses breadth at each of four points in time over a six-week period. Results are consistent with an interpretation of breadth as a highly stable individual difference variable, but one that is constrained by the nature of the stimulus sets employed. High stability coefficients were observed for each of two redundant tasks. In contrast, cross-task consistency indices were generally low. Personality correlates of breadth were similarly weak and inconsistent. Several variables are suggested that may attenuate the generality of categorization breadth. The existence of multiple styles of categorization is discussed.

The construct known as categorization breadth has emerged as the standard means of assessing individual differences in categorizing behaviors. In general, the breadth metric refers to the degree of variance along one or more dimensions of change that an individual is willing to assimilate within a single, homogeneous bundle. The subject's task in the categorization breadth paradigm is either to sort an array of stimuli into one of several possible categories, or to decide which of an array of stimuli can be assigned to a single established category. The former strategy has fallen prey to a number of methodological and theoretical criticisms (Kogan, 1977). The latter strategy, known as the band-width approach to categorization breadth, has been more widely accepted as the operationalization of choice.

Subjects who assign many stimuli to a category are known as broad categorizers. This label implies that, at least in regard to the particular

We would like to thank Mel Mark, Keith Nelson, Laura Landerman-Garber, David Shaffer, and two anonymous reviewers for their invaluable comments on previous versions of this manuscript. Requests for reprints should be sent to either author at the Department of Psychology, The Pennsylvania State University, University Park, PA 16802.

Journal of Personality 54:4, December 1986. Copyright © 1986 by Duke University Press.
CCC 0022-3506/86/\$1.50

category, such individuals incorporate a large degree of variance into a single bundle. In contrast, those who include relatively few stimuli into a given category are known as narrow categorizers. These individuals are hypothesized to be very restrictive in the range of variation that will be allowed within any single concept, preferring to create a great many highly refined categories rather than a few more global ones.

Pettigrew (1958) was among the first to investigate the concept of breadth, or what he termed category width. He developed a 20-item self-report inventory with which he assessed category width differences. Although he found the scale to be internally consistent ($r = .90$) and to have a fair degree of stability over a six-week interval ($r = .72$), he was unable to demonstrate hypothesized links to personality dimensions. The ecological validity of this task is also questionable.

Block and Block (1980) have proposed a two-dimensional model of the executive functions of personality which may map onto the categorization breadth continuum. The first of these dimensions is that of ego control: One's ability to contain impulse and remain insulated from the environment. Individuals representing one extreme of ego control are likely to be so responsive to the environment that differences and boundaries are overlooked. This individual is likely to be a broad categorizer. Individuals representing the other end of the ego control spectrum may prefer to create highly differentiated, narrow categories. Block, Buss, Block, and Gjerde (1981) have confirmed these relationships between ego control and categorization breadth among 4- to 7-year-olds, using four independent measures of breadth to generate an aggregate, more error-free (Epstein, 1983) breadth metric.

Block (1982) has recently gone a step further to provide the theoretical groundwork for relating the second dimension, ego resilience, and categorization breadth. Block proposed that the ego-resilient individual is likely to take a relatively strong accommodative approach to categorization, one that makes the individual's existing cognitive structures readily available to change based on information in the environment. Thus, the highly ego-resilient individual is hypothesized to be a narrow categorizer. The low ego-resilient individual will prefer an assimilative strategy, one that attempts to fit as much variation into existing cognitive structures in order to avoid the uncertainty of change. This individual is hypothesized to be a broad categorizer. Metzger and Miller (1985) provide data supporting this view.

The preceding considerations suggest that Block's two dimensions of personality may be related to categorization breadth behaviors. However, a number of important questions must be answered before such a conclusion can be validated. First, the preceding arguments suggest that categorization breadth should be stable across time and consistent

across tasks. In fact, we find very little data with which either to confirm or disconfirm this supposition. What does exist is too often uninterpretable due to methodological confounds. Second, the relationships between breadth and ego resilience and ego control are yet to be replicated. Finally, we are concerned that the data that Block et al. (1981) provide may be specific to the developmental levels of the subjects assessed. It may be the case that young children have not learned very many principles by which stimuli may be differentiated. Thus, a global categorization style may result. Adults, having acquired a larger variety of perceptual rules, may have a variety of categorization styles. The present study will begin to address these issues.

Hypotheses

At this point it is possible to summarize the goals of the present study. Our immediate questions concern the cross-task consistency and cross-time stability of categorization breadth, as well as the relationship of these indices to Block's ego constructs. Implicit in an examination of these issues will be a discussion of aggregation of measures as a means of reducing error variance. We propose to create aggregate breadth measures with which to evaluate better the foregoing questions.

Specifically, cross-task consistency is expected to be moderate to low (Block et al., 1981). However, as measurement error may contribute to this relative inconsistency, we posit that an aggregated measure of breadth will provide a truer index of consistency (Epstein, 1983). Cross-time stability within task is likely to be moderate to high (Block et al., 1981; Kogan, 1977). However, here again we expect that even greater stability should be evident when multiple measures are aggregated.

Note that these hypotheses will allow us to go far beyond Block et al.'s study by considering whether measurement error is indeed a source of significant confounding in the assessment of breadth as a unitary construct, or if, alternatively, breadth may be domain- or task-specific. Mischel's (1968) classic position on the consistency of behavior might favor the latter interpretation. Further, we will be able to determine the degree of stability inherent in the breadth metric. Each of these is a critical theoretical issue.

Finally, cross-task and cross-time breadth composites are predicted to be related to both ego control (Block et al., 1981) and ego resilience (Block, 1982).

Method

Subjects

Thirty-five undergraduate students enrolled in introductory psychology courses at The Pennsylvania State University agreed to participate in exchange

for course credits. Subjects understood that participation required four consecutive sessions of up to one hour each, spanning a period of approximately six weeks. The mean intersession interval was 10.75 days. Nineteen of the subjects were female; sixteen were male. The subjects' average age was 21 years, 3 months; the age range was 19 to 29 years.

Stimuli

Six stimulus sets were prepared for individual presentation to subjects in the present study. Each set was composed of stimuli representing points along a graduated continuum of discrepancy from a central member or prototype. The relative position of each stimulus along the ordinal continuum of each set was validated by one or more of three techniques. Depending upon the task, ratings, rankings, and indices based on the stimulus' psychophysical properties were used in pilot studies to develop the final stimulus sets.

Two stimulus sets were composed of line drawings depicting imaginary animals which we called *Beasties* and *Batties*. There were 25 members in each set. A third line drawing stimulus set depicted schematic faces very similar to those used by Strauss (1979). There were seven members in this set. All three sets were drawn in black ink on white, $8\frac{1}{2}'' \times 11''$ cards. Each was constituted by three or four attributes that were varied systematically in order to create the continua of variation. These varying attributes were set in a simple context of unchanging features. For example, for the *Faces* set, the forehead height, eye separation, nose length and mouth height covaried within the context of an unchanging, oval face with eyes of a fixed diameter.

For each of these stimulus sets, measurement of the variable features allows placement of a particular stimulus along a continuum of discrepancy from the category's prototype. In all cases where more than one feature varies, the features vary in a correlated manner. Thus, all three of these sets can be arranged along an ordinal, unidimensional continuum of change. Using the *Faces* example again, at one extreme is a face with close-set eyes, a small forehead, a stubby nose and seemingly without a chin. The opposite values constitute the stimulus at the other extreme of the *Faces* continuum. The intermediate valued stimuli can be reliably arranged in order based on these dimensions.

Two other stimulus sets were composed of color chips. Each stimulus in these sets consisted of a $1\frac{1}{8}'' \times \frac{3}{4}''$ rectangle of color mounted in the center of a 4- by 6-inch white, unlined card. There were nine stimuli in the *Yellows* set and eight in the *Blues* set.

The sixth and final stimulus set was composed of 18 color photographs of butterflies. In order to differentiate this set from butterflies in general, we refer to these stimuli as *GB-1s*. This was necessary so that our subjects would make categorical decisions that were relatively unbiased by previous experience. Each *GB-1* was mounted on an unlined, white $5'' \times 8''$ card. Because butterfly size varies in these photographs, it is impossible to standardize the size of each picture. However, it is possible to say that the final set of *GB-1s* contained butterflies ranging from approximately $7'' \times 4\frac{1}{2}''$ to $3'' \times 1\frac{1}{2}''$.

The central member or prototype from each of these six sets was set aside for initial presentation with each of the sets. Duplicates of the prototypes and information on the development of each stimulus set is available from the first author upon request.

Procedure

On each of the four occasions subjects were seated opposite the examiner with a 12-inch high barrier placed between them. This barrier allowed eye contact between subject and examiner but prevented the subjects from prematurely viewing the stimulus materials. The side of the barrier facing the subjects was painted black. Clips at the top of the barrier allowed vertical display of each category's central member or prototype.

Three stimulus sets were presented to each subject on each visit. The Beasties and the GB-1s sets were presented on all occasions in addition to a third, varied set. Order of presentation was counterbalanced across subjects and within subject over time. For occasions one through four, the third task was Faces, Yellows, Batties, and Blues, respectively.

Subjects were led to believe that the stimuli accompanying the Beasties and the GB-1s central members differed on each occasion of presentation. Thus, subjects were instructed to make new category decisions relative to each of the category's central members on each of the four visits. In fact, they were repeating the identical task. Subjects' uniform surprise when informed of this deception upon debriefing as well as the consistency data presented below both suggest that the deception was successful.

Each set was presented as a problem of differentiating members vs. nonmembers based on perceived similarity to the prototype. Brief, story-like contexts were created in order to present a context within which to make these differentiations. Each set's prototype was presented and remained visible from the introduction of the individual task until the subject indicated that he or she had completed the member/nonmember stimulus assignments, at which point the prototype was removed and the subject's stimulus assignments were set aside for later scoring.

After completion of the three tasks administered on the first visit, subjects were asked to complete two self-report measures, the ER (Ego Resilience) and EC-5 (Ego Control) questionnaires. The ER scale (Metzger & Miller, 1985) is an adaptation of two previous measures of ego resilience. The EC-5 scale is a questionnaire designed by Block (1965) to assess ego control. Ego resilience and ego control are complex constructs and so it is unlikely that self-report measures such as the ER and EC-5 scales will capture their full meaning. However, these measures have demonstrated extensive reliability and validity (e.g., Block, 1965; Metzger & Miller, 1985). Moreover, the preferred approach for indexing ego resilience and ego control, via the California Q-set (Block, 1978), was not feasible given the constraints of the present study.

Scoring

Each subject was assigned a breadth score for each task on each occasion. This score is the number of stimuli endorsed within each task. At the completion

Table 1. Coefficients of consistency for single task comparisons within occasion of measurement.

Categories	Week 1	Week 2	Week 3	Week 4
Beasties and GB-1s	.25*	.28**	.30**	.28**
Beasties and third	.07	.16	.08	.20
GB-1s and third	.12	-.14	.14	.02

Note. — "Third" refers to the nonredundant tasks administered on each occasion (i.e., Faces, Yellows, Batties, and Blues on Weeks 1 through 4, respectively).

* $p < .10$

** $p < .05$

of the study twelve such measures were available for each subject, three on each of four occasions. This includes the four measures of each of the two redundant sets.

Results

Consistency Across Tasks

Correlation coefficients were computed in order to determine the degree of cross-task consistency of the breadth measures. These coefficients are presented in Table 1. Clearly, the magnitude of these relations is not impressive. While Beasties and GB-1s are significantly related to each other on each occasion, no correlation exceeds $r = .30$. Correlations both across task and time (e.g., Beasties on visit 1 and GB-1s on visit 2) generate similar results: Beasties and GB-1s are significantly related, but no correlation exceeds $r = .35$. The median cross-task, cross-time correlation is $r = .29$, with a range of $r = .16$ to $.35$ (ns and $p < .05$, respectively). Of the other possible correlations, only one achieves significance (range, $-.22$ to $.28$, ns and $p < .05$).

These findings suggest a relatively weak relationship between the different categorization breadth tasks. While Beasties and GB-1s are more highly related at each occasion than either is with the third task, the magnitude of this relationship remains modest across all occasions. Had the Beasties-GB-1s correlations increased across time, we would have suspected that the deception employed had failed.

Stability Across Time

Table 2 presents the within-task, across-time correlation coefficients for Beasties and GB-1s, the two repeated tasks. For comparison, we have treated the third tasks on each occasion as generically equivalent and thus present these coefficients in Table 2 also. It is important to re-

Table 2. Coefficients of categorization breadth stability within task between all pairs of occasions of measurement.

Category	Occasions of Measurement					
	1:2	1:3	1:4	2:3	2:4	3:4
Beasties	.71*	.58*	.59*	.86*	.88*	.98*
GB-1s	.86*	.82*	.81*	.93*	.90*	.89*
Third	.04	.03	-.05	-.18	.62*	-.09

Note. — "Third" refers to the nonredundant tasks administered on each occasion (i.e., Faces, Yellows, Batties, and Blues on Weeks 1 through 4, respectively).

* $p < .001$

member, however, that these are essentially across-task and across-time consistency measures.

Stability of categorization breadth is remarkably high within the Beasties and GB-1 categories. The median correlations are $r = .78$ and $.88$ ($p < .001$ for each). Thus we conclude that categorization breadth is highly stable within task.¹

Across-time correlations among the third tasks yield coefficients of essentially zero. That is, subjects are responding highly independently to the novel task presented on each occasion. Given that these tasks do not even share an occasion of measurement, these findings can be taken as additional support for the finding of minimal consistency in the within-occasion measures discussed above.

The clear exception to this conclusion is the Yellows-Blues coefficient ($r = .62$, $p < .001$). This relationship is of the magnitude of the stability coefficients within-task. It suggests that breadth is not entirely category-specific. Rather, it appears that to the extent that two categories are related, individuals' breadth judgments will likewise be related. We take this finding as lending still further credence to the assumption that subjects treated each presentation of the two redundant categories as independent.

1. We should note that this *correlational* stability occurred in the face of mean changes in the number of stimuli assimilated within each category. *T* tests indicated that the change in breadth scores for Beasties (from 9.00 at week 1 to 4.54 at week 4) was significant, $t(34) = 4.14$, $p < .001$. Change for GB-1s (from 2.43 to 3.40) was similarly significant, $t(34) = -2.64$, $p < .05$. Such changes further verify that our subjects were treating each presentation separately and were not making an attempt at artificial consistency. The opposing trends for the two categories are interesting (especially given the already established correlational stability) but, since they were unexpected, we make no attempt at interpretation. Identification of the causes of such changes, however, would appear to be an important pursuit.

Categorization Breadth and Aggregation

To assess the potential effects of measurement error on the present breadth assessments, three aggregates were generated. Individual task breadth scores were standardized with a common mean and standard deviation. Two within-occasion aggregates were created as a means of possibly enhancing stability estimates by reducing extraneous variance. The first is an aggregate of all three within-occasion measures. The second aggregate includes the Beasties and GB-1s, eliminating the highly inconsistent third tasks. The median stability coefficients for the two- and three-task aggregates were .87 and .65, respectively. While both of these coefficients are highly significant, they do not improve upon the strong, single task stability coefficients reported above.

The third aggregate measure of breadth sums the four observations within each of the two redundant categories and the four third tasks. In this way, consistency between tasks can be studied. The summation across the four third tasks is artificial, as these are four different tasks. However, this aggregate serves as a comparison measure. Of the three possible relationships, only the Beasties and GB-1s are significantly related ($r = .31, p < .05$). The Beasties and GB-1 aggregates correlated with the third task aggregate at $r = .15$ and $r = .09$. Once again, aggregation did not greatly improve the observed correlations.

Categorization Breadth, Ego Resilience and Ego Control

In order to examine the potential relations between the present categorization breadth measures and the two personality measures, two correlational analyses were conducted. In the first, correlations were computed between the within-time aggregate of the two redundant categories and the two ego measures. Of the resulting eight coefficients, none were significant. This is also true for the within-occasion aggregates of three tasks. When gender is introduced as a mediator in each of the above sixteen relationships (creating 32 coefficients), only one coefficient achieves minimal significance. Thus, it appears that the within-occasion aggregates and the personality measures are unrelated.

In the second approach, the two within-task aggregates (Beasties, GB-1s) were correlated with the ego measures using gender as a likely moderator (Block et al., 1981). The resulting eight coefficients are presented in Table 3. Note that only three of these coefficients obtain even moderate significance. Also note the slight tendency for breadth to be more strongly related to ego control for males and to ego resilience for females. However, this post hoc observation is weak and nonsystematic. Rather than attempt to build an interpretation out of such findings, we prefer to conclude that there is at best a fragile and nonsystematic rela-

Table 3. Coefficients representing the relationship between the two within-task, across-time aggregates and each of the two ego constructs stratified by gender.

Aggregate	Ego Constructs			
	Males		Females	
	Ego resilience	Ego control	Ego resilience	Ego control
Beasties				
Across-time	-.26	-.41*	-.12	-.05
GB-1s				
Across-time	.04	.35*	.34*	-.13

* $p < .10$

relationship between categorization breadth, ego resilience and ego control among the present sample of college students.

Discussion

The importance of the present research lies in the central role that categorization processes play in human information processing. We have addressed three important questions concerning individual differences in categorization: The extent to which categorization breadth is consistent across tasks, stable across time, and the relationship between breadth and Block and Block's (1980) ego constructs.

In short, we have found that breadth is at best moderately consistent between tasks, with consistency increasing as the apparent content relatedness of the tasks increases. Breadth is highly stable across the period of time studied, both for individual task measurements and aggregations of two or more tasks. Finally, our search for personality correlates of breadth uncovered a labyrinth of interactions. Gender, category, and ego construct together contributed to a pattern of relations which we did not attempt to interpret. However, the extent to which these findings are specific to the sample and the assessment materials used is unclear. Differences in age, developmental stage, geographic background, and other factors may account for differences between the present findings and those of Block et al. (1981) or Metzger and Miller (1985).

Regarding consistency, we observed that of all the single task comparisons, only the Blues-Yellows comparison was highly significant. In general, this would suggest that subjects bring largely different criteria of category membership to each task. The single very significant relationship between breadth in the Blues and the Yellows categories, however, allows us to temper this generalization. The magnitude of the relation-

factors. We feel that differences in tasks, gender, and age of subjects, and the validity of the ego measures used here and in other studies are among the variables that serve to confound understanding of these relationships. Each of these variables will require much systematic research before uniform and replicable ego construct-breadth relationships can be observed.

Future Issues and Directions

In sum, the present study has documented highly stable (within task) categorization breadth among adults over a brief period of time. Consistency of these same measures between tasks is limited. Only those tasks sharing similar content demonstrated strong interrelationships. Relationships between breadth and ego measures were at best obscured by a number of confounding factors. In response to these findings, we are positing that breadth is domain-specific: a personal style difference, differentially deployed by category.

A much stronger conclusion could be drawn from these data: Categorization breadth is an irrelevant and useless concept. In our view, this perspective throws the Beasties out with the bath water. It seems much more useful to call for thorough and systematic research into the process of categorization and those variables that affect it.

Specifically, our conclusions suggest that future research efforts would be well spent investigating the nature of subjectively salient features that differentiate categories. For example, we have suggested that category content similarities and differences play an important role in determining the consistency of breadth judgments. We expect that certain content features are more relevant than are others. Thus, we are calling for careful investigation into the structures that relate categories into coherent domains, or meta-categories. Eventual specification of such subjectively determined clusters of categories would have important implications for the prediction of behavior across situations, for we would have some small insight into how situations are subjectively clustered.

Equally important to the cross-context issues inherent in an understanding of stimulus perception and subsequent breadth decisions is an understanding of cross-time issues in the development of breadth criteria. Block et al. (1981) have suggested that breadth is a highly generalized, undifferentiated style. Our results indicate that a multifaceted network of styles exists. Perhaps breadth proceeds from a global style early in development to a more complex set of criteria in adulthood. This would be highly consistent with Werner's (1957) conception of development. Such developmental questions likewise begin to open up issues of the relative universality or uniqueness of meta-categories (categories of

categories) to which similar criteria are applied. It may be more useful at this point to seek to understand individual differences in the perception of similarities between categories than differences in membership assignments within categories.

As well, future research in categorization will no doubt seek to discern the intricate patterns of relationships between particular breadth styles and personality variables. This effort will be greatly aided by a better understanding of the perceptions of relationships between categories as described above, as well as by a precise specification of the roles of such moderating variables as age and gender.

References

- Block, J. (1965). *The challenge of response sets*. New York: Appleton-Century-Crofts.
- Block, J. (1978). *The Q-sort method in personality assessment and psychiatric research*. Palo Alto, CA: Consulting Psychologists Press.
- Block, J. (1982). Assimilation, accommodation, and the dynamics of personality development. *Child Development*, 53, 281-295.
- Block, J., Buss, D. M., Block, J. H., & Gjerde, P. F. (1981). The cognitive style of breadth of categorization: Longitudinal consistency of personality correlates. *Journal of Personality and Social Psychology*, 46, 770-779.
- Block, J. H., & Block, J. (1980). The role of ego-control and ego-resiliency in the organization of behavior. In W. A. Collins (Ed.), *Minnesota symposia on child psychology* (Vol. 13, pp. 39-101). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, 51, 360-392.
- Kogan, N. (1977). *Cognitive styles in childhood and early infancy*. New York: Wiley.
- Metzger, R. L., & Miller, M. L. (1985). Individual differences in ego resiliency are a function of assimilation and accommodation. Manuscript submitted for review.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Mischel, W., & Peake, P. K. (1982). Beyond déjà vu in the search for cross-situational consistency. *Psychological Review*, 89, 730-755.
- Pettigrew, T. F. (1958). The measurement and correlates of category width as a cognitive variable. *Journal of Personality*, 26, 532-544.
- Strauss, M. S. (1979). Abstraction of prototypical information by adults and 10-month-old infants. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 618-632.
- Werner, H. (1957). *Comparative psychology of mental development*. New York: International Universities Press.

Manuscript received October 11, 1985; revised May 19, 1986.