

More on parts in object concepts: Response to Tversky and Hemenway

GREGORY L. MURPHY

University of Illinois, Champaign, Illinois

This article responds to the comments of Tversky and Hemenway (1991), who criticized the logic, stimuli, and data analysis of Murphy (1991). It is argued here that their objections do not mitigate the conclusions drawn by Murphy. In particular, the objection that the stimuli were not natural enough to reveal differences between category levels seems to presuppose an answer to the question under investigation. However, further experimentation with other stimuli might resolve this issue empirically.

Before beginning my response to Tversky and Hemenway's (1991) comments, I would like to emphasize the points that I believe we agree on. I think that we concur on the general nature of the basic-level phenomenon and on how that phenomenon predicts performance in tasks. I believe that we also agree on most of the causes of basic-level structure. Finally, we agree on the more specific point that in most natural categories, the basic level contains distinctive parts. Where we disagree is in the significance and interpretation of the last point, which the experiments in Murphy (1991) were designed to investigate. Thus, although this rebuttal will naturally focus on the areas in which we conflict, it is important to keep in mind the wide range of issues in which there is general agreement in the field.

Tversky and Hemenway (1991; henceforth, I will refer to this article as "T&H") have comments on the logic, stimuli, and results of my experiments. These issues are important ones for both theoretical and methodological reasons. I will address each of these issues in turn, somewhat reordering their points to ease exposition. To give a glimpse of my conclusions, I will argue that their concerns with the logic and results of the experiments are not actually problems. I will agree with them that the stimuli in Experiments 1-3 are not very natural. However, I will also argue that these problems do not mitigate the specific conclusions drawn from the experiments.

EXPERIMENTAL LOGIC

Previous research on the basic level of categorization has suggested a number of possible explanations for "what makes basic categories basic." One suggestion discussed by Murphy and Brownell (1985) was the following hypothesis:

H0: Basic categories are categories that are both informative and distinctive.

A category is informative if it is associated with a number of features, so that knowing that something is in the category communicates a large amount of information about it. A category is distinctive to the degree it differs from its contrast categories. If some categories overlap too much, it is difficult to tell which one an object is in. This hypothesis is consistent with most of the data on basic-level categories and was explicitly supported by experiments using natural categories in Murphy and Brownell (1985). Also, similar accounts have been adopted by other theorists (e.g., Corter, Gluck, & Bower, 1988).

The important findings of Tversky and Hemenway (1984), however, suggest a different possible explanation. In this paper, Tversky and Hemenway discovered that many of the attributes listed for basic categories were parts: leg, top, seat, wheel, nose, and so forth. These parts were not generally listed at the superordinate level; here, primarily functions or behaviors were listed. At the subordinate level, parts were listed, but the parts for contrast categories were often identical (cf. the distinctiveness criterion of H0). So, different kinds of tables almost all have four legs and a top. They differ not in their parts but in other attributes, such as size, shape, location, or function. This finding suggests a different explanation for what makes basic categories basic (though one that Tversky & Hemenway did not explicitly propose).

H1: Categories that have distinctive parts will be the basic categories in a hierarchy.

This hypothesis is similar in some ways to Biederman's (1987) recognition-by-components (RBC) model of object recognition. This influential model suggests that object identification proceeds by recognizing the parts of objects, implying that categories with distinctive parts should be easiest to identify.

A final point about these two hypotheses is that H1 refers to the content of a concept—what type of features it has. In contrast, H0 is a structural hypothesis that refers to abstract properties regardless of their content.

It is important to understand that Tversky and Hemenway's (1984) results are primarily correlational. That is, they correlated different levels of categorization (as defined by previous work) with the content of the features (functions, parts, and perceptual features). Thus, if one wants to evaluate H1, one cannot stop with their results. It could be that some other variable is causally responsible for the basic-level advantage, but that it and part configuration are themselves correlated. In order to discover whether parts are in fact causally responsible for the basic-level advantage, it is necessary to directly manipulate them and observe their effect on category structure.

What "third factor" could possibly be responsible for the basic-level advantage, if not part configuration? The obvious possibility is that suggested by H0: Perhaps the combination of informativeness and distinctiveness is

Please address correspondence to Gregory L. Murphy, Department of Psychology, University of Illinois, 603 East Daniel St., Champaign, IL 61820.

causing the basic-level effect, which happens to be correlated with the presence of parts at that level. And, in fact, this correlation is probably no coincidence. If parts are related to an object's function (Tversky & Hemenway, 1984), and if part configuration gives rise to shape, then having distinctive parts will often produce the situation described in H0: categories that are informative and distinctive. However, these two hypotheses are not identical. H0 says that it is the general properties of informativeness and distinctiveness that cause the basic-level effects and that the presence of parts is not necessary for the effects—it just happens that real objects within a category share parts. In contrast, H1 says that it is parts in particular that are crucial, not overall informativeness and distinctiveness.

These two hypotheses can be distinguished by two predictions that H1 makes but that H0 does not:

Prediction 1: To the degree a taxonomy does not have parts collected at one level, it will not display basic-level phenomena.

Prediction 2: To the degree a taxonomy has parts collected at one level, that level will tend to display basic-level phenomena.

These seem very reasonable implications to draw from H1, if it is to have any empirical consequences.

It is just these two empirical predictions that I intended to refer to when asking whether parts are "necessary" (Prediction 1) or "sufficient" (Prediction 2) to generate basic-level structure. This high-blown language may have given the impression that I was making an ontological claim about categories. T&H object that "It is puzzling to pursue questions of necessity and sufficiency in categorization, given that they are most likely unanswerable" (p. 441). However, no such ontological claim was intended. Of course, it is unlikely that a *single* variable will explain this phenomenon; more likely, a complicated causal analysis will be necessary. I suspect that this is part of T&H's point, and I concur.

Others have independently suggested the same test of H1. Markman (1989, p. 68), reflecting on the association of parts and basic categories, proposed that "What we really would like to know, however, is this: if we were to find or create categories that shared many nonpart features, would they serve just as well as categories even at the basic level? To my knowledge, this study has not been done, nor is it an easy one to design." Experiments 1–3 of Murphy (1991) take just this approach to understanding the connection between parts and basic levels.

EXPERIMENTAL STIMULI

To put it mildly, T&H are not happy with the stimuli used in the first three experiments of Murphy (1991). Their objections include the following: (1) the stimuli resemble postage stamps, (2) they are perceptually defined at all levels of abstraction, (3) they have no functions, and (4) they share the same general shape. Perhaps the best

stimuli to test the predictions of H1 would be stimuli that are just like regular objects, except that they have very few parts. However, any stimulus with very few parts looks quite different from most objects. The actual stimuli used in Experiments 1–3 give the impression of being visual patterns rather than solid objects. This is an advantage from the point of view of eliminating parts from the stimuli but a disadvantage from the point of view of simulating natural objects.

However, there is another side to this issue, namely that lack of naturalness is primarily a problem when different results are found than what are normally achieved with natural objects. T&H's objections, which are discussed in more detail below, ignore one major aspect of the stimuli—that they in fact generated statistically reliable basic-level effects using a number of different measures. In feature listings (Experiment 1A), in similarity ratings (Experiment 2), and in identification times (Experiment 3), the normal basic-level advantage was found at the expected level. This is strong evidence that the stimuli were natural enough *in relevant respects*. That is, although they differed from familiar objects in many ways, they were similar enough in some respects to yield the standard basic-level effects. Exactly what these "respects" are provides important information about the causes of basic-level structure. Since it was apparently not part structure (see below), it may well be that it was the informativeness and distinctiveness of the stimuli.

In short, the very artificialness of the stimuli places strong constraints on any account of basic-level effects. That is, no account of basic-level structure can depend on category members being just like natural objects if it is to explain these results. Analogously, Posner and Keele's (1968) and Rosch and Mervis's (1975) experiments with very artificial stimuli showed that prototype effects do not depend on stimulus naturalness.

One possibility T&H might be concerned about is that such artificial stimuli could engender specific strategies that are very different from those used with normal objects. Of course, it would be a surprising coincidence that these unusual strategies engender the expected basic-level results for three different dependent measures. However, to rule out this possibility conclusively, one could attempt to replicate Experiments 1–3 with other, perhaps more naturalistic, stimuli.

T&H list a number of quite specific complaints about the stimuli of Experiments 1–3. Some of these are details of the ways in which the objects do not resemble natural objects, as just discussed (e.g., their postage-stamp appearance). However, some of their objections appear to be prejudging the very debate that the experiments were designed to investigate. That is, they object that the stimuli lack some attribute X (which normal objects have), when it is exactly the importance of X that is being questioned. For example, they object that the stimuli "violate the qualitative pattern of features in natural categories, in which for superordinate categories, functional features outnumber perceptual ones" (p. 440). Since my stimuli did not use functional features, they obviously violate this

pattern. More generally, they argue that "artificial stimuli must adequately model the structure and features of natural categories and taxonomies" (p. 440).

But violating "natural" properties of taxonomies was not an error: The whole point of this series of experiments was to discover which aspects of natural categories are responsible for basic-level structure. This investigation *requires* the separation of different potential causes, thereby disrupting the natural co-occurrence of these causes. Thus, it is no mistake that these particular aspects of the stimuli are not "natural." T&H are making assumptions about which aspects of real-world categories are relevant, when these assumptions are the very subject of the investigation. For example, H0 does not say that superordinates must be defined by functions. If it is a correct explanation of basic-level structure, then the presence of functions or parts at one level is not essential.

Measuring Part Structure

T&H complain that Experiment 1B used an "unusual" procedure in deriving the number of parts in the stimuli. However, it is not clear that there is a "usual" procedure for part counting, and they only cite one paper that counted parts (Tversky & Hemenway, 1984). In particular, they are concerned about the use of one group of subjects to supply the attributes in general and another group to list the parts. Since "parts are, after all, a subset of attributes" (p. 440), they would have preferred the part listing to be derived from the attribute listing.

This issue is an interesting methodological point, and there are actually data that can help adjudicate it. In a study of event categories, Michael Morris and I (Morris & Murphy, 1990) asked subjects to list the parts of events (i.e., actions). We used the categories of Rifkin (1985), who asked subjects to list the attributes of events. Rifkin reported that 34 of the features in his taxonomy were parts (counted before any judge amending). However, our subjects who were instructed to list only parts produced 338 entries for the same taxonomy.¹ Unsurprisingly, asking subjects to focus on parts alone directed their attention more towards parts. Thus, it seems very unlikely that using different subjects to list "attributes" and "parts" in Experiments 1A and 1B reduced the number of parts produced. If anything, the opposite may be true.

In a more theoretical vein, I should note that it is questionable whether college students' pretheoretical notion of parts is necessarily the correct notion for a scientific understanding of categorization. For example, few, if any, subjects could articulate the notion of a part as defined in the object-recognition literature (see Biederman, 1987; Hoffman & Richards, 1985; Marr, 1982; Schyns & Murphy, 1991). The fact that some subjects say that a leopard's spots are parts (as pointed out by T&H) is good evidence that the technical concept of a part is not exactly the same as the naive notion.² Nonetheless, the judges in my study generally agreed that there were few parts in the stimuli and that they were not clustered at any one level.

To conclude this section, I believe that the stimuli used in Experiments 1-3 indeed lacked something in naturalism. However, I have argued that the specific problems listed by T&H do not mitigate the conclusions drawn. The fact that a basic-level advantage was actually obtained in three different dependent measures casts doubt on the argument that the stimuli were too unnatural to reveal anything about category structure. And others of their objections seem to prejudice the very question of what determines category structure. Nonetheless, I agree it would be worthwhile to attempt a replication of the experiment with more naturalistic stimuli, if they can be constructed.

RESULTS

T&H did not object as strenuously to the stimuli of the last two experiments, but they did argue that their results did not reflect changes in the basic level. This raises the important issue of how to measure basic-level structure. Thus, I will discuss statistical comparisons of category levels in some detail.

Experiment 4 compared "simple" stimuli (black-and-white drawings of artificial tools) with "enhanced" stimuli (colored drawings of the same tools, with size and texture variations). The enhancement was such that it added features at the proposed basic level without adding parts, thereby testing whether nonpart features have an effect, as H0 predicts they should. This manipulation made subordinates less distinctive (and superordinates slightly more distinctive) than in the simple condition. And the results showed that subordinate categories become reliably slower in the enhanced condition, thereby increasing the basic-level advantage.

As Murphy (1991) described explicitly, the measure of basic-level structure used was the reaction-time (RT) difference between basic categorizations and the other levels. T&H do not provide an explicit argument against this measure, nor do they suggest another measure of basic structure, but they make other comparisons that they argue show little difference in category structure. They point out that "There were no differences between the simple and enhanced conditions in reaction times to the middle [basic] levels or in the difference between the middle and highest levels. From this pattern of data, increased labeling time at the lowest level, one cannot infer, as Murphy did, that the 'basic' level has been enhanced. One can only infer that the lowest level has been made more difficult, and that is readily explained by decreased discriminability" (p. 441).

However, there is a problem with the comparisons that T&H make here. An example may help explain it. Imagine that you wanted to decide whether the concept *piano* was a basic concept. Suppose further that you knew that the average categorization time for identifying pianos was 700 msec. Now, you could not compare this time to other basic categories to find out whether *piano* was basic. That is, you could not say that this is about the same speed

as categorizing into *dog*, *chair*, or *car*, and so it is basic. Equally, you could not say that because it is slower than these times, it must not be basic. Rather, *piano* is only a basic category if it is the preferred category within its taxonomy. That is, *piano* must be compared with *musical instrument* and *grand piano* or *player piano*. In fact, the very notion of a basic level as described by Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) is a relative one—the level of categorization that is better than lower or higher levels.

The comparison between *piano* and *chair*, say, would be inappropriate because of the many confounding factors that might contribute to making one faster than the other: size, familiarity, perceptual properties, linguistic codability, and so forth. But comparisons within a hierarchy would be appropriate because the same stimulus item can be used for all three levels. That is, the identical picture of a grand piano could be categorized as a *grand piano*, *piano*, and *musical instrument*.

The comparison made in Experiment 4 was of the latter sort, namely the relative speed of the basic level and the other levels within a condition. As H₀ predicted, the basic-subordinate difference decreased, but the basic-superordinate difference increased slightly (nonsignificantly). T&H dismiss the first effect as being “readily explained by decreased discriminability” (p. 441), but it should be pointed out that this is exactly my explanation, as stated in H₀ (that basic structure is caused by informativeness and distinctiveness—i.e., discriminability).

The comparison T&H make of the enhanced and simple stimuli at the basic level has the same problems as comparing *piano* with *chair*. The enhanced and the simple stimuli differ in a number of perceptual variables that might slow or speed up their overall RT. For example, the enhanced stimuli were drawn in colored ink, whereas the simple stimuli were black and white. If the contrast of the two inks differed, this could create an RT difference between the two that has nothing to do with category level. Another confound is that some enhanced stimuli were smaller than the simple stimuli. Furthermore, there could be cognitive differences between the conditions as well: The enhanced items had many more features overall, and this might have influenced their identification times.

Thus, it is inappropriate to compare the enhanced and simple stimuli at a single level of categorization, since this comparison includes the confounds just mentioned. What is appropriate is to compare *differences* between levels in each condition (i.e., an interaction of condition and level), since these will be unconfounded by main effects due to item differences. The reliable interaction in the results shows that nonpart features did influence the preferred level of categorization, indicating that the relative speed of categorization is not dependent solely on parts.

T&H also object to the methods and results of Experiment 5. They argue that “adding a large number of visual features to the highest level again violates the structure of natural category taxonomies” (p. 441) (since natural superordinates have primarily functional features). How-

ever, this is another case of prejudging the question, as described above. If H₁ is correct, and parts determine basic category structure, then adding nonpart features should not influence it, even if natural categories do not have such a pattern.

As to the results of Experiment 5, T&H argue that the speedup in the highest level does not represent a change in basic-level structure: “‘no difference’ is not the same as eliminating a basic level” (p. 441). But if we define the basic-level advantage as the preference for one level relative to the others (as just argued), then such an effect is perforce changing the basic-level advantage. Perhaps T&H are simply reminding us that the null hypothesis cannot be proved. However, it is important to keep in mind that in two previous experiments (Murphy & Smith, 1982, and Experiment 4 of Murphy, 1991) using the simple versions of the same stimuli, the highest level was slowest (for true responses). In this context, a manipulation that makes the highest level the fastest seems quite dramatic.

CONCLUSION

T&H have made a number of important comments on the design and results of the experiments in Murphy (1991). I have argued that their criticisms of the results and the logic of the experiments do not require any change in the conclusions of the study. Their comments on the necessity and sufficiency of parts seems to be simply a misunderstanding of my intentions, probably caused by lack of detail in the writing of my original paper. The controversy over the results may be because of bona fide disagreement over how to measure the basic-level advantage. However, I have argued that the comparisons that T&H make contain possible item confounds, whereas the comparisons made in the original paper do not.

The problems with the stimuli are not so easily dismissed. Whereas some of T&H’s objections seem to prejudge the question of what determines basic-level structure, their more general concerns that the stimuli do not resemble natural objects should be taken seriously. Although I believe that it is unlikely that stimulus-specific strategies would just happen to produce the predicted basic-level phenomena, this matter could be resolved by testing very different, preferably more naturalistic stimuli. In spite of this limitation, as the first study that attempted to manipulate parts and observe the effects on levels of categorization, the experiments in Murphy (1991) provide a first step in understanding the causal connection of parts and basic-level structure.

REFERENCES

- BIEDERMAN, J. (1987). Recognition-by-components. A theory of image understanding. *Psychological Review*, *94*, 115-147.
- CORTER, J. E., GLUCK, M. A., & BOWER, G. H. (1988). Basic levels in hierarchically structured categories. *Proceedings of the 10th Annual Conference of the Cognitive Science Society* (pp. 118-124). Hillsdale, NJ: Erlbaum.

- HOFFMAN, D. D., & RICHARDS, W. (1985). Parts in recognition. *Cognition*, *18*, 65-96.
- MARKMAN, E. M. (1989). *Categorization and naming in children*. Cambridge: MIT Press.
- MARR, D. (1982). *Vision*. New York: Freeman.
- MORRIS, M. W., & MURPHY, G. L. (1990). Converging operations on a basic level in event taxonomies. *Memory & Cognition*, *18*, 407-418.
- MURPHY, G. L. (1991). Parts in object concepts: Experiments with artificial categories. *Memory & Cognition*, *19*, 423-438.
- MURPHY, G. L., & BROWNELL, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 70-84.
- MURPHY, G. L., & SMITH, E. E. (1982). Basic-level superiority in picture categorization. *Journal of Verbal Learning & Verbal Behavior*, *21*, 1-20.
- POSNER, M. I., & KEELE, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353-363.
- RIFKIN, A. (1985). Evidence for a basic level in event taxonomies. *Memory & Cognition*, *13*, 538-556.
- ROSCH, E., & MERVIS, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605.
- ROSCH, E., MERVIS, C. B., GRAY, W., JOHNSON, D., & BOYES-BRAEM, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382-439.
- SCHYNS, P. G., & MURPHY, G. L. (1991). The ontogeny of units in object categories. *Proceedings of the 13th Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- TVERSKY, B., & HEMENWAY, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, *113*, 169-193.
- TVERSKY, B., & HEMENWAY, K. (1991). Parts and the basic level in natural categories and artificial stimuli: Comments on Murphy (1991). *Memory & Cognition*, *19*, 439-442.
- WINSTON, M. E., CHAFFIN, R., & HERRMANN, D. (1987). A taxonomy of part-whole relations. *Cognitive Science*, *11*, 417-444.

NOTES

1. After judge amending, the figure was 476 parts. These figures cannot be derived from the results in Morris and Murphy (1990), because there we augmented Rifkin's taxonomy to increase the number of subordinate categories. The figures listed here include only the categories that Rifkin sampled and both are from the unamended tallies, so that they are comparable. It is also interesting to note that our subjects and Rifkin's listed similar numbers of (unamended) features as a whole: 290 in his study

and 338 in ours. It is just that most of Rifkin's features were not parts, since his subjects were not instructed to concentrate on them. Thus, studies of parts might be well advised to get listings of parts alone.

2. Although it is conceivable that the set of zebra's stripes as a whole or the set of leopard's spots as a whole serves as a part, it is not correct to suggest that any one stripe or spot is a part associated with an object concept. There are three reasons for this. First, such elements of a pattern are not perceived as individual features. Although the elements can be perceptually distinguished, the Gestalt laws of similarity and common fate result in their being perceived as an entire pattern rather than 100 individual spots or 12 individual stripes. For most objects, one does not identify and remember individual elements of this kind. I have a white shirt with blue and red vertical stripes, but I do not know how many stripes it has, nor have I encoded any individual stripes into memory, even though I am quite familiar with this shirt. Second, such elements are not constant across members of the category. Leopards do not all have the same spots in the same locations. If they did, then the concept *leopard* could be associated with each individual spot. But because of variation in their location, size, and shape, no individual spot is associated with the concept as a whole. Even worse, the number of these elements is not constant across category members. If one leopard has 100 spots and another has 120 spots, then it will be very difficult to assess which elements they have in common, since there is no one-to-one correspondence. (In contrast, because almost all leopards have four legs, which can be individuated by their locations, each leg could be encoded as a part.) Third, Winston, Chaffin, and Herrmann (1987) have pointed out that object components are in principle separable from the rest of the object. I would argue that these elements (and the stripes on my shirt) cannot be removed from the object without removing pieces of other, more coherent parts (e.g., skin or the shirt collar).

Note that the first two points apply to the texture elements in the stimuli of Experiments 1-3 of Murphy (1991), thereby arguing against any claim that they are individual parts. The speed of the identification times in Experiment 3 argues strongly that subjects were not counting or even identifying individual dots or stripes in the stimuli. The judges agreed that these elements were not parts, though they judged that the entire pattern formed a single part. I do not think that this is completely correct, either (the pattern is a surface property rather than a component), but it is their opinion that is reflected in the published part count. Even this liberal (by my standards) count did not create a cluster of parts at the middle level.

(Manuscript received May 28, 1991;
revision accepted for publication May 29, 1991.)