
Susan A. Gelman
University of Michigan

Douglas L. Medin
Northwestern University

Our first reaction to Jones and Smith (1993) reminds us of the classic paper in social psychology, "They Saw a Game: A Case Study," in which Hastorf and Cantril (1954) reported striking differences between Dartmouth and Yale students' perceptions of a hotly contested Dartmouth–Yale football game. The present circumstances are different in that understanding the development of cognition is anything but a game. Furthermore, we had thought that all of us were on the same team! Nonetheless, Jones and Smith see infractions where we see none, and sprinkled among their descriptions that we agree with were statements that made us cry "foul!" We do not imagine, however, that truth can be readily refereed. Therefore, our goal will not be to tally our grievances but rather to outline the different perspectives from which we approach the issue of conceptual development. We begin by underlining points of agreement. Then we shift to our point of view on the work described and criticized by Jones and Smith. Finally, we hope to converge on some critical issues that motivate the research agenda.

Points of Agreement
Jones and Smith see conflict and contrast between their results and those of our own. They seem to suggest that, because we propose that nonobvious features are important, we are claiming that obvious features are unimportant. Not so. We...
are instead struck by the commonalities between the two approaches. Both lines of research stress the following: (a) Perception, language, and conceptual knowledge mutually interact to determine task performance; (b) there is no simple developmental shift from perception to conception; (c) children do not reason about categories using a simple, context-free similarity metric; and (d) different categorization tasks lead to use of different kinds of information.

Both lines of research demonstrate that concepts are not unidimensional or fixed, but rather are multidimensional and open to restructuring. Both suggest that children may grasp the complex implications of perceptual features, and both suggest that the semantics of even familiar common nouns are not as obvious and uncontroversial as assumed on traditional accounts. Let us briefly expand on what Jones and Smith refer to as the “nonperceptual” view. Studies of induction have shown that, in addition to reliance on the category name, children attend to: (a) perceptual properties, making more inductions when within-category similarity is high than when it is not (Gelman, 1988; Florian, 1992; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990); (b) the semantics of the property being taught (Gelman, 1988; Gelman & Markman, 1986; Shipley, 1988); (c) category level (Gelman & O’Reilly, 1988); (d) ontological kind (Gelman & O’Reilly, 1988; Kalish & Gelman, 1992; Keil, 1989); (e) category structure (Davidson & Gelman, 1990); (f) category familiarity (Davidson & Gelman, 1990); and (g) (perhaps) syntactic information (Gelman & Coley, 1991; Kalish & Gelman, in press). The point from studies of induction is not that naming draws attention away from perception; indeed, ordinarily it does not (see Gelman & Coley, 1991; Medin, 1989; Medin & Ortony, 1989). Thus, we all seem to agree that conceptual development involves the interactive influences of perception, language, and conceptual knowledge.

How, if we agree on so much, can we be characterized as agreeing so little? In part this may be that the claim that concepts, perception, and language interact is not yet specified with sufficient rigor to allow one to detect if it is receiving support. We also suspect that within this broad framework there is room for a variety of perspectives, perspectives that can lead to very distinctive slants on the same issues. To outline our point of view we must turn to the broader question of what functions concepts serve.

Conceptual Functions
Concepts function in enormously varied ways. They can be used for extremely rapid identification (as when escaping from prey), organizing information efficiently in memory, problem-solving, analogizing, drawing inductive inferences that extend knowledge beyond what is known, embodying and imparting ideological beliefs, conveying aesthetic materials (e.g., metaphor, poetry), and so forth. These are different kinds of judgments that partake of different kinds of information. People rely strongly on perceptual similarity with some sorts of judgments, but focus on nonperceptual features more with others. One might
expect beliefs about essences to figure more prominently in judgments associated with Frank Keil's (1989) transformation experiments than in run-of-the-mill object recognition. In short, conceptual functions go beyond categorization.

One potential source of our differing perspectives is that different investigators focus on different conceptual functions, and often researchers fail to clarify which function is under study. Jones and Smith stress the identification function of concepts. In fact, they define "concept" as the represented structure "that allows members of a category to be recognized" (p. 114). We find this definition overly narrow. For example, interpreting novel conceptual combinations such as paper bee (presumably a bee made out of paper) or paper committee (presumably a committee concerned with paper but not made of it) is not a simple matter of conjoining two recognition functions (Downing, 1977; Medin & Shoben, 1988; Murphy, 1988). Nor is it clear that recognition routines reveal much about relations between concepts, a common concern in semantic analyses.

Different Slants
With all of the varied functions that categories serve, it is no wonder that the study of concepts cannot be distinguished from the study of the other kinds of information and cognitive systems they interact with. That being so, a critical issue is how to think about and perform experiments that describe the interplay of these factors. Smith and Jones suggest that the role of perceptual knowledge in concept formation and use has been systematically neglected and underestimated. In their view, when perceptual similarity has been contrasted with conceptual knowledge, the stimulus materials have been "highly impoverished" and, for that reason, the associated results have underestimated the importance of perceptual similarity in more normal, rich input conditions. We question several fundamental aspects of this view.

1. Comparing Perception with Knowledge? Our first point of disagreement is this: We do not think that asking about the relative importance of perceptual knowledge versus other forms of knowledge is a particularly meaningful question. The Gelman and Markman (1986, 1987) studies were aimed at the question of whether category membership information contributed to children's inductions in a manner that could not simply be attributed to perceptual similarity. The logic of this question about category membership dictates that perceptual similarity be held constant or biased against the hypothesis of interest. It is simply incorrect to suggest that these studies attempted to evaluate the relative importance of perception and conceptual knowledge to induction. If concept development and use reflect an interaction of perceptual knowledge, language, and conceptual knowledge, then it does not seem sensible to try to evaluate the relative magnitude of main effects.

It is also important to bear in mind the context for the original studies by
Gelman and Markman. Their work was in part designed to counter the view that perceptual similarity, uncontaminated by conceptual knowledge, is sufficient to account for the development of categorization. Their studies were designed to demonstrate that perceptual experience (more traditionally, stimulus generalization) is not the only factor controlling young children's reasoning about categories. Gelman and Markman showed that young children could use knowledge about category membership—even when it was pitted against perceptual similarity—as the basis for inferring other (hidden) properties.

2. Distorted Stimuli? It is incorrect to suggest that Gelman and Markman “distorted” similarities so that they did not coincide well with category membership. All items used in their studies of familiar categories were actual things that exist and pose a conflict between name and overall appearance (as rated by subjects). Legless “lizards” really do exist; “whales” actually resemble “fish”; flying “bats” are confusable with flying “blackbirds,” and so forth (quotes indicate the words used in the experiments). Indeed, the stimulus pictures were sufficiently realistic and accurate that adults were frequently able to identify them in the absence of experimenter labels and drew category-based inferences accordingly (Gelman & Coley, 1990). Even 4-year-old children were able to identify a subset of the animals when no labels were provided (Gelman & Markman, 1987).

3. Impoverished Stimuli? Demonstrations of category-based induction and essentialism rely on a wide variety of evidence, with stimuli ranging from line drawings to realistic color drawings to photographs of actual things. Gelman (1988) and Gelman and O’Reilly (1988), using photographs, found that children drew inferences to dissimilar category members. Gelman and Wellman (1991), using photographs, found evidence for a kind of psychological essentialism in preschool children. Rosengren, Gelman, Kalish, and McCormick (1991) employed photographs and realistic color drawings, and found that children assume that identity remains constant over growth, even when color change or metamorphosis occurs (see also Keil, 1989). Children's adherence to psychological essentialism cannot be explained simply as an artifact of deficient stimuli.

As a related point, Jones and Smith suggest that studies showing evidence of essentialism relied on pictures that were so “deeply impoverished” that by default children were forced to rely on the labels provided by the experimenter. The problem with this argument is that children did not uniformly rely on the labels. Control conditions were included in which pictures were named, but children did not use the names because the inferences were inappropriate (Davidson & Gelman, 1990; Gelman, 1988; Gelman & Coley, 1991; Gelman & Markman, 1986, 1987; Gelman & O’Reilly, 1988). Other conditions included pictures that were not named, however; subjects did use the names because they were able to detect
the category on the basis of subtle perceptual cues (Gelman & Coley, 1991; Gelman & Markman, 1987).

We are also puzzled by the argument that in the induction studies names provide obviously better clues than do pictures, but also that names are uninformative with respect to ontological type (e.g., that a toy bear is a bear). If the latter point is valid, then how would children know to make inferences based on names in Gelman and Markman's studies? What would prevent them from generalizing novel biological properties from real bears to toy bears? (We already know from Carey, 1985, that children do not draw inferences from a real animal to a toy animal.)

A final point in this regard is that category learning often takes place without direct (perceptual) experience. Children learn from syntactic information (Naigles, Gleitman, & Gleitman, in press), folktales, verbal instructions, and so forth. Although it would be interesting to vary the richness of the stimuli employed in these studies, and to determine how sensitive children are to subtle perceptual cues concerning category membership, the point remains that much of children's learning derives from picture books, bedtime stories, conversations, and other similar situations with "impoverished" visual information.

4. Disrespect for Perceptual Knowledge? Jones and Smith argue that perceptual knowledge has been denigrated—treated as insignificant, or as inferior to the deeper sort of knowledge associated with naive theories and essentialist ideas. Again, we have a different slant. First, in Gelman and Markman's work, one motivation for including perceptual similarity as a contrast to category membership (as conveyed by names), was precisely to capitalize on the demonstrated power and significance of perceptual similarity to children. It is well known that children are highly attentive to perceptual cues. If this were not true, then there would be no reason to use general appearance as a contrast in that work.

Second, we view the study of essentialism as attempting to discover the linkages between perceptual information and knowledge and beliefs. There are some deep issues and subtle problems associated with the most straightforward ways of integrating perceptually-driven and theory-based category learning (e.g., Wisniewski & Medin, 1991); we don't think that investigators in this area are at all guilty of ignoring perceptual knowledge.

A third point, however, is that there are important and systematic differences across studies in task and items, making it not very surprising that children attend to different kinds of information on one versus the other. Shape-bias studies ask children to categorize; induction studies ask children to make inferences. These are very different tasks that (normatively) take into account different kinds of information (see discussion in Gelman, Collman, & Maccoby, 1986). Moreover, shape-bias studies typically include nonmeaningful stimuli with no hidden properties. Indeed, the only kinds of information available to children on many of
these tasks were perceptual. Children would have to use perception on the shape tasks, in order to categorize at all.

5. What Drives Attention? One piece that seems to be missing from the Jones and Smith account is what the engine might be that could drive children to attend to shape, color, texture, or any other feature. Induction studies and the psychological essentialism perspective are valuable for reminding us of category function. Why do we classify as we do and not in other ways? One answer is that our categories are structured in part to foster induction, especially about properties not immediately discernible. Indeed, even the shape-bias results can be explained in these terms. That is, shape is probably a good predictor of other important properties (including internal parts, function, behavior, etc.). (See also Medin & Ortony, 1989, for a discussion of how appearance is thought to be a guide to underlying properties.)

Stability, Coherence, and Representation

Are concepts represented? What could it mean to say that concepts are not represented? If the point is that representations are meaningless to talk about without specifying associated processing principles, we quite agree. One cannot equate stability with fixed representations. Stability can arise from reliable computations, and instability can arise from unstable computations on stable representations (e.g., Barsalou, 1991; Medin & Schaffer, 1978). Indeed, a cardinal rule in cognitive psychology is that one can evaluate pairs of representation and processing assumptions, but not either entity in isolation.

Moreover, if the point is that categories and word meanings are fluid, contextually sensitive, and changing over time, we also agree. Jones and Smith provide interesting demonstrations of this point, as do many philosophers, linguists, and psychologists (Clark & Clark, 1977; Katz, 1964; Maloney & Gelman, 1987; Nelson, 1974). Nonetheless, categories are stable as well as fluid. We argue that a complete account of categorization must be able to address both stability and fluidity (see also Barsalou, 1987). Induction studies examine stability, in part reflecting the phenomenon that some concepts are lexicalized. Language is an important tool for fixing reference (at least for some time), and for concretizing the experience of past generations.

One corollary to the relative stability of language may be the assumption of category essences. It is important to clarify what is meant by “essentialism.” Psychological essentialism is a claim about folk metaphysics—beliefs about how the world is structured. An essentialist assumption is a realist assumption, that categories are discovered rather than invented. Thus, psychological essentialism is a claim about people’s (often implicit) belief systems. Essentialist assumptions appear to be common across history and across cultures. Essentialist beliefs appear to emerge in biology (Mayr, 1988), folk systematics (Atran, 1990), and everyday social concepts (Rothbart & Taylor, in press; Stoler, in press).
It is important to be clear about what psychological essentialism is not: It is not a metaphysical claim about the structure of the world, it is not a claim about the structure of concepts, and it is not a claim about the structure of word meanings. Beliefs are not the same as category structure. Indeed, according to current biological theorizing, essentialist beliefs about the concept of species may be wrong (Mayr, 1988; albeit very useful in supporting inductions about hidden (nonperceptual) properties). Essences are typically not known, almost always unobservable, and may not exist. So, the essence itself cannot usually serve as the basis of how people categorize or identify items. Indeed, Medin and Ortony (1989) suggest that often there may only be an essence “placeholder.” Rey (1983) has argued that one needs something like a “node” to provide for concept stability within and across people, and an essence placeholder is perfectly suited to perform that function.

**Perceptual Imperialism?**

The Jones and Smith analysis is valuable for forcing one to address questions about where perception leaves off and where conception begins. We agree that the sort of fuzziness that allows one factor to usurp an inappropriate explanatory role must be avoided. Jones and Smith appear to be worried that perception is not getting its fair share of credit, whereas we see a perceptual imperialism that may be stretching beyond its proper role. Indeed, if perceptual knowledge is asked to carry too much explanatory weight, it may end up paradoxically being unable to perform the role traditionally associated with it.

What do we mean by perception or appearance? Although certain types of perceptual experiences such as stereopsis evolve over time (many seconds), normally we think of perception as relatively immediate and direct in the sense that it is not mediated by reasoning or conscious computations. Sometimes people are made aware of perceptual experience by its occasional contrast with reality. For example, in the Mueller-Lyer illusion we see lines of unequal length that other computations (employing a ruler) indicate are equal in length, contrary to appearances. Why do we believe the ruler? In part, it is because we have a set of interconnected beliefs about rigid objects and procedures associated with measuring things, beliefs that have been reinforced by prior experience. Note, however, that this conceptual knowledge does not cause the Mueller-Lyer illusion to disappear—certain aspects of perceptual experience are impenetrable to certain forms of conceptual knowledge. This division of labor has the advantage of enabling us to overcome our expectations and beliefs when they conflict with reality.

Smith and Jones make the provocative and important point that perceptual similarity may be dynamic rather than fixed. It is provocative because it raises the possibility that similarity is “smarter” than we think, and that some of the contrasts between similarity and other forms of knowledge may have used an inappropriate assessment of similarity. The form of dynamism they suggest in-
volves a differential weighting of constituent features or dimensions and is nicely illustrated in their Figures 3, 4, and 5 (pp. 131, 133, and 135, respectively). In these figures they show how the Landau, Smith, and Jones (1988) results could be accommodated by assuming that a linguistic context increases the weight or significance of form features over other properties such as color or texture. That is to say, the effects of language can be thought of as mediated by a dynamic perceptual system that computes perceptual similarity in a flexible manner.

The idea that (the computation of) similarity is not fixed is a very intriguing avenue for exploring interactions between language, perception, and knowledge. Evidence suggesting that young children are less flexible about similarity than older children can be interpreted as showing that the perceptual system's conservatism serves to protect young children from forming the wrong categories (Medin, 1983). Smith (1989) showed that a formal model for this and related changes with age could account for a wide range of developmental findings. Jones and Smith extend this earlier view by suggesting that perceptual similarity is still more dynamic.

We believe that it is at least worth asking if the dynamic perceptual similarity described by Jones and Smith is still perceptual. We think the evidence from Landau et al. (1988) and related studies shows that the significance of different dimensions does change powerfully as a function of linguistic versus nonlinguistic contexts. A central question, however, is whether changes in the weighting of dimensions correspond to changes in perceptual experience. The Jones and Smith view seems to be that perceptual experience does change to reflect the difference in dimensional weighting. If so, then the implication that linguistic contexts dynamically determine perceptual experience raises some intriguing empirical questions that should be explored. Although we remain agnostic on this issue, we should point out the logical possibility that the change in weighting is the result of a slower, more conscious and deliberate weighting and ignoring of different aspects of the situation. To establish that contexts change perceptual experience, independent evidence is needed to separate strategic weighting from direct changes in perception.

If similarity is dynamically computed, then there is no single perceptual similarity that a comparison really has; there is only a space of possible similarities under different conditions. Why is this important? For us it is significant for limiting the imperialism of perceptual similarity. Note, however, that there are dangers of having perception be too penetrable—it risks having us see the world in accordance with our beliefs. We can live with rose-colored glasses; opaque glasses are more problematic.

We confess to being uncertain about how willing Jones and Smith are to link changes in dimension weights with changes in perceptual experience. It may be that they think of perceptual experience as being more stable, with dimensional weighting being more conceptual in character. In any event, one would like to have an account of the role of perceptual experience in concept formation and
use. It would be odd indeed if Jones and Smith thought of perceptual experience as irrelevant.

Finally, we feel we should add a footnote to the discussion of perception and similarity. Similarity theorists have not and should not necessarily restrict themselves to perceptual similarity. Conceptual knowledge also provides the basis for similarity. Furthermore, researchers such as Gentner (1983, 1989) have drawn the important distinction between attributional and relational similarity and have shown that they serve somewhat different functions in accessing knowledge and applying it. Thus the dialogue between similarity-based and knowledge-driven learning is both more complicated and more enriched than is implied by the simple distinction between perception and conception.

CONCLUSION

We end with the reminder that human categorization is extraordinarily varied and diverse (see also Waxman, 1991). Humans use a wide array of different kinds of concepts that vary in structure, content, linguistic expression, duration, and origins—concepts ranging from red to quark, from hypotenuse to game, from oak tree to justice. Inextricably tied to this variation is variation in the functions that concepts serve. We all seem to agree that studying concepts entails studying the contexts of their use, in all of their linguistic, perceptual, and conceptual intricacies. However, if we are to explain this variation and do more than describe (or celebrate) its complexity, it is necessary to examine how the functions and structures of categories vary systematically. We believe it is fruitful to ask, for example, how and why induction and identification lead to focus on different kinds of information; how and why concepts named by adjectives tend to differ in structure and function from those named by nouns; how and why a concept that has been encoded in language and then passed down from generation to generation (e.g., “dog”) may include different kinds of information and yield different kinds of performance than a concept newly invented for the purpose of an experiment (e.g., an abstract shape made out of wire or wood). In short, there are ample motivations for treating concepts as more than recognition routines.

REFERENCES


