Similarity and analogical reasoning

Edited by
STELLA VOSNIADOU
ANDREW ORTONY
Intraconcept similarity and its implications for interconcept similarity

LAWRENCE W. BARASLOU

A permanently existing "idea" or "Vorstellung" which makes its appearance before the footlights of consciousness at periodic intervals, is as mythical as an entity as the Jack of Spades.

William James, 1890/1950, p. 236

A central goal of cognitive science is to characterize the knowledge that underlies human intelligence. Many investigators have expended much effort toward this aim and in the process have proposed a variety of knowledge structures as the basic units of human knowledge, including definitions, prototypes, exemplars, frames, schemas, scripts, and mental models. An implicit assumption in much of this work is that knowledge structures are stable. Knowledge structures are stored in long-term memory as discrete and relatively static sets of information; they are retrieved intact when relevant to current processing; different members of a population use the same basic structures; and a given individual uses the same structures across contexts. These intuitions of stability are often compelling, and it is sometimes hard to imagine how we could communicate or perform other intelligent behaviors without stable knowledge structures.

But perhaps it is important to consider the issue of stability more explicitly. Are there stable knowledge structures in long-term memory? If so, are they retrieved as static units when relevant to current processing? Do different individuals represent a given category in the same way? Does a given individual represent a category the same way across contexts? Whatever conclusions are reached should have important implications for theories of human cognition and for attempts to implement increasingly powerful forms of machine intelligence.

The first four sections of this chapter lay the groundwork for the last two. The first section reviews demonstrations of instability in category representation, and the second reviews more systematic attempts at assessing just how unstable category representations are. The third section considers seven possible accounts of instability, and the fourth presents a retrieval-based framework for viewing the dynamic character of human knowledge. In the context of this framework, the fifth section introduces the concept of intraconcept similarity, namely, the similarity between different representations of the same category. The sixth section concludes by considering various implications of intraconcept similarity for interconcept similarity, the similarity between representations of different categories.

Demonstrations of instability

The well-known retention phenomenon of semantic encoding variability suggests that category representations are unstable (e.g., Anderson & Ortony, 1975; Anderson, Pichert, Goetz, Schallert, Stevens, & Trollip, 1976; Barclay, Bransford, Franks, McCarrell, & Nitsch, 1974; Geis & Winograd, 1975; Thompson & Tulving, 1970; Tulving & Thompson, 1973). Investigators have generally found that the cues most effective in accessing a memory trace at retrieval are those most similar to its encoding context. For example, Barclay et al. (1974) found that "something heavy" was a better cue than "something with a nice sound" for retrieving the sentence "The man lifted the piano." In contrast, "something with a nice sound" was a better cue for retrieving the sentence, "The man tuned the piano." Because weight was not an effective cue for the sentence about piano tuning, it must not have been incorporated in the representation of piano. But because weight was effective for the sentence about piano moving, it must have been. Rather than using an invariant knowledge structure to represent piano during the comprehension of the different sentences, subjects appeared to construct different representations, with the representations in different sentences focusing on weight and sound, respectively. Greenspan (1986) qualifies this result, showing that encoding variability occurs only for context-dependent and not for context-independent information, a distinction discussed later.

Numerous demonstrations of instability have been reported in the lexical priming literature. Barsalou (1982) demonstrated that certain knowledge about a category — context-dependent information — becomes active only if relevant in the current context. When people read frog in isolation, for example, eaten by humans typically remains inactive in memory. However, eaten by humans becomes active when reading about frogs in a French restaurant. Other researchers have reported similar
effects in various priming tasks, including Conrad (1978) and Whitney, McKay, and Kellas (1985). Like the retention literature, this work demonstrates that different information is incorporated into the representation of a category in different situations.

Not only do category representations vary across contexts, the properties composing these representations also vary. Halff, Ortony, and Anderson (1976) found that the representation of red varies widely across categories such as apple, brick, face, hair, light, soil, vine, and so forth. Contrary to Rips and Turnbull's (1980) suggestion that absolute properties like red should be relatively stable, even these properties are unstable. One could argue that red is stable only within particular categories rather than across all categories. But red clearly varies across particular exemplars within each of the categories just mentioned. Even one's memory of red for a particular exemplar could vary, depending on various contextual cues and subsequent exemplars encoded. As discussed by Barsalou (1987), if the properties that constitute concepts and exemplars are not stable, then it is difficult to know exactly what, if anything, is stable in human knowledge. Loftus and Loftus (1980) similarly argue that subsequent experience may continually change previously acquired information.

Kahneman and Miller (1986) review numerous findings on social decision making that can be interpreted as demonstrating instability. If one's current representation of waiter specifies that waiters are pleasant, for example, then seeing an unpleasant waiter causes surprise. However, seeing another unpleasant waiter soon thereafter does not cause as much surprise because the value for pleasantness in the representation of waiter has decreased. Because unpleasant waiters have become less surprising, the representation of waiter must have changed. Kahneman and Miller consider various phenomena in which emotions, decisions, and causal attributions vary as a function of cues in the current context and the accessibility of information in memory. They conclude that people represent categories with ad hoc representations rather than with static knowledge structures.

In the personality literature, instability of behavior has been the source of a central debate for years (e.g., Buss & Craik, 1983, 1985; Epstein, 1979; Mischel, 1968, 1983; Mischel & Peake, 1982). Rather than remaining consistent according to particular traits, people's behavior appears to be heavily determined by current context. Although it is not clear what implications such instability has for the nature of category representation, it is nevertheless consistent with the view that the human cognitive system operates in a highly dynamic manner.

Systematic assessments of instability

Whereas the studies reviewed so far demonstrate instability, studies reviewed in the next three sections assess more systematically the extent to which category representations are unstable.

Graded structure

A category's graded structure is simply the ordering of its exemplars from most to least typical. In birds, for example, typicality is generally perceived by American subjects as decreasing from robin to pigeon to parrot to ostrich. Much recent work has demonstrated that the graded structure of a category is unstable. Depending on the population, individual, or context, the ordering of a category's exemplars by typicality can vary widely. Whereas robin is more typical than swan when Americans take their own point of view, swan is more more typical than robin when they take the point of view of the average Chinese citizen. As argued by Barsalou (1985, 1987), changes in the graded structure of a category reflect changes in its current representation. This account assumes: (a) an exemplar's typicality increases with its similarity to the current category representation; (b) changes in the category representation alter its similarities to exemplars; (c) these changes in similarity alter graded structure. Similar accounts of typicality have been proposed by Smith, Shoben, and Rips (1974), Rosch and Mervis (1975), and Hampton (1979).

In most of the studies described in this section, subjects received category names followed by the names of exemplars and judged how typical exemplars were of their category. In some experiments, subjects ranked exemplars from most to least typical; in others, subjects rated exemplars from 1 to 7 for their amount of typicality. An average graded structure, as discussed shortly, was simply the average of all the individual graded structures for a category across subjects (i.e., the typicality of each exemplar was its average typicality across subjects).

A category's graded structure can vary widely across populations. Barsalou and Sewell (1984) correlated the average graded structures of Emory University undergraduates with the average graded structures of Emory faculty for the same categories and
observed an overall .2 correlation, indicating that these two populations had graded structures that were nearly unrelated. Similarly Schwanenflügel and Rey (1986) found that the average graded structures of Spanish and English monolinguals were quite different. Because a category's graded structure varies between populations, individuals in different populations appear to represent the same category differently.

Barsalou, Sewell, and Ballato (1986) explored the extent to which different individuals from the same population produce the same graded structures for a category. To assess agreement, Barsalou et al. computed the average correlation between the typicality judgments of all possible pairs of subjects for a category's exemplars. Across four experiments with American undergraduates, the average correlation was generally around .5, indicating that one subject's graded structure generally accounted for only around 25% of the variance in another subject's graded structure.

This low level of agreement was relatively unaffected by a variety of manipulations: Agreement was only slightly higher when subjects took the culturally shared point of view of the average American (.55) than when they took their own idiosyncratic points of view (.46). In another experiment, agreement was unaffected by the familiarity of the adopted point of view, with agreement being .53 for subjects' own point of view, .54 for the point of view of the average suburban housewife, and .57 for the point of view of the average country redneck. Agreement was the same when subjects judged typicality (.56) versus exemplar goodness (.53). Subjects judging 16 exemplars per category (.45) showed no less agreement than subjects judging 8 exemplars per category (.42). Agreement was somewhat higher for subjects ranking exemplars (.46) than for subjects rating exemplars (.41). Agreement was relatively unaffected by category type, with common taxonomic categories only sometimes being slightly more stable than goal-derived categories (see Barsalou, 1983, 1985, and Barsalou, Usher, & Sewell, 1985, for further discussion of these category types). Similarly Fehr and Russell (1984, Study 3) found .38 between-subject agreement for the typicality of emotion terms.

One might be willing to concede that different individuals use different representations for the same category but maintain that a given individual uses the same stable representation across situations. Consequently Barsalou et al. (1986) explored the extent to which a given individual produces the same graded structures for a category in different instances of the same context. To assess agreement, they correlated subjects' graded structure on one day with their graded structure in the same context a few weeks later. Across four experiments, they found that the average correlation was only around .8, indicating that a subject's graded structure in one session generally accounted for only about 64% of the variance in the other.

This low level of agreement was relatively unaffected by a variety of manipulations: Agreement was relatively unaffected by whether subjects took the self (.80), American (.79), housewife (.79), or redneck point of view (.78). Agreement was the same when subjects judged typicality (.80) versus exemplar goodness (.79). Subjects receiving the exemplars for each category in the same order across sessions (.85) showed no more stability than subjects receiving them in different orders (.84). Subjects judging 16 exemplars per category (.73) exhibited the same stability as subjects judging 8 exemplars per category (.74). Agreement was somewhat higher for subjects ranking exemplars (.78) than for subjects rating exemplars (.70). Agreement was relatively unaffected by category type, with common taxonomic categories only sometimes being slightly more stable than goal-derived categories. Similarly Fehr and Russell (1984, Study 3) found .55 within-subject agreement for the typicality of emotion terms.

Within-subject agreement was highly related to between-subject agreement, exhibiting an average correlation of around .75 across categories. The same factor(s) may be responsible for both between- and within-subject instability.

Graded structure also varies with context. Barsalou and Sewell (1984) found that subjects adopting different points of view generated very different graded structures for the same category. When American undergraduates judged typicality from the American and Chinese points of view, for example, they generated graded structures that were uncorrelated on the average across categories. Moreover, graded structures were inversely related for certain categories. Because between-subject agreement was comparable for the different points of view, these shifts in graded structure do not reflect random responding. Roth and Shoben (1983) similarly found that judging typicality and accessing exemplars in different linguistic contexts has a large impact on graded structure.

**Property generation**

One could argue that graded structure is a relatively indirect means of assessing category representations. Perhaps a more direct means is simply to have people describe category representations as they are thinking about them. Because people are reporting the contents of
working memory, these accounts are likely to be fairly reliable (Ericsson & Simon, 1980, 1984).

Barsalou, Spindler, Sewell, Ballato, and Gendel (1987) used a variety of tasks to explore property generation. In one experiment, some subjects were asked to produce *average characteristics* of a category's exemplars, whereas other subjects were asked to produce *ideal characteristics* that a category's exemplars should have. Subjects in each condition were given examples of averages and ideals for two categories and asked not to generate the other kind of information. In another experiment, *prototype subjects* were asked to produce information that was typically true of a category and that did not have to be definitional, whereas *definition subjects* were asked to produce a strict dictionary definition for each category that did not contain nondefinitional information. Subjects in both experiments produced responses in a written form and were asked to spend a few minutes per category; they were not to spend a tremendous amount of time, or to provide only one or two properties per category. Whereas average and ideal subjects tended to produce lists of properties (e.g., *pretty, sings, feathers, builds nests, and catches worms for birds*), prototype and definition subjects tended to produce sentence-like strings of properties (e.g., *something that people use to sit on in a house for furniture; slashes represent our division of the protocol into properties*). In scoring subjects' protocols, we attempted to combine different surface forms of the same property as much as possible, which worked against observing instability. For example, we combined *transporting, form of transportation, carrying, hauling, getting, move, takes, and travel* into a single property for *vehicle*. Across experiments and conditions, subjects produced about five properties per category on the average, reminiscent of Miller's magic number minus two (Miller, 1956).

Variability across individuals. To assess agreement between subjects, Barsalou et al. (1987) computed the average overlap in properties between all possible pairs of subjects producing properties for a category. Overlap was measured with the *common element correlation*, namely, the number of properties common to two protocols divided by the geometric mean of the total properties in each (Bellezza, 1984a, 1984b, 1984c; Deese, 1965; McNemar, 1969). Across experiments, different subjects' protocols generally exhibited an overlap of around .32, indicating that only about one-third of a subject's description overlapped with another subject's description for the same category.

Only one variable had a relatively large impact on agreement. Subjects who produced definitions and prototypes (.44) showed much higher agreement than subjects who produced ideals and averages (.20). Subjects producing definitions and prototypes may have generally reported information as it came to mind, whereas subjects producing averages and ideals may have had to perform more search and computation to produce properties (cf. Marbe's law; Woodworth, 1958; Woodworth & Schlosberg, 1960).

Otherwise this low level of agreement was relatively unaffected by a variety of manipulations: Agreement was only slightly higher for subjects taking their own point of view (.21) and the housewife point of view (.21) than for subjects taking the redneck point of view (.18). Subjects producing ideals (.22) showed more stability than subjects producing averages (.17). Ideals may be more stable because they are associated with relatively stable functional knowledge about categories, whereas averages may fluctuate with recently experienced exemplars. Stability may have been marginally higher for goal-derived (.22) than for common taxonomic categories (.18) because ideals represent goal-derived categories, whereas less stable averages also represent common taxonomic categories (Barsalou, 1985).

Subjects producing definitions (.44) showed no more agreement than subjects producing prototypes (.45). In addition, these two groups of subjects produced essentially the same properties overall and the same proportion of dictionary properties. On the basis of these null results, one could argue that no variable was manipulated. However, definition subjects took 1.41 sec longer on the average to generate a property than prototype subjects (9.40 vs. 7.99 sec per property). Even though definition subjects proceeded more carefully, they eventually retrieved the same information as prototype subjects. This pattern does not support the core-plus-identification view of categories, which proposes that category representations contain definitional cores and typicality-based identification procedures (Armstrong, Gleitman, & Gleitman, 1983; Osherson & Smith, 1981, 1982; Smith & Medin, 1981). According to this view, definitional cores, which reflect natural and logical invariants, should be more stable than identification procedures, which reflect personal experience. However, definition subjects did not show more stability or produce more definitional information than prototype subjects. Even though definition subjects were apparently more careful than prototype subjects, neither showed sensitivity to a distinction between definitional and nondefinitional information.

Agreement did not vary between well-defined concepts having necessary and sufficient conditions (e.g., *bachelor, mammal*), fuzzy concepts not having necessary and sufficient conditions (e.g., *game, furniture*), subordinates of the well-defined and fuzzy concepts (e.g., *dog, chair*), and concepts intermediate in definability between well-defined and
fuzzy concepts (e.g., window, dentist). Nor did this factor interact with whether subjects produced definitions or prototypes.

Bellezza (1984b, 1984c) also observed the stability of property generation, as well as the stability of exemplar generation (1984a). Similar to our results, he found low levels of between-subject agreement for abstract categories (.18), superordinate categories (.20), basic-level categories (.28), and descriptions of famous people (.21).

**Variability within individuals in the same context.** To assess agreement within subjects, Barsalou et al. (1987) computed the overlap between a subject's protocol for a category on one day and his or her protocol in the same context a few weeks later. Across two experiments, they generally found an average overlap of around .55, indicating that only a little more than half a subject's protocol for a category in one session overlapped with his or her protocol in the other.

Similar to between-subject agreement, subjects producing definitions and prototypes (.67) were more stable than subjects producing ideals and averages (.43). Again this may have reflected a difference between subjects reporting information as it came to mind versus searching for — and possibly computing — less accessible types of information.

Otherwise this low level of agreement was relatively unaffected by a variety of manipulations: Agreement was unaffected by whether subjects took the self (.44), housewife (.45), or redneck point of view (.40). Agreement was the same when subjects produced ideals (.45) versus averages (.41) but was higher for goal-derived (.45) than for common taxonomic categories (.41) (only one of these identical differences was significant because the two tests varied in statistical power).

Within-subject agreement was the same for definition (.66) and prototype subjects (.67), offering further evidence against the core-plus-identification view. Only minor differences occurred among fuzzy (.64), subordinate (.65), intermediate (.69), and well-defined categories (.69). Bellezza (1984b, 1984c) observed comparable levels of agreement for abstract categories (.63), superordinate categories (.46), basic-level categories (.54), descriptions of famous people (.55), and descriptions of personal friends (.58).

Similar to graded structure, within-subject agreement was related to between-subject agreement, exhibiting average correlations of .3 for average and ideal subjects and .77 for prototype and definition subjects. Again the same factor(s) may be responsible for between- and within-subject instability.

**Variability across contexts.** Property generation also varies with context. Barsalou et al. (1987) computed the average overlap between all possible pairs of subjects across contexts and compared these values to the average overlap within contexts. Subjects taking different points of view (.16) agreed less than subjects taking the same point of view (.2). Subjects producing ideals versus averages (.15) agreed less than subjects producing the same kind of information (.2). The agreement that remains in these between-context comparisons suggests that certain core information is produced for a category regardless of context.

One might assume that subjects used the same representations for property generation and graded structure. However, when Barsalou et al. (1987) compared the relative stability of the same categories across the two tasks, they found no relationship. Because different categories were highly stable for each task, subjects must have represented the same category differently in each. Individual categories varied widely in stability, indicating that lack of variability was not responsible for no correlation. If categories had not varied, the correlations of between- and within-subject agreement within tasks would have not been as high as they were. Rips (this volume) similarly demonstrates that people use different information in different tasks.

**Category membership**

One could argue that instability in property generation simply reflects random sampling of information from invariant representations. Perhaps invariant representations manifest themselves only in more logical tasks, such as determining category membership or reasoning about categories (Armstrong et al., 1983; Osherson & Smith, 1981, 1982; Smith & Medin, 1981; but see Cherniak, 1984; Rips, 1975). If people have invariant representations but use them only in more logical tasks, then these tasks should demonstrate stability. If invariant representations are used in membership decisions, for example, then membership judgments should be relatively stable. However, McCloskey and Glucksberg (1978) found a pattern of instability similar to the one reported here: Different subjects showed substantial differences in how they assigned membership for a particular category; and given subjects frequently changed their minds about whether an entity was a category member in two sessions across a 1-month period. Similar to the representations used for graded structure and property generation, the representations used for category membership judgments are unstable.
Conclusion

Taken together, these findings illustrate substantial instability in category representations. Different people do not use the same representation for a particular category, and a given person does not represent a category the same way across contexts. Instead the representation of a given category varies substantially between and within individuals. This suggests that the invariant knowledge structures that many researchers attempt to identify through scaling, property listing, and linguistic analysis are analytic fictions. Instead of being actual representations that people sometimes use in a particular context, these theoretical constructs are averages or ideals of representations abstracted across a wide variety of people and contexts. Moreover the postulation and use of such constructs tend to obscure many important mechanisms and sources of information that underlie people’s dynamic representational ability (also see Barsalou, 1987; Barsalou & Medin, 1986).

Sources of instability

What factors underlie this widespread instability in category representation? This section considers several possibilities.

Differences in knowledge

Instability between people could reflect differences in knowledge. Because different people acquire different knowledge for a category over their lifetime, they retrieve different knowledge when representing it. Differences in knowledge must certainly exist between members of different populations, between experts and novices, and even between particular individuals from the same population to some extent. Moreover these differences must certainly cause people to view categories differently on some occasions.

However this explanation does not appear to account for the instability in Barsalou et al.’s (1987) experiments. All the properties produced across subjects for a category seemed to be basic facts that would be known by all subjects. Although less than 20% of the subjects in one study produced has a beak, builds nests, or lives in trees for bird, all subjects must be familiar with these properties. Although subjects probably had idiosyncratic knowledge for the various categories, they apparently did not produce it. Instead, nearly all of the properties produced by subjects in these experiments were clearly facts and beliefs about the category known by all subjects. Some other mechanism appears responsible for the instability reviewed earlier.

Atypical exemplars

Instability between and within individuals for graded structure could reflect uncertainty about the status of atypical exemplars. Because subjects lack knowledge for these exemplars, or are unsure about their category membership, they could frequently change their minds about these exemplars’ typicality and thereby produce instability. According to this view, stability should increase monotonically with typicality.

Barsalou et al. (1986) assessed the stability of exemplars at various levels of typicality and found, across all experiments, that typical and atypical exemplars were equally stable and that both were more stable than moderately typical exemplars. Judgments for moderately typical exemplars were most likely to change across sessions within a given subject; and when judgments changed, judgments for moderately typical exemplars moved the farthest, relative to the average distance they could potentially move. Between subjects, variability was again highest for moderately typical exemplars. Consequently, atypical exemplars did not underlie the substantial instability we observed for graded structure. Judgments for exemplars at all levels of typicality frequently changed, with judgments for moderately typical exemplars changing the most.

This finding complicates unitary views of categorization that assume a single mechanism underlies typicality and category membership (e.g., Hampton, 1979; Zadeh, 1965). Because stability increases from moderate to low typical exemplars for typicality, whereas stability decreases from moderate to low typical exemplars for membership (McCloskey & Glucksberg, 1978), different mechanisms may underlie these two tasks to some extent.

Forgetting

Instability within individuals could reflect forgetting. On finding that subjects produced a smaller number of properties in Session 2 than in Session 1, Bellezza (1984c) suggested that subjects were trying to remember their Session 1 protocols but failed to remember all of them. Barsalou et al. (1987) also observed a decrease in the number of properties from Session 1 to Session 2.
Other findings argue against this interpretation. Bellezza (1984c) also found that between-subject agreement increased from Session 1 to Session 2, as did Barsalou et al. (1987). It is not clear how forgetting could improve agreement. An alternative interpretation is that subjects converge on core information across sessions. In Session 1, subjects may retrieve and report somewhat irrelevant information in the process of finding information they perceive as relevant. This initial experience may enable subjects to be more efficient in Session 2 by retrieving only relevant information. To the extent that relevant information is shared among subjects and irrelevant information is not, agreement should increase across sessions. In further support of this interpretation, Barsalou et al. (1987) also found that the proportion of dictionary properties retrieved for words increased from Session 1 to Session 2.

Barsalou et al. (1986) tested the forgetting hypothesis more directly. In one experiment, half the subjects received the exemplars for a category in the same order in both sessions, and half received them in different random orders. Almost any theory of memory would predict that receiving stimuli in the same order should provide higher retention of previous typicality judgments than receiving stimuli in different orders. Consequently, if forgetting underlies within-subject instability, more stability should be observed for subjects receiving the same order. However, order had no effect, with same-order subjects showing .85 agreement and different-order subjects showing .84 agreement.

Barsalou et al. (1986) provided one other test of the forgetting hypothesis. Half the subjects in another study judged 8 exemplars per category, and half judged 16. Again almost any theory of memory would predict that having to remember more judgments across sessions should increase forgetting. However, subjects receiving 16 exemplars (.73) were just as stable as subjects receiving 8 exemplars (.74).

As discussed in the next section Barsalou et al. (1986) also manipulated the delay between sessions from 1 hour to 4 weeks and observed a decline in stability. Notably the shape of this decline did not resemble standard forgetting functions (e.g., Bahrick, 1984; Bahrick, Bahrick, & Wittlinger, 1975; Ebbinghaus, 1885/1964; Rubin, 1982). Whereas standard forgetting functions show substantial change between 1 and 4 weeks, Barsalou, Sewell, and Ballato's function was asymptotic during this period. In general, forgetting does not appear to underlie instability.

Measurement error

Perhaps the least interesting explanation of instability is that it reflects measurement error, what Winer (1971, p. 283) defined as instability due “in part to the measuring device itself and in part to the conditions surrounding testing.” But if measurement error were responsible for the low levels of stability observed in Barsalou et al. (1986), then we should always observe these low levels regardless of the items being ranked or rated. However, Galambos and Rips (1982, Experiment 1) report a result that disconfirms this prediction. Some of their subjects ranked script actions from most to least central, where the centrality of actions in a script is analogous to the typicality of exemplars in a category (Barsalou & Sewell, 1983). Other subjects ranked these same actions by temporal position, beginning with the action occurring first in the script and concluding with the action occurring last. Subjects ranking centrality exhibited between-subject agreement of .35, which is comparable to the between-subject agreement reported earlier for typicality.

In striking contrast, subjects ranking temporal position exhibited between-subject agreement of .89, indicating that the low agreement for centrality was not due to measurement error. Instead low agreement for centrality appears to reflect less stability in the information subjects retrieved to make those judgments. Furthermore it is highly plausible that Galambos and Rips (1982) would have also found higher stability within subjects for temporal position than for centrality if they had collected those data. Given the likely possibility that subjects would show nearly perfect between- and within-subject agreement for some dimensions of categories (e.g., weight), it follows that measurement error is not solely responsible for the instability of typicality judgments. If it were, then there should be no dimension for which subjects exhibit higher stability.

Mapping judgments onto response scales

Rather than reflecting variability in category representations, instability in graded structure could simply reflect uncertainty in mapping judgments onto response scales. When ranking exemplars, subjects may perceive two exemplars as tied in typicality but be forced to give one a higher ranking randomly. Between-subject instability could result because different subjects make different random responses to
such ties. Within-subject instability could result because the same subject makes different random responses in different sessions. To test this, Barsalou et al. (1987) compared ranking of typicality to ratings of typicality. In contrast to rankings, ratings allow subjects to have ties because the same rating can be applied to more than one exemplar. As discussed earlier, however, between- and within-subject agreement were both significantly less for ratings than for rankings. Consequently the problem of exemplar discriminability does not appear to underlie instability.

Barsalou et al. (1987) provide further evidence that mapping judgments onto response scales does not underlie instability. If instability reflects subjects' difficulty in deciding which exemplar of a particular subset is more typical, then increasing the number of exemplars that subjects judge should aggravate this problem and thereby increase instability. As the number of exemplars increases, the number of ties between exemplars should increase thereby increasing the rate of random responding. As discussed earlier, however, both between- and within-subject stability were unaffected by whether subjects judged 8 or 16 exemplars per category. Moreover, number of exemplars did not interact with whether subjects were performing ratings or rankings. Again the problem of exemplar discriminability does not appear to underlie instability.

It should be noted that problems in mapping judgments onto response scales cannot explain instability in property generation, which does not involve a response scale. Moreover, patterns of instability were basically the same for property generation and graded structure (in terms of between-subject agreement, within-subject agreement, and contextual shift). If response scales were responsible for instability in graded structure, then it seems unlikely that we would observe such a similar pattern of instability in another task that does not involve a response scale.

Stochastic retrieval mechanisms

Instability for graded structure both between and within individuals could reflect random fluctuation in retrieval, or, more technically, stochastic retrieval mechanisms. Whenever an individual represents a category, every property has some probability of being retrieved, and a different random subset represents the category on each occasion.

One argument against this view is that it does not make much sense in terms of achieving everyday goals. If only a subset of information represents a category, then why would a mechanism have evolved that has a high probability of retrieving information that is irrelevant to current goals, as well as a high probability of not retrieving information that is relevant?

Barsalou et al. (1986) attempted to test the stochastic retrieval hypothesis directly by manipulating the delay between sessions from 1 hour to 4 weeks. If the stochastic retrieval hypothesis is correct, then within-subject agreement should not be affected by delay. Because properties are retrieved randomly, each session is an independent event, and sessions should not be related in any way. Consequently the delay that separates sessions should not affect the extent to which retrieved properties differ. However, within-subject agreement showed a large initial decrease from 1 hour to 1 week and then showed no change thereafter (.92 after 1 hour, .87 after one day, .81 after 1 week, .81 after 2 weeks, and .79 after 4 weeks). Contrary to the stochastic retrieval hypothesis, within-subject stability after a week's delay does not reflect random fluctuation. Other factors must be involved. One might suggest forgetting as such a factor, but this seems unlikely, as discussed in a previous section. A recency-based explanation is proposed shortly.

The .92 agreement in the 1-hour-delay condition indicates that the maximum amount of instability potentially reflecting stochastic retrieval is small. Moreover other mechanisms besides stochastic variability could underlie instability in this condition. For example, Session 1 processing could have changed category knowledge such that the knowledge underlying performance in Sessions 1 and 2 differed. This possibility receives further discussion later.

Deterministic retrieval mechanisms

Instead of reflecting random fluctuation, instability may reflect the systematic operation of basic retrieval mechanisms. Two such mechanisms — accessibility and contextual cuing — appear to provide a reasonable qualitative account of people's dynamic representational ability.

Accessibility provides a central source of instability. Even though most people in a population may have the same basic knowledge for a category (as discussed earlier), the accessibility of this information may vary widely between individuals. Because the highly accessible information for a category varies from individual to individual, different individuals retrieve different information when initially ac-
cessing it, even though they all share most of the information any one of them retrieves. People might eventually converge on the same basic information if allowed to produce properties long enough, although studies of precuing suggest that initially retrieved information would prevent convergence (Brown, 1968; Karchmer & Winograd, 1971). Accessibility may similarly underlie instability within individuals and between populations: Within individuals, everyday experience may constantly change the accessibility of category information and thereby produce instability. Between populations, patterns of accessibility may vary widely and thereby produce greater instability than found between individuals of the same population.

Contextual cuing makes category representations still more dynamic. Even though most people may have the same basic knowledge for a category, the contexts that different people experience have the potential to cue different information and thereby cause between-subject instability. Within an individual, experiencing a category in different contexts would similarly cue different information and thereby cause within-subject instability.

**Stable knowledge use**

Although accessibility and contextual cuing both produce instability, it is important to note that they can also produce stability under certain conditions. Consider the following examples. Imagine that a group of people all regularly use dial-in modems to access a computer via the same telephone number. Once everyone has memorized the number, they should show perfect between- and within-subject agreement in retrieving it, although occasional errors are a reminder of the potential for instability. Similarly imagine that an experimenter asks a group of people, "How many legs do dogs have?" We would again observe perfect between- and within-subject agreement for the information subjects retrieve.

Accessibility and contextual cuing provide a natural means of accounting for such stability. High levels of accessibility may sometimes produce high stability. Repeatedly retrieving particular information may cause it to become sufficiently accessible and integrated that it becomes retrieved as a static structure. Practicing a phone number, for example, should cause a relatively static structure to become well established in memory.

Contextual cuing can similarly produce high stability. Because highly specific cues focus on only a small amount of information in memory, they reduce variability in what is retrieved. Dog, for example, is not a specific cue because it is associated with a tremendous amount of information. Consequently a wide variety of information is retrieved that thereby produces instability. In contrast, dog conjoined with number of legs is so specific that it focuses on a single piece of information and results in perfect stability.

Much remains to be learned about how accessibility and contextual cuing simultaneously produce flexibility and stability. When are categories represented dynamically, and what purpose does this flexibility serve? When are category representations stable, and what purpose does this stability serve? Although Rey's (1983) claim that category representations are unstable appears true, so does Smith, Medin, and Rips's (1984) counterclaim that mechanisms exist for establishing stability.

**A retrieval-based framework for dynamic knowledge representation**

This section presents a more specific formulation of how accessibility and contextual cuing combine to produce instability (see Barsalou, 1987, for further discussion). According to this view, a person possesses a tremendous amount of loosely organized knowledge for a category in long-term memory. Much of this knowledge may be widely shared by a population, and its content may remain relatively stable over time within individuals. However, only a very small subset of an individual's total knowledge for a category is ever active on a given occasion to represent the category in working memory. Such subsets may contain many kinds of information, including abstracted properties, exemplars, fragments of exemplars, and fragments of intuitive theories. Although certain core information may occur in most subsets for a category, much of the information in a subset is either context-dependent or reflects recent experience. Because contexts and recent experience are rarely the same, the same subset of information for a category is rarely, if ever, activated as its representation. For this reason, instability, rather than invariance, better characterizes the representations of a category.

Following the terminology of Barsalou (1987), I use the term concept to refer only to temporarily constructed representations in working memory; concept will never refer to information in long-term memory. Instead, a concept is simply a particular individual's conception of a category on a particular occasion. And rather than being definitional — as they are often assumed to be — concepts simply provide an individual with useful expectations about a category based on long-term
past experience, recent experience, and current context (Barsalou & Medin, 1986). This usage is similar to Kahneman and Miller's (1986) use of norm, which refers to standards constructed in working memory to evaluate events.

Viewing concepts as temporary constructs by no means implies that people do not have well-established knowledge. Instead this view does assume the presence of relatively stable knowledge in long-term memory. But rather than being a collection of invariant structures used over and over again across situations, category knowledge provides the material from which concepts in working memory are dynamically constructed.  

Types of information in concepts

This formulation proposes that three basic types of information provide concepts with both stability and flexibility: context-independent information constituting conceptual cores (CI information); context-dependent information activated by the current context (CD information); and recent context-dependent information activated in recent contexts (CD_rec information). These distinctions are retrieval-based in the sense that they classify information by its accessibility and cue specificity. In addition, they are somewhat orthogonal to content-based distinctions (e.g., whether a piece of information is an ideal, an average, an exemplar, or part of an intuitive theory).

Context-independent information. A number of studies have shown that certain information is automatically activated every time a concept is constructed for a category (Barsalou, 1982; Conrad, 1978; Greenspan, 1986; Whitney, McKay, & Kellas, 1985). When reading skunk, for example, unpleasant smell is activated automatically on all occasions independent of context and incorporated into the representation of skunk. The same appears true for valuable with respect to diamond and for poisonous with respect to rattlesnake. In terms of the retrieval mechanisms just discussed, CI information is high in accessibility and low in context specificity. As suggested by Barsalou (1982, p. 87), subsequent strategic processing may later inhibit this information if it is irrelevant in the current context (as for sour in plastic lemon).

Barsalou and Bower (1980) propose that information becomes CI after it has been incorporated into a concept on numerous occasions. Frequent incorporation of a property causes it to develop an automatized status with respect to the category such that its activation becomes obligatory every time the category is processed (cf. Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Barsalou and Bower further suggest that highly discriminative properties for a category, and properties that are functionally important, are likely to be processed frequently and thereby become CI. Medin (personal communication, March 1986) further suggests that properties become CI because they are central to an intuitive theory for a category and are therefore processed frequently. Because the CI information for a particular individual depends on his or her frequent experience with a category, the CI information for different individuals can vary widely, and the CI information for a particular individual can change over time.

The graded structure and property generation studies reviewed earlier corroborate the presence of CI information in concepts. Within an individual, concepts maintain a certain amount of stability over time, as measured by both graded structure and property generation. Concepts also possess a lesser amount of stability between different individuals in the same population. One interpretation of these stabilities is that they reflect the presence of CI information. Because the same CI information is incorporated into all of a person's concepts for a particular category, some stability is maintained across contexts. Because a lesser amount of CI information is shared by members of a population, some stability is maintained across individuals.

According to this view, a category's conceptual core contains CI information, namely, highly accessible information that is incorporated into all of an individual's category representations across contexts. However, most current theories view conceptual cores differently. Some propose that conceptual cores are definitional (e.g., Armstrong et al., 1983; Osherson & Smith, 1981, 1982; Smith & Medin, 1981). Others propose that conceptual cores contain intuitive theories and idealized cognitive models (e.g., Lakoff, 1987; Michalski, this volume; Miller and Johnson-Laird, 1976, pp. 280–301; Murphy & Medin, 1985; Nelson, 1974, 1978; Medin and Ortony, this volume, Rips, this volume). Whereas these latter views can be construed as stressing the importance of category competence, the CI-property view stresses performance, focusing on those properties most frequently central to category use.

CI information may often contain information from definitions and intuitive theories. However some information from definitions and intuitive theories may be relevant so rarely that it does not become CI. The chemical composition of water (H₂O), for example, may be relevant so rarely for water that it does not become CI, even though
it is definitional. Moreover information that is neither definitional nor relevant to intuitive theories may become CI if processed frequently enough with a category. If someone frequently encounters brown dogs, for example, then brown may become CI for dog, even though it is neither definitional nor relevant to intuitive theories. Much remains to be learned about the relations among CI information, definitional information, and intuitive theories.

**Context-dependent information.** Contrary to CI information, other information becomes incorporated into concepts only because of its relevance in the current context. As proposed by Barsalou (1982), such information may become incorporated into a concept in two ways. First, CD information may be stored with knowledge about a category in long-term memory but be activated so rarely that its accessibility is far below the level necessary for context independence. Such information may be incorporated into a concept only when activated by highly associated cues in the current context. For example, edible may typically be inactive for frog and be incorporated into concepts only when cued by relevant contexts such as French restaurants. As suggested by Brooks (personal communication, December 1986), current contextual information may also retrieve category exemplars, both recent and from the distant past (Brooks, 1987; Jacoby & Brooks, 1984). The relatively inaccessible exemplars that underlie reminders may also fall within this class of CD information (cf. Ross, 1984, this volume; Schank, 1982).

Alternatively, CD information may not be stored in memory for a category but may instead by inferred. When reading about a zebra orienting to noises in the brush, for example, someone may infer ears for zebra by using a cognitive economy inference procedure: Because a zebra is a mammal, and because mammals have ears, zebras have ears. Michalski (this volume) similarly proposes that much of the information in category representations is inferred from base representations. These two formulations differ in that Michalski assumes that CD information is only inferred, whereas the view proposed here assumes that some CD information is also retrieved.

These accounts of CD information primarily concern focal context, namely, the specific tasks and information currently focal to a person's attention (e.g., reading about someone ordering frog legs at a restaurant). However, peripheral characteristics of a context may also activate CD information. Someone's current psychological state (e.g., euphoria), physical context (e.g., being in a cafe), recent mental activity (e.g., planning a vacation), and so forth, may also be sufficiently salient to act as retrieval cues and thereby select information about categories as they are processed.

A final source of CD information is the perception of exemplars. All the CD mechanisms mentioned so far involve retrieving or computing information in memory that becomes incorporated into a concept. If actual exemplars of the category are present, however, their perceptual characteristics may also become part of the CD information on that occasion.

**Recent context-dependent information.** As found in the graded structure and property generation studies described earlier, a given person shows substantial change in his or her concept for a category after a delay of a week or more, even though experimental conditions remain constant. In Barsalou et al. (1986), for example, within-subject agreement declined from .92 after an hour to .87 after a day to .8 after a week. Because the experimental contexts of the two sessions were the "same," these declines, along with instability after a week or more, probably do not reflect changes in CD information. Instead they appear to reflect another factor, CDrec information.

Once CD information becomes activated for a category, it may temporarily remain at a high level of activation such that it acquires a temporary CI status. Within a short temporal window — perhaps a day or two — this information may be automatically incorporated into every concept constructed for the category, even when irrelevant. But beyond this window, CDrec information loses its temporary CI status and once again becomes incorporated only if relevant to the current context (i.e., it returns to being CD). For example, if someone had frog legs at a French restaurant one evening, an encounter with a frog in the back yard the following day might bring the edibility of frogs to mind. But encountering a frog a week after consuming frog legs may no longer do so. Examples of such a recency effect have been reported in the social cognition literature, where subjects are more likely to use a trait during impression formation when it has been recently activated (e.g., Bargh, Bond, Lombardi, & Tota, 1986; Higgins & King, 1981; Wyer & Srull, 1986).

CDrec information may also be created indirectly. Instead of being created by the construction of previous concepts for a category, it may also be created by the construction of concepts for related categories. A recent encounter with a vicious dog, for example, may temporally elevate the activation level of viciousness in concepts for other animals. Evidence that events perturb more knowledge than directly applies has been reported in the frequency literature (e.g., Barsalou...
whilst we think, our brain changes, . . . and its whole internal equilibrium shifts with every pulse of change. The precise nature of the shifting at a given moment is a product of many factors. . . . But just as one of them certainly is the influence of outward objects on the sense-organs during the moment, so is another certainly the very special susceptibility in which the organ has been left at that moment by all it has gone through in the past. Every brain-state is partly determined by the nature of this entire past succession. Alter the latter in any part, and the brain-state must be somewhat different. . . . It is out of the question, then, that any total brain-state should identically recur. Something like it may recur; but to suppose it to recur would be equivalent to the absurd admission that all the states that had intervened between its two appearances had been pure nonentities, and that the organ after their passage was exactly as it was before.” [James, 1890/1950, p. 234]

The following example for frog shows how these three types of information account for the effect of delay on within-subject agreement for graded structure when context is held constant. Consider the 1-hour-delay condition. When a subject constructs a concept for frog during Session 1, CI information becomes automatically incorporated by virtue of its high accessibility (e.g., green, hops). CD_{rec} information that reflects recent processing also becomes incorporated automatically. Edible, for example, might be CD_{rec} if frog legs had been consumed recently. CD information activated by focal context becomes incorporated. If the subject’s task is to classify exemplars into biological classes, for example, then focusing on class with respect to frog will activate amphibian, which might be CD for many subjects. Finally, CD information activated by peripheral context becomes incorporated. If the previous category judged were insects, for example, then the subliminal activation of insects that remained might activate eats insects for frog.

When the subject constructs another concept for frog an hour later in Session 2, the same CI information is incorporated as in Session 1 (green, hops). However, the CD_{rec} and CD information, although highly similar to their counterparts in Session 1, differ somewhat. First consider CD_{rec} information. The CD_{rec} information incorporated in Session 1 (edible) is also incorporated in Session 2, because both sessions are roughly within the same temporal window for the CD_{rec} information that accrued prior to Session 1. However, additional CD_{rec} information is also incorporated, namely, the CD information from Session 1 that was activated by peripheral context (eats insects). Consequently the CD_{rec} information is not identical in the two sessions.

CD information behaves similarly to CD_{rec} information, showing both constancy and change. Because experimental conditions were held constant in both sessions, the focal context should be the “same,” and focusing on class should activate the same CD information (amphibian). In contrast, it is likely that peripheral context has changed such that new CD information is activated in Session 2. If the previous category judged were musical instruments, for example, then the subliminal activation of musical instruments that remained might activate the sound “rivet” for frog.

According to this account, the high stability within a subject over an hour reflects three sources of common information: CI information, CD_{rec} information that accrued prior to Session 1, and CD information activated by focal processing. The small amount of instability at this delay reflects two sources of unique information: CD_{rec} information in Session 2 that resulted from peripheral context in Session 1 and CD information resulting from a new peripheral context in Session 2.  

At longer delays, this account assumes that changes in CD_{rec} information are primarily responsible for instability. As delays become longer than the lifetime of CD_{rec} information, this information only serves to make concepts different, whereas within its lifetime it makes concepts similar. To the extent the CD_{rec} information active for Session 1 (edible) has become inaccessible and been replaced with new CD_{rec} information, the CD_{rec} information active for Session 2 will be different. Because within-subject stability stops decreasing after a week’s delay, the lifetime of CD_{rec} information appears to be a week or less.

Different CD information activated by peripheral context may also play a role in decreasing stability. At very short delays, peripheral context may remain relatively constant and thereby cause stability, whereas it may differ at longer delays and thereby cause instability. Because a person’s peripheral context changes so quickly, however, this factor may affect stability only at very short delays.
Factors affecting the volume of particular information types

Within an individual, the volume (amount) of CI information in his or her concepts for a category should not vary much across contexts. Because it takes much experience to establish and change this information, it should change slowly over time such that the same amount generally enters into all concepts for a category (although intense experience with a category, such as on-the-job training, may cause rapid change).

Between individuals, the volume of CI information in the concepts for a category may vary widely. Consider Figure 3.1(a) in which the width of a box represents the volume of information. As depicted there, someone who once owned a dog should have developed much more CI information about dogs than someone who never owned a dog (or who has but little experience with them). As is also depicted, this person has not had much recent experience with dogs and therefore has the same amount of CD\textsubscript{rec} information as someone who never owned a dog (and who also has not had much recent experience with dogs).

Another variable that probably affects the volume of CI information is population. Because a certain population may make extensive use of a category, its members may generally develop a larger volume of CI information for the category than members of another population that makes less use of the category. For example, people from Michigan may have a larger volume of CI information for snow plow than people from Georgia, and even more so than people from New Guinea.

Contrary to CI information, the volume of CD\textsubscript{rec} information in a person’s concepts can change substantially from occasion to occasion. Consider the example in Figure 3.1(b). On the left is a person’s concept prior to ever owning a dog. On the right is that person’s concept after recently acquiring her first one. Although she has not had the dog long enough to increase the amount of CI information for dog, she has established a large amount of CD\textsubscript{rec} information. Presuming that the dog is around for a long time, she will continue to have a large volume of CD\textsubscript{rec} information that will be continually changing, at least somewhat. The volume of CD\textsubscript{rec} information can vary widely both within and between individuals, depending on their recent experience.

Similar to CD\textsubscript{rec} information, the volume of CD information in a person’s concept can change substantially from occasion to occasion. Consider the example in Figure 3.1(c). On the left is a concept for dog held by someone who has never owned a dog, who has not had much recent experience with dogs, and who is reading the word dog in no specified context. “No specified context” in this and later examples means reading a word or perceiving an exemplar in isolation (e.g., in the middle of an otherwise empty computer screen). In such situations, not much contextual information exists to activate CD information focaly.

On the right of Figure 3.1(c) is the same person’s concept after being asked to think about dogs from the Eskimo point of view. Because this person has previously acquired knowledge about the kinds of dogs that Eskimos have, about the roles these dogs play in Eskimo culture, and so forth, this information exists in his knowledge about dogs. However because he has not thought about this knowledge much in the past, and has not thought about it recently, Eskimo-related knowledge about dogs is not present in the CI and CD\textsubscript{rec} information for the concept. When Eskimo is processed as the focal context, however, it cues this information by virtue of associative structure in memory. As a result, the total volume of CD information increases.

Intraconcept similarity

The first two sections demonstrated that concepts are unstable, and the next two presented a theoretical framework for viewing people’s dynamic ability to represent categories. The final two sections explore implications of this framework for similarity.
Theories of similarity traditionally address the similarity of concepts for two different categories, what I will refer to as intercategory similarity. If category representations are unstable, however, then theories of similarity must consider another kind of similarity: the similarity between different concepts for the same category, what I will refer to as intracategory similarity. Before discussing either intra- or intercategory similarity, it is first necessary to state several assumptions regarding similarity in general.9

**Assumptions about similarity**

Based on the analysis in previous sections, the similarity of two concepts depends on the amount of overlap between their CI, CD\textsubscript{rec}, and CD information. For example, the two concepts juxtaposed along the vertical line on the left side of Figure 3.2(a) have less overlap for three kinds of information than the two concepts on the right (where overlap is represented by the amount of shaded area). Consequently the concepts on the left are less similar than those on the right.

However, the total volume of each kind of information must also be considered when assessing similarity. Although the two concepts on the left of Figure 3.2(b) have the same amount of overlap as the two concepts on the right, the total volume of the two concepts on the right is much greater. Following Tversky (1977), similarity is also a decreasing function of the amount of unique information in each concept, as well as the amount of information in common. Consequently the two concepts on the right are less similar than those on the left, even though their overlap is the same.

Also following Tversky (1977), various factors can cause subsets of information to be weighted differently. The relative weighting of common and unique information, for example, can vary across contexts. In addition, the relative weighting of CI, CD\textsubscript{rec}, and CD information may also vary. For example, when people judge typicality from their own point of view, they may weight CD\textsubscript{rec} information highly to maximize the contribution of their personal experience (assuming that similarity of exemplar and category concepts underlies typicality). In contrast, when taking the American point of view, people may weight this information closer to zero to minimize the contribution of personal experience. To simplify matters, nothing further will be said about weighting. However it is certainly an important factor that a more complete analysis would have to consider.

**Factors affecting intracategory similarity**

Cultural transmission of knowledge may largely involve establishing shared cores of CI information for categories within members of a population. For example, our culture may generally establish a core for cow that in part contains information about cows being a source of meat. Other cultures may transmit different cores for the same category. For example, the core in cow for citizens of India may contain information about cows not being a source of meat and about their religious significance.

As a baseline, then, intracategory similarity should generally be lowest (although not necessarily) between individuals from different populations. As shown in Figure 3.3(a) there may generally not be much overlap on the average between two individuals from different populations with respect to CI, CD\textsubscript{rec}, or CD information (e.g., as in Barsalou & Sewell, 1984; Schwanenflugel & Rey, 1986).
CD overlap may increase because exemplars are more similar within populations, because actions performed on exemplars are more similar, and so forth. CD overlap may increase because the associative knowledge that underlies CD activation may be culturally bound, similar to CI knowledge. For example, people in the United States and the Soviet Union may have different stereotypes for Eskimos. Consequently, when thinking about dogs from the Eskimo point of view, these two populations may activate somewhat different CD information.

We turn now from populations to consider individuals. Because different people in a population often have different experiences with a category, they may develop different CI information to some extent. Consequently, instances of a concept constructed by the same person should be more similar than instances constructed by different people. As suggested by Figure 3.3(c), virtually the same CI information should be incorporated into a person’s concepts for the same category across occasions. In support of this, we have generally found substantially higher agreement within subjects than between subjects, both for graded structure (.8 vs. .5) and for property generation (.55 vs. .32).

Again this gain in CI information may be accompanied by gains in CDREC and CD information. CDREC overlap may increase because exemplars are more similar within individuals, because actions performed on exemplars are more similar, and so forth. CD overlap may increase because the associative knowledge that underlies CD activation may be somewhat idiosyncratic.

Within the same individual, recent experience may increase intraconcept similarity in two ways. First, if a person constructs two concepts for the same category within the short lifetime of particular CDREC information, then the presence of this information in both concepts should increase their similarity. When focal context also remains constant, intraconcept similarity may be quite high. As depicted in Figure 3.3(d), these may be the conditions under which intraconcept similarity is highest. As suggested by Barsalou (1987), such recency effects may facilitate goal-directed behavior by maintaining conceptual constancy while a particular goal is being achieved.

Second, recent experience may increase intraconcept similarity within the same individual across a long time period if similar events occurred prior to constructing instances of the same concept. For example, concepts constructed for dog on two occasions a year apart may be similar if the person was bitten by a dog just prior to both.

Similar recent experience may also increase intraconcept similarity...
between individuals. If two people were recently bitten by a dog, for example, then their concepts for dog should be more similar than if their recent experiences had differed. Analogously, if two people experience intense recent training with a concept, such as in a laboratory or on the job, then this should increase not only the proportion of CD_{rec} overlap but also the total volume of CD_{rec} overlap. As can be seen in Figure 3.3(e), this could substantially increase intraconcept similarity between individuals.

Context affects intraconcept similarity through the activation of CD information. To the extent contexts activate the same information, either within or between individuals, intraconcept similarity should increase. As shown in Figure 3.3(f), for example, reading a word in isolation may activate relatively little CD information, whereas reading a word in the context of a point of view may activate much CD information that is completely unrelated to the CD information in the other context. Conversely, constructing a concept from the same point of view on two occasions should result in a substantial amount of shared CD information, thereby causing a large increase in intraconcept similarity.

In summary, intraconcept similarity depends on the overlap of CI, CD_{rec}, and CD information. Population and individual subject factors may have large effects on the overlap of CI information and accompanying effects on the overlap of CD_{rec} and CD information. Recent experience is primarily responsible for the overlap of CD_{rec} information, which is maximized within individuals at short delays and between individuals who have had similar recent experiences. Current context is primarily responsible for the overlap of CD information. Depending on the composition of these factors, intraconcept similarity can range from extremely high to extremely low. Assuming invariant concepts for categories not only is unjustified but also obscures the presence of dynamic and important representational mechanisms. As discussed next, it also distorts interconcept similarity and obscures similar mechanisms there.

Implications for interconcept similarity

Theories of similarity generally address the similarity of concepts for two different categories and generally seem to assume that invariant concepts represent these categories (e.g., Krumhansl, 1978; Shepard, 1962a, 1962b; Tversky, 1977). But if the concepts representing a particular category are unstable, then it follows that the similarity of two categories is also unstable. Contrary to an implication of traditional theories, the similarity of two categories is not invariant.

Barsalou's (1982) Experiment 2 provides direct evidence that the similarity of two categories varies across contexts. Raccoon and snake, for example, were much less similar when judged in no explicit context than when judged in the context of pets. Because pets activates similar CD information in raccoon and snake, their number of common properties increases, which thereby increases their similarity. Barsalou (1982) also showed that this effect does not occur when a context is relevant to CI information. The similarity of robin and eagle, for example, was the same in no explicit context as in the context of birds. Because the common properties relevant to birds are context-independent, they are active in the absence of the birds context.

Moreover, interconcept similarity varies when context remains constant. Hutchison and Lockhead (1977) obtained .55 between-subject reliability and .73 within-subject reliability for subjects' judgments of interconcept similarity (e.g., the similarity of chicken to eagle, of plum to robin). These levels of agreement are reminiscent of those reported earlier for graded structure and property generation, indicating that concept instability also has a substantial effect on interconcept similarity.

It should be noted that the weighting of properties typically provides theories of similarity with potential for generating variability in interconcept similarity (e.g., as in Tversky's, 1977, diagnosticity principle). For example, by assuming that many properties in a concept are weighted to zero in a particular context, and by assuming that the properties weighted to zero can vary widely in different situations, the content of concepts can functionally change across contexts. However, the actual content of the concept remains the same, unless a further assumption is made that properties weighted to zero are not part of the concept in that context. Generally speaking, most theorists have not viewed weighting as a means of varying the content of concepts but instead have viewed it as a means of varying the effects of an invariant set of properties.

Factors affecting interconcept similarity

All the factors that affect intraconcept similarity also affect interconcept similarity. As shown in Figure 3.4(a, b, c), the primary difference between intra- and interconcept similarity is the average amount of
Intraconcept and interconcept similarity

gories. Most important, the concepts for dog and raccoon may be more
similar than the two concepts for cow, such that interconcept similarity
is higher than intraconcept similarity. Dornbusch, Hastorf, Richardson,
Muzzy, and Vreeland's (1965) classic study of interviewers' percep-
tions of campers demonstrates a similar effect in social cognition.
They found that a given interviewer's personality assessments of two
different campers were generally more similar than two interviewers'
assessments of the same camper.

Figure 3.5 demonstrates how various factors influence interconcep-
tsimilarity. These factors generally parallel those for intraconcept
similarity in Figure 3.3. As a baseline, interconcept similarity
should generally be lowest between concepts constructed by mem-
ers of different populations. As shown in Figure 3.5(a, b), simi-
arity should increase when constructed by members of the same
population because of gains in CI overlap. Similar to intraconcept
similarity, such gains may also be accompanied by gains in CD,rec
and CD overlap. Interconcept similarity should generally increase
further when concepts for two categories are constructed by the
same individual as seen in Figure 3.5(c). Again, such gains may ref-
lect gains in CI information, along with accompanying gains in
CD,rec and CD information.

Similar recent experience with two categories, either within or
between individuals, can further increase interconcept similarity.
As shown in Figure 3.5(d), for example, someone who recently be-
gan work in a pet store may have had similar recent experiences
with boas and dogs such that concepts constructed for these cate-
gories are similar at that point in time. As shown in Figure 3.5(e),
once having worked in a pet store — but not recently — may also in-
crease similarity, but in this case by increasing the overlap of CI
information. Finally, interconcept similarity can be increased
substantially by current context as in Figure 3.5(f). For example, if
someone who has never worked in a pet store views boa and dog in
the context of pets, the increase in CI overlap will increase inter-
concept similarity.

In summary, interconcept similarity, like intraconcept similarity,
depends on the overlap of CI, CD,rec, and CD information. The pri-
mary difference between these two kinds of similarity may simply be
that intraconcept similarity is generally higher than interconcept simi-
arity, although not necessarily. Otherwise, intra- and interconcept
similarity may both be produced by the same cognitive processes and
reflect the same basic variables.
Averaging data across subjects and across trials within the same subject is standard practice in empirical studies of human knowledge. But given the substantial instability of concepts, what do these averages represent? According to the analysis here, the answer depends on what is averaged. Consider averaging the typicality judgments of different subjects in the "same" context. As shown in the left two panels of Figure 3.6(a), the concepts constructed for an exemplar by different subjects overlap on CI information shared by the population, have no CD_{re} overlap, and overlap on CD information shared by the population. Assume for simplicity that the effects of nonoverlapping information on typicality are uncorrelated (i.e., nonoverlapping CI effects are uncorrelated, all CD_{re} effects are uncorrelated, and nonoverlapping CD effects are uncorrelated). Then typicality averaged across subjects should reflect only overlapping information between subjects, as shown in the right panel of Figure 3.6(a). In other words, averaging across subjects nets the effects of population CI information for the exemplar plus population CD information. All other information has no effect.

Next consider averaging multiple typicality judgments made by the same subject in the "same" context at two times more than a week apart. As shown in the left two panels of Figure 3.6(b), concepts constructed for the same exemplar should have a complete CI overlap, no CD_{re} overlap, and a partial CD overlap. Again assume for simplicity that the effects of nonoverlapping information on typicality are uncorrelated (i.e., all CD_{re} effects are uncorrelated, and nonoverlapping CD effects are uncorrelated). Then typicality, as averaged across sessions, should reflect only overlapping information, as shown in the right panel of Figure 3.6(b). In other words, averaging across trials nets the effects of the individual's CI information for the exemplar, plus the individual's CD information that is affected only by focal context. All other components have no effect.

Averaging across subjects and contexts obscures the effects of certain sources of information in concepts. What remain are the effects of information common to the particular concepts observed. There may be certain situations in which we would like to know this information, and averaging may be a good way to determine it. To the extent we assume that this information represents a person's concept in a particular situation, however, we are misinterpreting data. If we wish to identify actual concepts constructed by an individual, and the
relations between them, we cannot do so by averaging across individuals or contexts. Instead we must assess concepts in specific individuals in specific contexts.

Two other points should be noted. First, averaging does not necessarily identify CI cores. As suggested in Figure 3.6, averaging may net not only CI information occurring across individuals and contexts but also CD information occurring across contexts. Moreover, if subjects have had similar recent experience with a category, such as laboratory training, then averaging may also net CD_rec information across individuals. The point is that using averaging to identify CI cores may be difficult, and care should be taken to cancel out all other sources of information. Second, averaging may not yield accurate estimates of information occurring across contexts or individuals when the effects of nonoverlapping information are correlated. Under these conditions, it is impossible to know what averages mean without knowing the nature of the correlations.

Specific applications

When the inputs to scaling procedures such as multidimensional scaling and hierarchical clustering procedures are data averaged across subjects and contexts, only information common to the particular concepts observed enters into scaling solutions. Nonoverlapping information between individuals and contexts will not have an effect, to the extent the effects of this information are uncorrelated. Although these procedures may be valuable for identifying common information across individuals and contexts, they do not provide complete information about any individual's concepts in any particular context.

Another scaling procedure, INDSCAL, does exhibit sensitivity to the concepts of particular subjects by identifying weightings that individuals place on dimensions. When the inputs to this procedure are averages across contexts, however, the basis of solutions will only be information common to the individual's concepts across contexts. Nonoverlapping information unique to specific contexts will not be captured, to the extent its effects are uncorrelated. In contrast, when the inputs are an individual's nonaveraged data from specific contexts, the solution will reflect all the information in concepts constructed during the experiment. Because of changes that can occur in CD_rec and CD information, however, care should be taken in generalizing these results to other contexts for that subject.

In comparing group versus individual points of view, Barsalou and Sewell (1984) found that the average graded structures constructed by undergraduates taking the faculty point of view were identical to the average graded structures constructed by faculty taking their own point of view. Sewell (1985), however, found that specific individuals were far from perfect at constructing graded structures from the point of view of a close friend. How could the average graded structure of faculty and undergraduates taking the faculty point of view be identical when the graded structures of close friends were highly divergent? One possible account concerns the effects of averaging. Averaging across subjects may eliminate the effects of all information not common to the population. Consequently, if information common to one population is the same as information common to the other, then the effects of this information on average graded structures should be identical, because all nonoverlapping information cancels
out. In contrast, when comparing the nonaveraged concepts of individuals, idiosyncratic information is not canceled out, and differences between individuals are observed. The apparent identity of population averages is therefore misleading: A concept held by a member of one population is not identical to a concept held by a member of the other—in fact, there is probably substantial nonoverlapping information. The illusion of identity exists only because averaging eliminated the nonoverlapping information.

In Barsalou, Sewell, and Ballato (work in progress), subjects taking their own point of view generated average graded structures for common taxonomic categories that were identical to those of subjects taking the American point of view. However, subjects taking their own point of view showed lower between-subject agreement, suggesting that these subjects' judgments were more idiosyncratic. Again the equivalence of average graded structures reflects nonoverlapping information being canceled by averaging. Even though the self condition had more nonoverlapping information than the American condition, this difference was not captured by average graded structures.

Conclusion

Categories are not represented by invariant concepts. Different individuals do not represent a category in the same way, and a given individual does not represent a category in the same way across contexts. Instead there is tremendous variability in the concepts that represent a category. This variability engenders the consideration of intracategory similarity, namely, how instances of the same concept are related. Intracategory similarity depends on (a) whether concepts are constructed by members of the same or different populations, (b) whether concepts are constructed by the same or different individuals, (c) recent experiences with the category, and (d) current context. Because concepts for the same category are not invariant, neither is intercategory similarity. The similarity of the concepts for two categories can vary widely as a function of the same factors that determine intracategory similarity. Accurate assessments of how an individual in a given context perceives the similarity of two concepts cannot be based on data averaged across either individuals or contexts. Averaging tends to obscure important sources of information in concepts and to leave only sources that are constant across the concepts observed. Though obtaining such information may be of value for certain purposes, it should not be construed as completely representing a person's concept in a particular context. As James (1890/1950, p. 236) states:

No doubt it is often convenient to formulate the mental facts in an atomistic sort of way, and to treat the higher states of consciousness as if they were all built out of unchanging simple ideas. It is convenient often to treat curves as if they were composed of small straight lines, and electricity and nerve force as if they were fluids. But in the one case as in the other we must never forget that we are talking symbolically, and that there is nothing in nature to answer to our words.

NOTES

I am grateful to Lee Brooks, J. Wesley Hutchinson, Douglas Medin, Andrew Ortony, Jayna Spindler, and especially Stella Vosniadou for helpful comments on previous drafts of this chapter. I am also grateful to Jayna Spindler, Daniel Sewell, Susan Ballato, and Elaine Gendel for their contributions to the research described herein, and to Ulric Neisser for pointing out that William James had much to say about instability. Work on this chapter was supported by Grant IST-8308984 to the author from the National Science Foundation. Correspondence and reprint requests should be sent to Lawrence W. Barsalou, School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332.

1 For examples of relevant literature, see Armstrong, Glei
tman, and Gleit-

2 Although people make extensive use of stereotypes such as American, house-
wife, and redneck, and even though such stereotypes may lead to accurate inferences on some occasions, they often lead to inaccurate inferences and offer a narrow and prejudiced view of the world. Our use of stereotypes is not meant to condone their use. Instead we have chosen to study stereotypes because of the extensive role they play in cognition (Fiske & Taylor, 1984).

3 Rosch (1975) and Armstrong et al. (1983) report that between-subject agreement is .9 and higher. As discussed by Barsalou (1987) and Barsalou, Sewell, and Ballato (1987), these values do not actually represent between subject agreement but instead represent the stability of average typicality judgments. Because the reliability of averages increases monotonically with the number of subjects in a sample, a .9 level of agreement can be obtained with a modest number of subjects, even though between-subject agreement may actually be quite low.

4 It is important to note that retrieval has been central to previous theories of knowledge. Quillian (1968), Collins and Loftus (1975), and McCloskey
and Glucksberg (1979), for example, all assume continuous retrieval of category information during classification and question answering; that is, information is retrieved until a sufficient amount is available to make a response. Moreover the first two of these theories, through intersecting search, have some capability of handling context effects (as a conjunction of probes). Exemplar models are also oriented toward the importance of retrieval in representation and have much potential for explaining instability (e.g., Brooks, 1978, 1987; Hintzman, 1986; Jacoby & Brooks, 1984; Medin & Schaffer, 1978). In general, however, none of these theories has yet addressed issues associated with between-subject instability, within-subject instability, and contextual shift.

5 CD information at Session 2 that resulted from peripheral context in Session 1 (see inserts) need not necessarily be a source of instability. Because this information was also present during Session 1 as CD information, it could increase stability rather than decrease it. This depends on the extent to which memory of CD information from Session 2 is the same as its initial activation in Session 1.

6 Previous theories have discussed the similarity of a concept to itself, which could be construed as a kind of intraconcept similarity (e.g., Krumhansl, 1978; Tversky, 1977). Because this previous work has addressed a stimulus' overall confusability with other stimuli, however, rather than how instances of the same category are related, it has addressed a very different kind of intraconcept similarity than the kind addressed here.

REFERENCES


