17

The Structure of Experience

BARBARA TVERSKY, JEFFREY M. ZACKS,
& BRIDGETTE MARTIN HARD

Continuous Sensation/Discrete Mind

One way to regard the body is as a moving set of sensors, continuously capturing light, sound, smell, touch, heat, and more from the surrounding world. Yet more sensors are inside the body, capturing information from the body's own movements and processes. Comprehending everyday action and experience requires integrating that information and making sense of it. Despite the fluctuating flow of information to our senses, the impression is of a stable world. From the ever-changing multimodal stream of information, the mind carves fixed entities, organizing and integrating sensations of light, sound, smell, and touch into entities that are distinct from other lights, sounds, smells, and touches. The perception is not just of separate entities but of specific objects and organisms, each with its own shape, size, and parts. Although the sensations are continuous and changing, the impression is discrete and enduring. Activity, too, is discretized. Although the perception of activity is one of change over time, the change is thought of as chang*es*—not as constant change but as sequences of key moments.

10/6/2007   8:49:23 PM

Why the mind discretizes is a question that has answers on many levels. On the neurological level, neurons fire or don't. On the cognitive level, the continuous input is so rich and complex that much of it must be, and is, ignored; the input must be categorized to be effectively processed and understood. The categories the mind forms are not random; rather, the mind organizes information into packets that are easily recognized on the one hand and informative on the other. Such packets are useful not just for understanding what is happening but also for predicting what will happen. This case has been made forcefully for objects (Rosch, 1978; Tversky & Hemenway, 1984). As shall be seen, basic objects such as tables and dogs and violins are easy to detect by their shapes. These perceptual packets are not only readily discriminated but also readily associated with significant information about the objects—their behavior and functions. This correspondence between the appearance of things and their behaviors or functions provides a way for people to learn the information essential to organizing and planning their own behavior. It enables a working hypothesis that things that look alike behave alike and things that look different behave differently. Of course, this is a working hypothesis, a starting point, and the world presents both examples and counterexamples. Ultimately, the link between perceived features and functions or behavior renders the world more comprehensible and predictable. The link from appearance to function or behavior has been established for objects. One question raised here is whether the link holds for other important entities, notably events.

Ongoing experience is discretized in multiple ways. The perceptual world is parsed into distinct scenes, objects, and people. Interactions among people and objects are segmented into events. Discourse is decomposed into sentences, phrases, words, and sounds. As is evident from these examples, people do more than extract whole entities from ongoing experience; they go on to divide these wholes into parts. People perceive discrete objects and decompose them into discrete components (e.g., Hochberg, 1978). Similarly, people perceive discrete events and decompose them into discrete stages (e.g., Newtson, 1973; Newtson & Engquist, 1976; Zacks & Tversky, 2001; Zacks, Tversky, & Iyer, 2001). The process of partitioning is significant for many reasons: knowing the parts of a whole, how the parts are determined, how they are related, and what they do is a crucial part of understanding the whole. This analysis leads to a series of questions to be considered here:

- *Wholes:* How are wholes determined—that is, how are they distinguished from backgrounds?
- *Parts:* How are wholes partitioned into parts, and on the basis of what kind of information? Parts may be further partitioned into subparts; do the same bases for partition hold for the subparts?
- *Configuration:* How are the parts of a whole arranged?
- *Composition:* Each whole entity has a set of parts, which may be parts of other wholes as well. How does the entire set of parts get distributed to wholes?
- *Perception-to-function:* Are there relations between perception and appearance on the one hand and behavior and function on the other?

Although our central concern is the structure of events, considering the structure of other categories—specifically language, objects, and scenes—provides insights by comparison. Moreover, these categories interact in actual experience. Objects, scenes, and events are the primary categories forming the stage for human activities. They also compose central topics of language as it is used: people are usually someplace in some activity with something. Objects have served as the paradigm example of entity since antiquity (e.g., Casati & Varzi, 1996, 1999; Jacob & Jeannerod, 2003; Quine, 1985) as their many uses, including the concept "reify," attest. Objects are typically inseparable parts of events, and events invariably occur in scenes. But before we begin, a short detour into alternative ways of organizing knowledge, by parts and by kinds.

Partonomies and Taxonomies

As indicated, the focus here is event partonomies. Partonomies are hierarchies formed by "part of" relations. The human body is a prototypical example. Some of the body's parts are arms, legs, head, feet, hands, chest, and back. Hands, in turn, have parts: thumb, fingers, palm, and back. Partonomies contrast with another hierarchical organization of the world: taxonomies (see Miller & Johnson-Laird, 1976; Tversky, 1990; Tversky & Hemenway, 1984). Taxonomies are hierarchies formed by "kind of" relations. A familiar example is the animal kingdom: vertebrates and invertebrates are kinds of animals; fish, amphibians, reptiles, birds, and mammals are kinds of vertebrates. In a taxonomy the

same individual is simultaneously in all the classes superordinate to it, so a robin is a bird, and also a vertebrate, an animal, and a living thing. Given that an individual belongs to so many nested categories, a question that has fascinated psychologists is how to choose the level for reference (Brown, 1958; Rosch, 1978). It turns out that one level is preferred across many contexts and tasks, perceptual, behavioral, and linguistic: the *basic level*, the level of bird and chair (Rosch, 1978). The basic level bridges perception or appearance with function or behavior, allowing inferences from one to the other; moreover, parts distinguish the basic level, linking partonomic and taxonomic organizations (Tversky & Hemenway, 1984).

Language

Wholes and Parts

We begin with language, whose structure and organization have been studied for generations. Indeed, the structure of language has served as an instructive analogy for that of bodies, scenes, and events, and vice versa. Here, we overview the features of language that have served the analogies. Language has distinctive characteristics—at the levels of frequency, phonology, and more—that allow it to be distinguished from the background of other sound. Language decomposes into parts on many levels: discourse has as parts sentences or utterances, utterances have as parts words or morphemes, morphemes have as parts sounds or phonemes. Each higher level serves as the whole for a more elementary level. The bases for segmenting into components as well as the rules of combination of components change at each level, and in fact vary with the individual language. The level of phonemes (units of sound) and the level of morphemes (units of meaning) are most relevant here.

Phonemes: Configuration, Composition, and Perceptual-Function Links

For every language, there is a small set (20 to 40) of phonemes that combine to make words (see Chapters 4 and 10 in this volume). There are strong perceptual correlates for phonemes, so much so that continuous changes in sounds are perceived to have categorical boundaries corresponding to phonemes. Phonemes are at the same time a unit of speech

perception, as well as a unit of speech production. This, along with other findings, has led some to claim that the same perceptual-motor mechanisms that underlie production of speech also underlie perception of speech—that is, that we understand speech through the motor mechanisms that produce it (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985). Every language has rules for arranging phonemes. These rules of combination do not allow certain sequences of phonemes, such as *tv* (in English), at the beginning of words, but they are free enough to allow far more combinations than any language is likely to need, even with a small alphabet. Phonemes are indeed combined in a multitude of ways, challenging poets and delighting readers. Within languages, the sequencing constraints are strong enough that there are statistical dependencies for sequences of phonemes that even infants and other new language learners rapidly pick up, providing a basis for distinguishing words (e.g., Brent, 1999; Saffran, Aslin, & Newport, 1996; Saffran, Newport, Aslin, & Tunnick, 1997).

Morphemes: Configuration, Composition, and Perceptual-Functional Links

Despite their complexity, distinguishing and configuring phonemes seems easy in contrast to distinguishing and configuring morphemes or words. The perceptual basis for distinguishing words from utterances is multifaceted and language-dependent, as anyone acquiring a new language can confirm. Discerning individual words or morphemes in spoken language relies not only on the phonemes of the particular word but also on the phonemes of the surrounding morphemes, the syntactic structure, and the semantic context. Languages have a large vocabulary of morphemes or, loosely, words, numbering in the tens of thousands. The rules of combining morphemes—the syntax—are intricate and constrained, expressed sometimes in arrangements of words and sometimes in inflective changes within words. There are statistical dependencies in arrangements of morphemes just as there are in the arrangements of phonemes (e.g., Landauer, 1998; Landauer & Dumais, 1996; Miller, 1963). We *peel* apples but not books, just as we *read* books but not apples. The statistical dependencies or redundancies at both levels, phonemes and morphemes, may facilitate comprehension as well as production. In contrast to phonemes, it is not straightforward to tie

mechanisms used to distinguish morphemes with those used to produce them.

The cases of phonemes and morphemes have been revealing. Within language, the principles of segmentation, composition, and configuration vary depending on the level at which language is analyzed. Each creates large numbers of wholes, but differently. Phonemes combine to create many morphemes by using a small number of elements and loose constraints on combination. Morphemes combine to create many utterances by using a large number of elements and tight constraints on combination. Phonemes have a strong perceptual basis, tightly linked to production; not so for morphemes.

As we have seen, language is structured on many levels—on the level of sound, on the level of meaning, and on the level of discourse. These levels cooperate and interact but are not completely reducible (e.g., Clark, 1996). Language serves many human activities. One important service of language is providing a means for describing and remembering the things, places, and activities that occur in the world. We turn now to those.

Objects

Wholes: Distinguishing Objects from Backgrounds

Thinking about partonomies brings us to the first question about objects, an old one that continues to challenge researchers: how are objects partitioned from a scene—that is, how are figures distinguished from ground? More simply, what makes a good object? To answer that question, the Gestalt psychologists proposed principles for organizing perception, providing insights that continue to fascinate artists and scientists alike. Good objects are more likely to have closed, continuous, convex contours; they are also more likely to have parts with a common fate—that is, parts that move together (e.g., Hochberg, 1978; Peterson, 1994; Spelke, Gutheil, & Van der Valle, 1995). Contours, if presupposed, are nevertheless key to object integrity: contours that are continuous and closed, especially under movement, suggest that what is contained by the contour has an existence independent of the background. This is not to say that there are necessary and sufficient conditions for objecthood; there are borderline cases and cases that are ambiguous in context, and these are provocative, puzzling, and illuminating. Despite

such ambiguities, many common objects can be recognized from their contours, especially at canonical orientations (e.g., Palmer, Rosch, & Chase, 1981; Rosch, 1978)

Parts: Partitioning Objects

Partitioning objects from backgrounds leads to the next question: how are objects themselves partitioned? There is more than one way to partition an object: an object can be partitioned into the stuff it's made from; it can be partitioned into its sides, front, back, top, and bottom; it can be partitioned into the pieces it breaks into when it falls. Here, we are interested in a different sense of part from any of those; we are interested in what might be called *compositional* parts or *integral* parts, the kinds of parts that people name when asked to give the parts of an object, say a body or a car. Because external, visible parts are those available to direct perception, we are not interested here in internal parts such as hearts and lungs.

The clues the Gestalt psychologists provided for distinguishing objects seem to apply to distinguishing parts of objects; what makes a good part is also what makes a good object—continuity, closure, convexity, and common fate. This suggests that the same principles that underlie discriminating objects appear to underlie discriminating parts of objects—that is, the same perceptual features that serve as clues to wholes should serve as clues to parts. Although the features that make an object good seem also to be the features that make a part of an object good, the analyses of object segmentation have come from perspectives other than Gestalt. Some analyses of object segmentation have been inspired by Attneave's (1954) observation that natural boundaries are likely to be points of large changes in information. For objects, one important change in information is relative discontinuity in contour, particularly local minima in the curvature of contours (e.g., Hoffman & Richards, 1984; Hoffman & Singh, 1997). These local minima occur at the junctures of parts—for example, where the fingers attach to the hand or where the arms and legs attach to the body. The parts picked out by these local minima in curvature are relatively closed, continuous, and convex. Good parts have a perceived independence—detachability from their wholes analogous to the detachability of objects from scenes, if not in actuality then in perception.

Compelling as this view is, it has limitations. For one thing, objects have an infinite number of contours, depending on the point of view. Some points of view are far easier to recognize than others, notably those that show the critical features of the objects (Palmer, Rosch, & Chase, 1981). Because objects are typically three-dimensional, they may have parts that do not affect the contour from certain views—for example, noses from frontal views of faces. Even so, a frontal view of a face will provide some information about the shape of the nose; that is, even a single view of an object has clues that reveal the three-dimensional structure of the object. Thus, the visual system has more to work from when parsing objects than inflection points in contours.

Another approach to object partition relies on local convergences of edges irrespective of viewing angle. As Biederman and his collaborators noted in the "recognition by components" theory of object recognition, the visual system is sensitive to a host of local properties of object contours, such as lines at various orientations, pairs of lines, vertices, convex curves, and more (Biederman, 1987; Hummel & Biederman, 1992). These and other attributes are called nonaccidental properties, as they are likely to arise from enduring features of objects rather than accidents of perspective. Groupings of such attributes activate *geons*, generalized cones that form shapes such as cylinders, blocks, wedges, and cones. Geons can take many meanings, depending on their size and their configuration in objects. A cylinder might be the leg of a person or chair or an ear of corn. A block might be a brick or a layer of a stupa. A curved cylinder might be the handle of a coffee cup or a piece of macaroni. An ovoid might be a Brancusi head or an egg.

Composition: Components of Objects

Biederman (1987) has drawn analogies between phonemes and geons. Just as phonemes are the building blocks of words, geons are the building blocks of objects. Just as there is a small set of phonemes that can be combined to form the words of a language, there also is a small set of geons that can be combined to form all objects. Just as phonemes vary depending on the neighboring phonemes, geons vary depending on the neighboring geons. For example, whether a *p* is aspirated or not depends on the neighboring phonemes. However, the particular characteristics of geons—size and shape—seem to depend more on global than local

features of objects. The size and the shape of a curved cylinder seem to derive from qualities of an entire object, even its function, whether it's the spout of a teapot or the handle of a suitcase. As is the case for phonemes, there appear to be statistical dependencies among parts of objects (e.g., Rosch, 1978; Malt & Smith, 1982). Animate things that have legs also have heads with eyes and mouths; things that have feathers also have wings and beaks. Certainly the integrity, and hence recognizability, of parts (or geons) is important in object recognition. When portions of line drawings of contours of objects are deleted so that part boundaries—the nonaccidental properties—are intact, objects are more readily recognized than when the same amount of contour is deleted at part boundaries (Biederman, 1987).

From Perceptual Parts to Functional Parts

Geons are perceptually defined object parts, but geons may or may not correspond to the parts people name when asked to list parts of objects. The horizontal and vertical components of the tail of an airplane are probably two geons, but they form a single part. When asked to rate object parts for "goodness," people give high ratings to those parts that have contour distinctiveness and functional significance, such as legs of pants or tables, either for the object or for the user of the object (Tversky & Hemenway, 1984). Parts that are perceptually distinct also have different functions; people hold the handle of a knife and slice with the blade; they blow into the mouthpiece of a clarinet and open and close its holes with their fingers. The legs of a chair or a person or a pair of pants have different functions from the seat of a chair or the arms of a person or the waist of pants. In some cases, the appearances of parts give clues to their functions. For example, long, thin parts are likely to be good for reaching, and flat, horizontal ones of a certain size and height are likely to be good for putting or for sitting. The very names of parts suggest links from perception to function: *seat*, *leg*, and *handle* refer sometimes to appearance, sometimes to function, sometimes to both; on many occasions it is not clear, or doesn't even matter, whether part names refer to appearance or to function.

Connecting geons, the perceptual parts of objects, to the functional parts of objects is not straightforward and may not be feasible. Functional parts such as *handles*, *legs*, *bodies*, and *frames* have a broad range

of specific forms, differing in geons. The seat of a bicycle bears little resemblance to the seat of an armchair. The leg of a chair may be close to a cylinder, but the leg of a horse and the leg of a crab are not. This points to a difference between geons and phonemes. Phonemes are at once parts of perceiving language and parts of producing language. Geons are perceptual parts but not functional ones. Moreover, the building blocks of objects may be closer to functional parts than to perceptual parts. Rabbit legs and camel legs look different, but they serve rabbits and camels in similar ways, respectively, just as bicycle seats and armchair seats look different but serve humans similarly.

Configurations of Parts

Whatever their view on the status of object parts, most approaches recognize that an object is more than just a collection of parts; the parts must be properly arranged. A pile of arms, legs, torsos, and heads is a pile of parts, not a set of bodies. Names of parts reflect the significance of configuration; many part names derive from their position in a configuration (*top*, *bottom*, *middle*, *side*). As for phonemes and morphemes, or morphemes and phrases, the organization of the parts of objects is critical to meaning. A highly constrained configuration of parts does not appear in all domains, as shall be discussed in the analysis of scenes.

Objects of all kinds are all around us, but they are not distributed helter-skelter. If there's a refrigerator, there's probably a sink and a stove nearby. Objects that serve related ends typically appear together in contexts, specifically in scenes.

Scenes

Wholes and Parts

The first thing to notice about scenes in contrast to objects is that they don't have shapes or clear boundaries. Scenes are the contexts for objects, the grounds from which objects are distinguished. They are also the contexts for events. Scenes typically surround us, include us. Perhaps for this reason, the problem of distinguishing a scene from its background has not occupied psychologists. The background would have to be an even larger context encompassing more than one scene. The related problem of recognizing scenes, distinguishing one scene from another,

has occupied psychologists, and scene recognition is surprisingly quick, requiring less than a second of exposure, even for schematic line drawings of scenes (Biederman, 1981).

Scenes don't seem to have shapes, in contrast to objects. This being the case, how are they distinguished and recognized? What features of scenes underlie their rapid recognition? There are clues from research on scene taxonomies and partonomies (Tversky & Hemenway, 1983). In that research, one group of participants was asked to list categories of scenes and subcategories of scenes. For those scenes that were frequently mentioned, other informants listed parts of scenes and activities performed in scenes. The top-level categories were *indoors* and *outdoors*; outdoor scenes included beaches, forests, and cities; indoor scenes were schools, restaurants, and stores, each of these with subcategories. Could it be that scenes are recognized by their parts? The parts of scenes informants listed were the objects that are common in them: sand and water for beaches, desks and tables and blackboards for schools. Informants also listed activities appropriate for different scenes. Activities were the things people do in scenes: hike in forests and eat in restaurants. The features of scenes—that is, their appearance—and the activities performed in scenes—that is, their functions—are linked. For example, swimming and boating go with water and sand, writing and reading go with desks and blackboards, and eating goes with tables and dishes.

Scenes, then, are both different from and similar to objects. Gestalt features like closure or continuity, or features analogous to them, do not partition the world into scenes. Unlike objects, scenes don't have shapes; they seem to be recognized by the kinds of things they contain—the objects large and small—and not shapes.

Configuration, Composition, and Perceptual-Functional Links

For objects, the spatial arrangement of the parts is highly constrained: the legs of a chair must be under the seat, at far corners from each other; the back of the chair must be above the seat and at its edge. Similarly, the legs of a giraffe must be below its body and at far corners. The spatial arrangement of the parts of a scene is more loosely constrained, partly by gravity and the physics of the world, partly by function. For schools,

desks should be on the ground, blackboards on the wall, lights on the ceiling. Desks should have chairs nearby and chalk should be near the blackboard, but the exact configuration of objects in scenes is not as constrained as, say, the configuration of the parts of a desk (e.g., Mandler & Parker, 1976). A kitchen needs a stove, refrigerator, sink, cabinets, and countertops, but the relative positions of the major appliances and the overall shape of the kitchen do not matter within a large range. Scenes have a potentially large number of parts, but they appear in a correlated fashion. Schools have desks and chairs and books and chalkboards, and supermarkets have produce and canned goods and aisles and cash registers. Thus sharp changes in information may distinguish one scene from another, but the information changes are in the objects and activities, not in anything analogous to a contour. The features that allow scenes to be recognized so quickly are not perceptual features like contour minima or geons but rather larger features, objects, which have been interpreted and assigned meanings. Like object parts, scene parts, mostly objects, have different appearances and serve different functions.

Scenes are characterized not just by the objects appearing in them but also by the activities occurring in them. In fact, the characteristic objects determine the activities. Stoves, refrigerators, countertops, and tables support cooking and eating. Desks, chairs, books, and chalkboards support teaching and learning. Just as object parts afford actions, linking perception and function, so scene parts (objects) afford actions and link perception and function.

Scenes, like objects and words, are rapidly recognized, a strong indication of their perceptual distinctiveness and significance. Scenes have characteristic parts, typically objects, but the configuration of those parts is loosely constrained. Moreover, recognition of parts does not seem to occur prior to recognition of scenes (e.g., Biederman, 1981). Although the features underlying the rapid discrimination and recognition of scenes are not yet well understood (e.g., Epstein & Kanwisher, 1998; Henderson & Hollingworth, 1999), there does appear to be a part of the cortex selective for recognizing scenes (Epstein & Kanwisher, 1998). Scenes link objects and activities; they provide the settings for objects and for human activities, with different objects and different activities associated with different scenes. We now turn to human activities, from categories existing in space to categories existing in time.

Events

Our lives are a string of events, from the mundane—for example, going to a movie—to the extraordinary (e.g., getting married). One way of looking at events is as segments of time, analogous to objects as segments of space—but this view is misleading. It is misleading first because events have a spatial status in addition to a temporal one, and objects have a temporal status—buildings are constructed, remodeled, destroyed, reconstructed—as well as a spatial status. But the view is misleading for yet another reason: objects aren't merely segments of space, they exist in space and also in time; what's more, their positions in space can change. Similarly, events aren't segments of time; they exist in time and also in space; events as types rather than specific episodes can also change their spatial–temporal positions. A wedding, a meeting, or a parade can be held in many places or times. Events contrast with activities; running is an activity, but running a race is an event. Events have been characterized as achievements or accomplishments; as such, events are associated with outcomes as well as processes. Thus, as Casati and Varzi (Chapter 2 in this volume) argue, it is appropriate to regard both objects and events as entities with internal spatiotemporal structure, not as homogenous regions. Within the structure, some parts are more central to function than others—the seat of a chair is more central than the armrests, and blowing out candles is more central to a birthday party than pouring juice.

Wholes and Parts: Distinguishing and Partitioning Events

Just as a scene can contain many objects, it can also contain many events, like a three-ring circus. Think of a busy parent preparing dinner after work, monitoring a toddler, answering the phone, setting the table, and chatting with a spouse about the events of the day. Intuition suggests that each of these events can be comprehended, and the actions associated with each distinguished, so that reaching for the phone is not usually confused with chopping the vegetables, either in enactment or in perception. Of course, these events are familiar to us; how those unfamiliar with such events would separate them awaits investigation. As will be seen, what probably allows partitioning a scene into separate, coherent events is that events are typically characterized as sets of related actions on the same object or associated objects.

Event Contours

Do events have shapes, as objects do? Or are events like scenes, without clear shapes? For objects, contour serves as a one-dimensional description, one that is powerful, if incomplete; it is the boundary between an object and the surrounding world, but for observers this boundary exists only from a particular perspective. A candidate for an analogous one-dimensional description of an event is an activity contour, the moment-to-moment change in amount of physical activity over time. By analogy, an activity contour can be regarded as the "boundary" between the activity of the event of interest and the background activity. Abrupt changes, either increases or decreases, in moment-to-moment levels of activity may signal changes in event parts, just as contour discontinuities are a clue to object parts. Why might this happen? Event segments correspond to completions of goals and subgoals (Zacks, Braver, et al., 2001). After a goal or subgoal is accomplished, such as putting on a sheet or scrambling an egg or buying a movie ticket, there might be an increase in activity in preparation for the next subgoal or goal. Or, after a large task is finished—for example, vacuuming the living room—there might be a pause, a slowing down, before another task is begun—for example, washing the clothes. Either way, there would be a dramatic change in level of activity. Seen this abstractly, events, like objects, can be partitioned at many levels and still be made of the same stuff: activity. This one-dimensional summary ignores the numerous qualitative differences that characterize events and their segments. However, the question raised here is whether there is any psychological validity to an event contour.

One common way to study parts of events is to ask observers to segment films of events, such as a person making a bed or assembling a saxophone, into parts as they watch them (e.g., Newtson, 1973; Newtson & Engquist, 1976; Hard, 2006; Hard, Tversky, & Lang, in press; Zacks, 2004; Zacks, Tversky, et al., 2001; see Chapter 15 in this volume for an overview). In many experiments, observers have been asked to segment events at two levels on separate viewings: coarse and fine. In *coarse* segmentation, they are asked to identify the largest segments that make sense, and in *fine* segmentation, the smallest. There is remarkable agreement across and within observers on locations of segment boundaries, called *breakpoints*. Knowledge about events is hierarchically

organized; that is, boundaries of fine units coincide with boundaries of coarse units more often than could occur by chance (e.g., Hard, Lozano, & Tversky, 2006; Zacks, Braver, et al., 2001). Hard et al. developed a variant measure of hierarchical organization indexing the frequency with which the corresponding fine unit occurs at or before the related coarse unit; this *enclosure* measure better reflects the containment of fine units in coarse units.

Observational studies of naturally occurring behavior suggest that sharp perceptual discontinuities form the basis for identifying parts of events. In a large in vivo study aimed at capturing what ordinary people do on ordinary days, observers recorded people's behavior throughout the day in "behavior episodes." These behavior episodes corresponded to events at varying levels of granularity: eating a meal, reading a book, crossing a street. The changes from one behavioral episode to another were characterized as changes in the "sphere" of behavior—verbal to social to intellectual, for example— changes in the active part of the body, changes in the object interacted with, changes in the spatial direction, changes in the tempo, or changes in the behavioral setting or scene (Barker & Wright, 1954). These are physical changes that signal changes in the nature, especially the purpose, of the activity.

This project was observational, and the observations were insightful, but the approach was atheoretical. It did not consider the possibility that these kinds of changes may be correlated—for example, that different parts of the body may be active in different spheres and with different objects. For objects and scenes, key features are correlated. Nor did this approach distinguish different breadths or levels of events. The event of getting through a day can include making a bed, going to work, and eating in a restaurant, and eating in a restaurant includes the events of ordering, eating, and paying. The discontinuities at different levels may well differ in quantity as well as quality of activity, and evidence from several studies of a variety of events suggests that they do.

In one project directed at investigating the link between perceptual change and event boundaries, participants segmented everyday events filmed from a single camera angle several times, each time at a different level of granularity (Hard, 2006). Still frames were selected from the filmed events at 1 s intervals and filtered for contours, eliminating irrelevant factors such as ambient lighting and yielding sharpened images of people and objects on backgrounds. The pixel-to-pixel change from

frame to frame was computed. Comparisons of the segment boundaries to the pixel-change measure revealed that segment boundaries corresponded to large physical changes—in fact, to local maxima. Averaging the relative pixel change over all the coarse breakpoints for all the events yielded a regular event contour: a sharp rise in pixel-to-pixel activity at each coarse breakpoint, followed by a slow decline. The pattern was the same for the finer breakpoints, but far less dramatic. Thus, the coarse event segment boundaries were physically distinctive from the other captured moments of the events. In a companion experiment, participants watched the still frames from the video in sequence, free to examine each slide as long as they liked. Looking time was longer at breakpoints, even controlling for pixel change. This finding shows that breakpoints elicit heightened attention, suggesting that they are especially informative. The finding also makes it apparent that high relative change, while characteristic of perceived event boundaries, is not the only factor contributing to their meaning.

The dramatic correlation between event boundaries and relative degree of pixel change suggests that discontinuities in activity might allow people to segment events that are novel or difficult to understand. A physical basis for segmentation would allow observers to bootstrap the perceptual information to segment novel events into parts to start making sense of them. Another project has addressed that process. That project required segmenting films of abstract events in which geometric figures moved in ways that were difficult to comprehend on first viewing but became meaningful after repeated viewings (Hard, Tversky, & Lang, in press). One video was based on the well-known film of Heider and Simmel (1944) in which a large geometric figure is perceived to bully and chase two smaller ones who taunt the larger one; another was based on hide-and-seek. Observers segmented the videos at both coarse and fine levels. Observers' verbal descriptions of what happened in each segment indicated that these interpretations were not evident on first viewing. However, after viewing the animations five times and writing a narrative describing them, most observers were able to interpret the actions as a related sequence of intentional actions.

In spite of the differences in interpretation, observers both familiar and unfamiliar with the animations segmented the events the same way, suggesting that movement change rather than comprehension was the basis for segmentation. To ascertain whether event boundaries

corresponded to changes in physical activity, the videos were coded by type of movement rather than measuring pixel-to-pixel change. The movements coded were when the geometric figures stopped, started, turned, and so on. As before, though with a different measure, the quantity of movement changes distinguished breakpoints. Analogous to line junctures and line angles in objects, the nervous system seems tuned to such changes in motion: stops, starts, changes in direction, changes in velocity, etc. Coarse segment boundaries were associated with more changes in movement, as for the previously discussed work on human events. This makes sense; completing a relatively large goal should be associated with greater change in activity than completing a subgoal.

A third project, using yet different stimuli and different measures, also showed correspondences between degree of physical change and event segment boundaries. This project compared moment-to-moment movement of objects and event segmentation quantitatively (Zacks, 2004). Participants viewed movies of simple abstract animations in which the movements of two geometric objects were determined either by two people playing a video game or by a stochastic algorithm. As they watched, participants segmented the movies into fine or coarse segments. The two-dimensional trajectories were analyzed to provide a detailed quantitative characterization of the object's movement, including the velocity and acceleration of each object, the distance between the objects, and their relative velocity and acceleration. Features of the objects' movements were associated with observers' segmentation in all conditions. The most predictive features were discontinuities: local minima in the distance between the objects and points of high acceleration. Thus, three projects using different stimuli and different measures provide support for the claim that activity contours are correlated with event boundaries, and that segment boundaries tend to occur when there are sharp changes in amount of activity. Event contours bear some analogies to object contours.

Conceptual Influences on Event Segmentation

As we have seen, there is compelling evidence for a bottom-up, perceptual basis for event segmentation, though it is by no means a complete account. The fact that events can be segmented fairly well on the basis of the perceptual input alone simplifies the task of understanding complex

human activities. The correspondences between changes in physical activity and event boundaries, impressive as they are, are nevertheless insufficient for identifying events or for understanding them. Top-down knowledge of goals and causes also influences event segmentation. These influences are apparent even in unfamiliar, abstract events that are difficult to interpret, such as those studied by Hard, Tversky, and Lang (in press) and by Zacks (2004). Those movies can be interpreted as actions of agents relative to each other and to the environment, but such interpretations must be achieved without the rich set of cues available in real-life behavior. In them, the geometric figures are without faces, bodies, or limbs, and the context is minimal. These studies also suggest that social interactions and intentional states are important for partitioning events. In the project just described, showing relations between motion change and event segmentation, participants commonly reported looking for the achievement of goals in order to decide when to segment activity (Zacks, 2004). One experiment provided a focused look at how the segmentation of random movements is influenced by observers' beliefs about the intentionality of the activity (Zacks, 2004, Experiment 3). One group of observers was told that the activity was generated randomly, but the other group was told that the activity was generated by humans trying to achieve goals. For the group who thought the activity was randomly generated, the movements of individual objects were the best predictors of event segmentation, particularly moments of high acceleration. For the group who thought the activity was intentional, the distance between objects was the single best predictor of segmentation. This suggests that when people believe activity is intentional, they are sensitive to features of the activity that are relevant to the intentions of the actors—in this case, configural movement features that capture the objects' interactions. Similarly, in Hard et al.'s (in press) study, participants segmented abstract events by relying on the same discontinuities that predict changes in goals and intentions for people, such as initiations of movement, reorientations, and contact with objects. In this case, when the events had been viewed five times as opposed to one, intentional descriptions of the actions increased from 45% to 75%.

A final piece of evidence (Schwan & Garsoffky, 2004) that event breakpoints correspond to discontinuities in physical or conceptual information comes from a manipulation analogous to eliminating contour information at part (geon) junctures or between part junctures

(Biederman, 1987). In Biederman's studies, recognition of objects declined when part junctures were eliminated. Schwan and Garsoffky (2004) tested memory for filmed events that had deletions at event boundaries or deletions between event boundaries. Later memory of the events was poorer when frames were deleted at event boundaries than between event boundaries—more evidence that event boundaries are especially informative.

Together, these projects highlight an important fact about the relationship between physical change and conceptual interpretations: they often are correlated, and so have discontinuities in the same places (Zacks & Tversky, 2001). In other words, when one goal is completed and another initiated, there is a change of activity.

Events by Feet and Events by Hands

The schematic films discussed so far were of motion paths of geometric figures, interpreted as agents moving around in environments. Thus they can be viewed as "events by feet," in contrast to events that can be viewed as "events by hands," such as making a bed or doing the dishes (Tversky, Zacks, & Lee, 2004). The actions that make up events by feet are relatively simple; they consist of whole-body movements and can be summarized by a moving dot or a stationary line. A paradigmatic example is the route taken from home to work. Thinking of events by feet as cumulative paths inspires comparisons of event paths to object contours (Shipley, Maguire, & Brumberg, 2004). Here the analogy from object contours is to the actual path inscribed by the event, not to the more abstract degree of activity. Inflection points in both would presumably underlie segmentation. However, a naked path is not the same kind of summary of an event by feet as a contour is of an object. The line inscribing the outline of an object is closed, summarizing a three-dimensional form, in the best case, of a canonical view (e.g., Palmer, Rosch, & Chase, 1981). The line inscribing the path of an agent is an abstraction of a path of motion, not an entity in and of itself. On a finer level, the meaning of changes in line direction, or inflection points, is different for object contours than for event paths. For objects, the changes in contour reflect the internal structure of the object, its inherent parts. For events, the changes in path are a consequence of the intentions of the agent in an environment, for example, to approach or avoid

or accompany other agents or features of the environment. Without clues as to the external terrain, a path can not be interpreted (Gelman, Durgin, & Kaufman, 1995). Thus, an object contour can be understood on its own, but an event path can not. The differences between object contours and abstractions of motion paths are revealed in studies by Shipley and colleagues (2004) in which segmentation of the two types of stimuli differed.

Event Contours and Object Contours

This difference between lines that inscribe event paths and object contours illustrates a fundamental difference between objects and events. Actions that are parts of events typically occur with respect to something else, usually an object; that is, of necessity they involve not just an action but also an object that is acted on or with respect to. For paths, actions occur with respect to objects in the external environment—for example, turning at landmarks or chasing another agent. Studies of events by hands will draw out this point further. The actions that compose events performed prototypically by hands are far more complex than those performed by feet, entailing intricate interactions with objects and object parts rather than simple turns.

As we have seen, event contours, whether conceived concretely as paths of motion or more abstractly as activity contours, bear analogies to object contours. Sharp changes of contour signal new parts for both. However, there is a yet another critical disanalogy between object and event contours: objects can normally be recognized by their contours— that is, when the view is canonical and recognition occurs at the basic level, for example, the identification of or differentiation between chairs, giraffes, and trees (Palmer et al., 1981; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Events such as going to the store or making a bed, by contrast, can not be identified or recognized by either their path contours or their activity contours. Research suggests that recognition of everyday events depends on articulated actions on objects in scenes.

Action–Object Couplets as Event Parts

Earlier, we described research showing that event segments selected by observers correlate with points of relatively large changes in overall

activity. However, event parts marked by activity changes are not in themselves meaningful event segments. Descriptions of event segments while segmenting reveal that there is more to segmenting events than quantitative changes in action. For these studies, common basic-level events were chosen (e.g., Hemeren, 1996; Morris & Murphy, 1990; Rifkin, 1985). In one project, observers segmented basic-level events such as making a bed or assembling a saxophone at both coarse and fine levels (Zacks, Braver, et al., 2001). Some observers described what happened in each segment as they segmented. As before, there was considerable agreement across observers in event boundaries, event levels, and event descriptions. The descriptions were illuminating. More than 95% of the descriptions were actions on objects: "he spread the sheet," "she attached the mouthpiece." In effect, the descriptions were of achievements of subgoals. The descriptions also distinguished coarse and fine levels. The changes from one coarse segment to another were primarily changes in objects: the bottom sheet, the top sheet, the pillowcases. In contrast, the changes from one fine segment to another were primarily changes in actions on the same object: "he spread the bottom sheet," "he tucked in one corner," "he tucked in another corner," "he smoothed the sheet." These play-by-play descriptions of ongoing action also correspond to event descriptions produced from memory of the films (Zacks, Braver, et al., 2001) as well as to descriptions of generic events such as going to a restaurant or visiting a doctor (Bower, Black, & Turner, 1979).

This evidence reveals that events are understood as action–object couplets. The entire set of action verbs used in the descriptions of four very different events was not large; that is, the same verbs were used in many different contexts (Zacks, Braver, et al., 2001). Common verbs include *putting*, *taking*, *lifting*, *inserting*, *pushing*, *pulling*, and *spreading*. These action verbs do not describe components of events the way nouns alone can describe components of objects and entire objects. Verbs are relational terms (e.g., Gentner, 1981; see also Chapters 7 and 8 in this volume); a list of action verbs alone is difficult to understand. Consider, for example, the following list: *take*, *spread*, *fold*, *put*. Without knowledge of the objects being acted on, we can not know if this is about baking a cake or putting away the laundry (in Chapter 8 of this volume, Maguire and Dove argue that this is why verbs are harder to learn than nouns, as Gentner argued earlier). Folding flour into a batter and folding a sheet, or spreading icing on a cake and spreading a sheet on a

bed, are achieved with very different movements of the body, as well as different auxiliary objects. Indeed, the very movements depend on the object. Events are not simply partitioned into movements, they are partitioned into action–object couplets (this observation provides support for the contention of Casati and Varzi in Chapter 2 of this volume that objects play a special role in determining the structure of events, whereas events may determine objects only in a weaker sense). Neither movements nor objects alone suffice as parts of events. This fact means that event parts differ in a fundamental way from object parts. Event parts include two ontologically different categories: movements on the one hand and objects on the other.

Perception-to-Function Hypothesis

Just as for objects, for events perception is connected to function through parts. Parts of events are typically actions on objects—that is, they include perceptually identifiable behaviors as well as perceptually identifiable objects. At nearly every stage there is an accomplishment, a goal, a function. That is how events and event parts are understood—as a sequence of accomplishments. For unfamiliar events, this may be effortful. The work on unfamiliar events suggests that with repeated exposure, the actions that are first perceived as movements come to be perceived and comprehended in terms of goals—in other words, functions (Hard et al., in press).

How might event segmentation support understanding? Segmenting reduces the amount of information into manageable chunks. Segmenting appears to occur naturally, even under passive viewing of everyday events (Zacks, Braver, et al., 2001). The perception-to-function hypothesis proposes that a large change in activity signals that something important has happened. Increased inspection of what happened in the segment reveals clues as to the nature of what has happened. The illuminating clues are actions on objects. With increasing familiarity, actions on objects come to be understood as completions of goals or subgoals. The bootstrapping from large perceptual changes to functional understanding in events parallels processes linking part structure to part function in objects (Tversky, 1989). This reasoning can be extended to scenes as well, where the parts are objects, and the objects present give clues to the likely behaviors and functions.

Composition of Event Parts

Are action–object couplets like the phonemes of the sound system of language or like the morphemes of the meaning system of language? In other words, is there a small number of them that are used in many combinations to form different events, as phonemes combine to yield an abundance of words? Or is there a large number that are used in correlated fashion as words combine to utterances? The idea that events consist of parts that have internal coherence and can be excised and reassembled in different temporal configurations is appealing. It has been proposed and has received some support through cartoons (Avrahami & Kareev, 1994; see Chapter 4 in this volume) and primate behaviors (Byrne, 2002). Classical ballet is to some extent composed that way: a sequence of steps that have names and are used and reused to create many different dances. But for typical events that fill human lives, it appears that there is a large vocabulary of action–object couplets and that there are strong correlations between the parts that co-occur. Making a bed involves a different set of actions, objects, and object–action couplets than going to a restaurant or seeing a doctor. Mixing and matching the parts won't create sensible events. Actions and objects constrain each other; not every action can be applied to every object. Actions constrain the objects they can be applied to; eating a meal entails a different set of objects than making a bed or assembling a saxophone. Conversely, different objects afford different actions; balls and Frisbees can be tossed, bananas and bread can be sliced, milk and wine can be poured. Object–action couplets co-occur in events. On this analysis, events appear to be more like scenes than like objects or language. Both have a large set of components that aren't mixed and matched but, rather, appear in a correlated fashion (see also Chapter 4 in this volume). Scenes like schools have a different set of parts—objects and activities—than beaches, and beaches than movie theaters.

At a higher level, events can be arranged and rearranged to some extent. The bed can be made before or after fixing breakfast (or not at all). These events, along with many others, constitute the larger event of living a day. This is not meant to imply that making a bed and eating breakfast are necessary parts of living a day, just that they are typical ones. Even so, the events of making a bed and eating breakfast tend to occur at the beginning of a day, so their position is somewhat

constrained, much as the positions of chairs and desks are constrained in a classroom. At the basic level of making the bed or eating breakfast, there appear to be a multitude of event parts, and these tend to occur in a correlated fashion; tucking in sheets and fluffing pillows go with the bed, and making toast and brewing coffee go with the breakfast.

Just as object parts have a spatial structure—the parts of the body have a specific spatial arrangement—so event parts have a temporal structure. They also have a spatial structure. In making a bed, the bottom sheet goes on before the top one. The temporal and spatial configurations of many events, however, are flexible. At a birthday party, the games can come before or after the cake and ice cream. Grocery shopping can be done in any order, though some are more efficient than others.

Seen this way—the way observers see it—the set of event parts seems more like the set of morphemes than the set of phonemes. There is a large and open class of events and event parts. Parts are correlated within events; ice cream and cake go with the birthday party, and sheets and pillowcases with making the bed; they co-occur just as the morphemes used to describe those events co-occur.

Returning to the Questions

Returning to the opening questions, what can be said about the structure of events? Events have two structural bases: one at the raw level of changes in amount of activity, the other at the level of understanding; one bottom-up, one top-down. Observers' segmentation of events corresponds to sharp changes in level of activity, suggesting that either people are using changes in activity for segmentation or that the features they are using correlate with changes in activity. For unfamiliar events, descriptions of what occurs in event segments are in terms of movement, but as events become familiar they are described in terms of accomplishments of goals and subgoals. Events are distinguished from activities by achievements or accomplishments. Events can be conceived of at many levels: a lifetime can be a single event, but so can eating a meal or folding a shirt. What distinguish events from backgrounds are their accomplishments or achievements. Similarly, accomplishments or achievements partition entire events into segments and subsegments. The parts of events have both temporal and spatial configurations, but in many cases those configurations are flexible and can be rearranged. The number of event

parts is enormous in contrast to the number of phonemes, yet, rather than being combined and recombined like phonemes, event parts tend to co-occur. Finally, the parts of events (when viewed as action–object couplets), like the parts of objects, connect perception and appearance on the one hand and behavior and function on the other.

Pulling It Together

The world provides a multitude of sensations from which the mind delineates a multitude of experiences. Life is experienced as a series of events, events that ordinarily involve objects in places and which are facilitated by language. To reduce the overwhelming inundation of information, to make sense of it, and to predict and prepare for what will happen next, the mind segments, groups, and categorizes. The mind structures each of these domains critical to existence—language, objects, events, and scenes. They are experienced as meaningful, organized, and related wholes and parts, distinct from backgrounds. The comparisons of the structures across domains have been instructive.

The focus here has been events—not monumental events such as the French Revolution but events that involve one person and one place, events short enough to be studied in real time in the laboratory. These typify the events that fill the day, the events that people readily enact and comprehend even though they weren't born with those skills or that knowledge. Events are about action—not simple action but action that ends in accomplishment or achievement. The sheer amount of activity ebbs and flows as the events progress; the ebbs and flows correlate with the parsing of events into segments. The segment boundaries also coincide with achievements and accomplishments of goals. Objects, chunks in space, are the closest analogue to events, which are chunks in time and space. For objects, contours are distinctive and informative, so much so that many can be recognized as silhouettes (e.g., Rosch et al., 1976). Discontinuities in object contours correlate and seem to serve as a signal to parse object parts. The ebbs and flows of activity in events form one-dimensional contours that bear analogies and disanalogies to object contours. As for objects, the partition boundaries fall at points of change. The analogy extends in that the separate parts of objects and events are both salient in perception and serve as clues to behavior or function. These links from the perceptually salient to the

functionally significant promote understanding of new events. But the contour analogy fails at an essential point: objects can be recognized by spatial contour, but events can not be recognized by activity contour. For events, higher-level qualitative information, namely actions on objects, is needed for recognition.

Partitioning the "blooming, buzzing confusion" the world presents is the first step to comprehending it. Some partitioning is so instantaneous and automatic that perception of the world is not of multimedia mixtures of continuously changing sensations but rather of coherent objects, events, and scenes. The mind goes on to parse those elements and to look for structure among the parts. Typically there is a perceptual basis for part structure. Truly understanding each of these elements of our lives requires assigning meaning to their parts. Simply chopping up the flow of experience into chunks does not, by itself, allow comprehension of the world or action in it. However, in the world, the parts of objects, of scenes, of language, and of events covary with the functions of those things. According to the perception-to-function hypothesis, the perceptual identification of parts allows bootstrapping to meaning. What is remarkable about the segmentation of activity into events, then, is that the discovery of parts in perception provides links to their significance in conception.

Acknowledgments

References

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*, 183–193.

Avrahami, J., & Kareev, Y. (1994). The emergence of events. *Cognition*, *53*, 239–261.

Barker, R. G., & Wright, H. F. (1954). *Midwest and its children: The psychological ecology of an American town*. Evanston, IL: Row, Peterson and Company.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–252). Hillsdale, NJ: Erlbaum.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–117.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177–220.

Brent, M. R. (1999). An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning*, *34*, 71–105.

Brown, R. (1958). How shall a thing be called? *Psychological Review*, 65, 14–21.

Byrne, R. (2002). Seeing actions as hierarchically organized structures: Great ape manual skills. In A. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution and brain bases* (pp. 122–130). Cambridge: Cambridge University Press.

Casati, R., & Varzi, A. C. (1996). *Events*. Aldershot, England; Brookfield, VT: Dartmouth.

Casati, R., & Varzi, A. C. (1999). *Parts and places: The structures of spatial representation*. Cambridge, MA: MIT Press.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Epstein, R., & Kanwisher, N. (1998) A cortical representation of the local visual environment. *Nature*, *392*, 599–601.

Gelman, R., Durgin, F., & Kaufman, L. (1995). Distinguishing between animates and inanimates: Not by motion alone. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 150–184). Oxford: Clarendon Press.

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory*, *4*, 161–178.

Hard, B. M. (2006). Reading the language of action: Hierarchical encoding of observed behavior. Ph.D. Dissertation. Palo Alto, CA: Stanford University.

Hard, B. M., Lozano, S. C., & Tversky, B. (2006). Hierarchical encoding: Translating perception into action. *Journal of Experimental Psychology: Genera1*, *135*, 588–608.

Hard, B. M., Tversky, B., & Lang, D. (in press). Making sense of abstract events: Building event schemas. *Memory and Cognition*.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, *57*, 243–259.

Hemeren, P. E. (1996). Frequency, ordinal position and semantic distance as measures of cross-cultural stability and hierarchies for action verbs. *Acta Psychologica*, *91*, 39–66.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271.

Hochberg, J. (1978). *Perception*. Englewood Cliffs, NJ: Prentice-Hall.

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition*, *18*, 65–96.

Hoffman, D. D., & Singh, M. (1997) Salience of visual parts. *Cognition*, *63*, 29–78.

Hummel, J. E., & Biederman, I. (1992). Binding in a neural network for shape recogniton. *Psychological Review*, *99*, 480–517.

Jacob, P., & Jeannerod, M. (2003). *Ways of seeing: The scope and limits of visual cognition*. Oxford: Oxford University Press.

Landauer, T. K. (1998). Learning and representing verbal meaning: The latent semantic analysis theory. *Current Directions in Psychological Science*, *7*, 161–164.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Malt, B. C., & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory & Cognition*, *10*, 69–75.

Mandler, J. M., & Parker, R. E. (1976) Memory for descriptive and spatial information in complex pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 38–48.

Miller, G. A. (1963). *Language and communication*. New York: McGraw-Hill.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

Morris, M. W., & Murphy, G. L. (1990). Converging operations on a basic level in event taxonomies. *Memory & Cognition*, *18*, 407–418.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*, 28–38.

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, *12*, 436–450.

Palmer, S., Rosch, E. & Chase, P. (1981). Canonical perspective and the perception of objects. In J. B. Long & A. D. Baddeley (Eds.), *Attention and Performance*, IX. Hillsdale, NJ: Erlbaum.

Peterson, M. A. (1994). Shape recognition can and does occur before figure-ground organization. *Current Directions in Psychological Science*, *3*, 105–111.

Quine, W. V. (1985/1996). Events and reification. In R. Casati & A. C. Varzi (Eds.), *Events* (pp. 107–116). Aldershot, England: Dartmouth. Reprinted from E. LePore & B. P. McLaughlin (Eds.), *Actions and events: Perspectives on the philosophy of Donald Davidson* (pp. 162–171). Oxford: Blackwell.

Rifkin, A. (1985). Evidence for a basic level in event taxonomies. *Memory & Cognition*, *13*, 538–556.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*, 101–105.

Schwan, S., & Garsoffky, B. (2004). The cognitive representation of filmic event summaries. Applied Cognitive Psychology, 18, 37–55.

Shipley, T. F., Maguire, M. J., & Brumberg, J. (2004). Segmentation of event paths. *Journal of Vision*, *4*, 562.

Spelke, E. S., Gutheil, G., & Van der Valle, G. (1995). The development of object perception. In D. Osherson (Ed.), *An invitation to cognitive science*, Vol. 2 (pp. 297–330). Cambridge, MA: MIT Press.

Tversky, B. (1989). Parts, partonomies, and taxonomies. *Developmental Psychology*, *25*, 983–995.

Tversky, B. (1990). Where partonomies and taxonomies meet. In S. L. Tsohatzidis (Ed.), *Meanings and prototypes: Studies in linguistic categorization* (pp. 334–344). London: Routledge.

Tversky, B., & Hemenway, K. (1983). Categories of scenes. *Cognitive Psychology*, *15*, 121–149.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, *113*, 169–193.

Tversky, B., Zacks, J. M., & Lee, P. (2004). Events by hand and feet. *Spatial Cognition and Computation*, *4*, 5–14.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*, 979–1008.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., et al. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*, 651–655.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, 127, 3–21.

Zacks, J., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*, 29–58.