Barsalou, L.W. (2005). Abstraction as dynamic interpretation in perceptual symbol systems. In L. Gershkoff-Stowe & D. Rakison (Eds.), *Building object categories* (389-431). Carnegie Symposium Series. Majwah, NJ: Erlbaum.

Chapter 15

Abstraction as Dynamic Interpretation in Perceptual Symbol Systems

Lawrence W. Barsalou
Emory University

INTRODUCTION

If a scientific construct's centrality reflects the variety of forms it takes, then abstraction is a central construct in cognitive science, taking at least the following six senses:

Sense 1: Abstraction as categorical knowledge. Abstraction can simply mean that knowledge of a specific category has been abstracted out of the buzzing and blooming confusion of experience. Just about any account of knowledge is comfortable with this sense, including rule-based, prototype, exemplar, connectionist, and embodied theories.

Sense 2: Abstraction as the behavioral ability to generalize across instances. Another relatively uncontroversial sense is that people can summarize the properties of one or more category members *behaviorally*. All theories agree that people state generics, such as "Bats live in caves," and state quantifications, such as "Some birds fly." Behaviorally, people clearly produce abstractions.

The material in this chapter was first presented at the Workshop on Abstraction, supported by the Cognitive Science Program, French Ministere de la Recherche, CNRS, Gif sur Yvette, France, September 2001. This research was supported by National Science Foundation Grants SBR-9905024 and BCS-0212134 to Lawrence W. Barsalou. Address correspondence to Lawrence W. Barsalou, Department of Psychology, Emory University, Atlanta, GA 30322 (barsalou@emory.edu, http://userwww.service.emory.edu/~barsalou/).

Sense 3: Abstraction as summary representation. Much more controversial are the cognitive bases of the behavioral abstractions in Sense 2. According to some theories, behavioral abstractions reflect underlying summary representations of category instances in long-term memory. On these views, when people generalize behaviorally, they read out an underlying summary representation, such as a declarative rule, a statistical prototype, or a connectionist attractor. Notably, however, the summary representations in Sense 3 are not necessary to produce the behavioral abstractions in Sense 2. In exemplar models, only exemplars are stored in memory—no summary representations—and behavioral abstractions result from scanning and summarizing exemplars online (e.g., Hintzman, 1986).[1]

Sense 4: Abstraction as schematic representation. Another controversial sense is that schematic representations describe categories in memory. According to this sense, summary representations are sparser than exemplars, abstracting critical properties and discarding irrelevant ones (e.g., Biederman's, 1987, geons). Alternatively, properties may be distorted in various ways to idealize or caricature a category, thereby increasing its distinguishability relative to other categories (e.g., Posner & Keele, 1968; Rhodes, Brennan, & Carey, 1987; also see Barsalou, 1985; Palmeri & Nosofsky, 2001).

Sense 5: Abstraction as flexible representation. Another controversial sense of abstraction is that summary representations can be applied flexibly to a wide variety of tasks, including categorization, inference, language comprehension, reasoning, and so on. According to this sense, increasing abstractness allows a representation to become increasingly flexible (e.g., Winograd, 1975).

Sense 6: Abstraction as abstract concepts. Finally, abstraction can refer to the concreteness of concepts, ranging from concrete (e.g., *CHAIR*) to abstract (e.g., *TRUTH*).[2] As concepts become increasingly detached from physical entities, and more associated with mental events, they become increasingly abstract (e.g., Barsalou, 1999; Barsalou & Wiemer-Hastings, in press; Paivio, 1986; Wiemer-Hastings, Krug, & Xu, 2001).

[1]A classic problem for this view is why these abstractions do not subsequently become stored in memory along with exemplars. To the extent that abstractions require deep processing to produce, they should become well established in memory (e.g., Hyde & Jenkins, 1969).

[2]Italics will be used to indicate concepts, and quotes will be used to indicate linguistic forms (words, sentences). Thus, *CHAIR* indicates a concept, and "chair" indicates the corresponding word. Within concepts, uppercase words will represent categories, whereas lowercase words will represent properties of categories (e.g., *CHAIR* vs. *seat*) and relations between properties (e.g., *above* for the relation of a *CHAIR's seat* to its *back*).

As these senses illustrate, abstraction is a central construct in cognitive science. The focus here, however, is on the most controversial sense, namely, Sense 3. From here on, "abstraction" will mean *summary representations* in long-term memory.

The first of the five remaining sections describes three properties of abstractions. The second section reviews existing approaches and problems that they encounter for these properties. The third and fourth sections present the DIPSS theory of abstraction (Dynamic Interpretation in Perceptual Symbol Systems). The fifth section shows how DIPSS can be applied to various abstraction phenomena in categorization, inference, background knowledge, and learning. The final section revisits the other five senses of abstraction, and applies DIPSS to them.

PROPERTIES OF ABSTRACTIONS

Three properties of abstractions appear central to their nature: interpretation, structured representation, and dynamic realization.

Interpretation

In a classic paper, Pylyshyn (1973) argued that cognition is inherently an interpretive process. Addressing the nature of mental imagery, Pylyshyn argued that cognitive representations are not like the holistic bit-mapped recordings in cameras, video recorders, and audio recorders. Many perception researchers would agree (e.g., Hochberg, 1998). Rather than being recordings, Pylyshyn argued, cognitive representations are interpretations of experience. To produce an interpretation, concepts in memory type the components of sensorimotor experience to produce type–token propositions. On walking into a living room, for example, the concepts for *SOFA*, *RUG*, and *LAMP* become bound to particular objects, thereby creating type–token propositions of the sort, *SOFA*(object-98), *RUG*(object-32), and so on. Such propositions essentially make claims about the world that can be true or false, such as the belief that object-98 is a *SOFA* (e.g., Church, 1956).

A given component of experience can be interpreted in infinite ways. Thus object-98 could be interpreted alternatively as *FURNITURE*(object-98), *CONTEMPORARY SOFA*(object-98), *PLACE TO CRASH*(object-98), *PLACES THE DOG CAN'T SIT*(object-98), and so forth. Not only are there infinite true interpretations of an individual, there are infinite false ones as well, with each interpretation providing a different spin on how to think about it.

Once a type–token proposition is constructed to interpret an entity or event, the proposition provides a wealth of inferential knowledge. Once something is interpreted as a *SOFA*, inferences follow that it's soft, comfortable, and heavy, extending the object's interpretation. If the object were interpreted instead as *A PLACE THE DOG CAN'T SIT*, different inferences would follow. All such inferences constitute propositions linked to the type–token mappings that triggered them.

On this view, propositions underlie representations of the world, not bitmapped recordings (also see Barsalou, 1999; Dretske, 1995; Haugeland, 1991). A representation of a chair is not a holistic recording of it, but a set of propositions that interpret it. Most importantly for our purposes, Pylyshn assumed that abstractions underlie this process. The types in his type–token propositions are abstractions for properties, objects, events, relations, and so forth. Once a concept has been abstracted from experience, its summary representation enables the subsequent interpretation of later experiences. Thus abstractions are linked closely to interpretation.

Structured Representation

When concepts interpret experience, they typically do not do so individually. Instead they become organized into structured representations that capture relations between individual type–token propositions. Rather than *SOFA*(object-98) and *RUG*(object-32) being independent, a spatial concept, such as *on*, might organize them into a structured proposition, such as:

on(upper-region = *SOFA*[object-98], *lower-region* = *RUG*[object-32])

Much empirical evidence demonstrates the extensive presence of structured representations in human knowledge. Perhaps the most direct evidence comes from work on concepts and categorization, where researchers have explicitly assessed the presence of such structure and found robust evidence for it (e.g., Goldstone & Medin, 1994; Markman & Gentner, 1997; also see Barsalou, 1992; Barsalou & Hale, 1993). Assigning exemplars to categories, judging the similarity of exemplars, and drawing categorical inferences all rely heavily on structured relations—not just on independent properties. Furthermore, the process of conceptual combination is essentially the process of combining individual concepts into structured representations (e.g., Hampton, 1997; Rips, 1995; Wisniewski, 1997).

Much additional evidence comes from research on analogy, where structured representations are strongly implicated in people's ability to extend relational systems from one domain to another (e.g., Gentner & Markman, 1997; Holyoak & Thagard, 1997). Similar evidence comes from the literature on language comprehension, where complex propositional structures

provide the standard scheme for representing meaning (e.g., Graesser, Singer, & Trabasso, 1994; Kintsch & van Dijk, 1978). Fodor and Pylyshyn (1988) offered general theoretical arguments for the necessity of structured representations, and for the related constructs of productivity and systematicity.

Thus a second important property of abstractions is that they enter into complex interpretive systems. Rather than interpreting isolated components of experience, abstractions assemble into structured representations that interpret complex structure in the world.

Dynamic Realization

The abstractions that represent a category are notoriously difficult to pin down. In my own research, I have continuously experienced the slipperiness of abstractions, referring to it as *linguistic vagary* (Barsalou, 1993). Artificial intelligence researchers who program knowledge into intelligent systems chronically experience similar vagaries in articulating abstractions. Specifically, three problems arise in trying to specify the abstraction that represents a category:

Identifiability. What particular information should be included in an abstraction? Consider Schank and Abelson's (1977) attempt to specify the abstraction that underlies the restaurant script. Of everything that could possibly occur in a restaurant, what should be included in a summary representation? Only the most invariant properties across restaurant visits? What about important properties only true occasionally? What about differences between cultures and individuals? When is an abstraction complete? Specifying the content of an abstraction is an extremely challenging task.

Motivation. Why is a particular abstraction the correct one? Typically artificial intelligence researchers intuitively select the abstractions that best serve a specific application. Problematically, however, no principled account of how to do this exists, nor is it clear that such an account is possible.

Rigidity. How does one handle all of the exceptions that arise for an abstraction? When Schank and Abelson proposed the restaurant script, a common criticism was that it could not handle unexpected deviations and unusual restaurant visits. Schank and Abelson replied that different *tracks* through a script handle special cases, but the counter-reply was that infinitely many tracks are required to handle all the possibilities. Moreover, how does one handle cases never encountered, which people seem to do effortlessly? No compelling account of how abstraction can handle such variability exists.

Conclusion. One could view the identifiability, motivation, and rigidity problems for abstractions as a sign that we simply need a better methodology for discovering them. Alternatively there may be no correct abstractions to discover. Rather than a single abstraction representing a category, diverse abstractions may be constructed online to represent a category temporarily (Barsalou, 1987, 1989, 1993). If so, then studying the *skill* to construct temporary abstractions dynamically may be more fruitful than attempting to establish one particular abstraction that represents a category. In this spirit, I assume that a third important property of abstractions is their dynamic realization (this is clearly not a standard assumption in the literature).

THEORIES OF SUMMARY REPRESENTATION

Later sections develop a theory of dynamically realized abstractions. First, however, it is useful to briefly review existing theories, and the status of abstraction as a theoretical construct.

GOFAI Theories of Abstraction

Haugeland (1985) dubbed classic abstraction theories as "Good Old Fashioned Artificial Intelligence" (i.e., GOFAI), an approach that dominated the early history of the field, and that continues alongside other approaches currently. Classic examples can be found in Winograd (1972), Newell and Simon (1972), Schank and Colby (1973), Bobrow and Collins (1975), Schank and Abelson (1977), and Charniak and McDermott (1985).

GOFAI provides a powerful account of interpretation and structured representation. Through the mechanisms of argument binding and recursion, GOFAI implements these processes elegantly and powerfully. Simple interpretation results from binding a predicate to an individual, and structured representation results from binding higher order predicates to lower order ones. Thus, *SOFA(X)* and *RUG(X)* can be bound to object-98 and object-32, thereby interpreting those objects in particular ways. Similarly, *on(upper-region* = x, *lower-region* = y) can be bound to *SOFA*(object-98) and *RUG*(object-32), thereby forming a structured representation.

The problem that has bedeviled GOFAI theories for decades is dynamic realization. Identifying the content of the abstractions in GOFAI theories has constituted a daunting and sobering challenge. Clearly, adequate abstractions can be developed for specific tasks, yet few would argue that they offer definitive accounts of human knowledge. Motivating these accounts has also been difficult, given their reliance on programmer intuition. Perhaps most critically, these accounts are known for their brittleness. Although they work in some situations, they don't work in all, given the difficulty of handling exceptions and novel cases.

Another major problem for GOFAI concerns the basic nature of their symbols. On the one hand, connectionist theories argue that the discrete symbols in GOFAI representations don't exist—instead knowledge is distributed statistically across continuous neural-like processing units (e.g., McClelland, Rumelhart, & the PDP Research Group, 1986; Rumelhart, McClelland, & the PDP Research Group, 1986; Smolenksy, 1988). On the other hand, embodied theories argue that the arbitrary amodal symbols in GOFAI don't exist—instead simulations of sensorimotor processing represent knowledge (e.g., Barsalou, 1999; Glenberg, 1997; Mandler, 1992).

For all these reasons, accounts of human knowledge look increasingly less and less like GOFAI representations. Theorists find it increasingly implausible that knowledge takes this form.

Connectionist Theories of Abstraction

In GOFAI theories, abstractions are clear and explicit, spelled out in predicate calculus-like expressions. Connectionism offers a radically different approach, where abstractions are relatively fuzzy and implicit. In a network of neural-like processing units, an abstraction is an attractor for a statistically likely combination of properties. When a set of learned exemplars shares correlated properties, the network's weights evolve to recognize this pattern and its variants, establishing an attractor for the category. Within the dynamical system that constitutes the network's state space, activation prefers to follow trajectories toward learned attractors.

The active units that characterize an attractor implicitly represent an abstraction. When these units are distributed and course-coded, the content of an abstraction can be difficult to specify precisely, but this is the beauty of the approach: It is not necessary or even desirable to specify abstractions explicitly or precisely, thereby avoiding the brittleness of GOFAI abstractions.

Two properties of connectionist abstraction further allow it to avoid brittleness. First, many activation states around an attractor can each represent the same category. Depending on the current context, the representation of the category can vary dynamically (e.g., Smolensky, 1988). Second, as experience with a category changes, an attractor can adapt quickly. Connectionist learning algorithms offer powerful ways to revise abstractions as the input changes. For all these reasons, connectionist approaches offer a compelling account of dynamic realization.

Where connectionist theories struggle is with the first two properties of abstraction: interpretation and structured representations. Connectionist theories do offer a basic form of interpretation. When an attractor becomes active, it provides an interpretation of the input. Because the attractor represents multiple category instances, it is a type that can interpret tokens. Implicitly, the simultaneous activation of an attractor and its input constitute a type–token relation, albeit of a much different variety than in GOFAI theories. Furthermore, if the attractor represents information not found in the input, the attractor provides inferences via pattern completion (e.g., unseen parts, actions, contexts, etc.; Rumelhart, Smolensky, McClelland, & Hinton, 1986).

Although connectionist nets have basic interpretive ability, limitations become apparent as interpretations become increasingly complex and structured (e.g., Fodor & Pylyshyn, 1988). In a complex situation with many exemplars present, it becomes tricky to bind all of the active attractors to their respective instances. Furthermore, when arguments must be bound in structured relations, connectionist nets don't naturally accomplish this, at least not nearly as naturally or easily as in GOFAI. Connectionist theorists have offered various solutions. For example, a separate bank of units can be set aside for each argument in a relation, with its bound value being the currently active state (e.g., Miikkullainen & Dyer, 1991). A problem with this approach is its assumption that the set of arguments is finite, small, and known in advance—an unlikely possibility (Barsalou, 1993).

Another approach is to translate each element of a predicate calculus expression into a vector, superimpose all these individual vectors into a single vector, and then extract the string of symbols as needed later (e.g., Pollack, 1990; Smolensky, 1990; van Gelder, 1990). Technically, this algorithm can implement structured representations, although there are problems for it too (Barsalou, 1993). Problematically, this approach has not struck many researchers as psychologically plausible—it primarily seems like an engineering solution, mostly having computer science applications. Furthermore, this approach assumes implicitly that predicate calculus expressions—at some level of functionality—are the right way to think about knowledge. Given the troubled status of GOFAI representations as psychological accounts, their connectionist cousins could be viewed similarly.

Lack of a Viable Account

Abstraction in the classic sense has gone out of fashion—at the least, it has become a dubious construct. On the one hand, classic GOFAI theories that champion abstraction are falling by the wayside. On the other hand, abstraction plays a minimalist role in the theories replacing them. As we just saw, abstraction exists in connectionism, but in a much less powerful form. Simple interpretation arises naturally and elegantly, but complex structured interpretations do not.

Furthermore, the other reigning theories of knowledge—exemplar models and latent semantic analysis—similarly relegate abstraction to the periphery. In standard exemplar models, no abstractions exist—only exemplar representations (e.g., Brooks, 1978; Lamberts, 1995; Medin & Schaffer, 1978; Nosofsky, 1984). Furthermore the empirical literature has largely failed to support the existence of summary representations for categories (although see Nosofsky, Palmeri, & McKinley, 1994; J. Smith & Minda, 2002; Spalding & Ross, 1994). The message is that there is no need for summary representations in memory. When an abstraction is needed, it can be constructed online behaviorally and then discarded. In latent semantic analysis, the message is similar: From a large data base of low-level associative knowledge, it is possible to compute online abstractions as needed—it is not necessary to store them explicitly (Burgess & Lund, 1997; Landauer & Dumais, 1997).

All of these approaches do a great job capturing the dynamic realization of abstractions; indeed, they can all be viewed as reactions to the rigidity and brittleness of GOFAI theories. Where they fall down, though, is in handling structured representations. Some might say that we don't need to worry about structured representations, but throwing out babies with bath water appears applicable here. Researchers who study conceptual combination, language, and thought all know that structured representations are not only a signature property of human cognition, but are essential for adequate accounts of these phenomena. Thus it is important to search for theories of abstraction that not only explain dynamic realization but that also explain structured interpretation.

SIMULATORS AND SIMULATIONS IN PERCEPTUAL SYMBOL SYSTEMS

One tack would be to develop more plausible connectionist accounts of abstraction. Indeed, this is an important direction to explore. The theme of this article, however, is that an elegant and natural account of structured interpretation can be found elsewhere, namely, in theories of embodied cognition. Furthermore, these theories naturally exhibit dynamic realization, thereby offering the potential for a complete account.

This current section lays the groundwork for the embodied theory of abstraction in the subsequent two sections. Obviously a working computational model is desirable. The goal here, however, is to outline the mechanisms of such a model, and to show how this architecture implements the three properties of abstraction. First, simulators and simulations are defined (i.e., two of the basic constructs in perceptual symbol systems; Barsalou, 1999). The focus then turns to simulators for properties and relations, and the empirical evidence for them.

Simulators and Simulations

Sensorimotor Reenactment. The account of abstraction later relies on simulators and simulations, which more basically rely on the mechanism of sensorimotor reenactment. Damasio (1989) presented this mechanism in his convergence zone theory (also see Simmons & Barsalou, 2003, for further development of this approach). Sensorimotor reenactment has also been adopted widely in neural accounts of mental imagery (e.g., Farah, 2000; Jeannerod, 1995; Kosslyn, 1994). As Fig. 15.1 illustrates, this mechanism has two phases—storage and reenactment—each addressed in turn.

When a physical entity is perceived, it activates feature detectors in the relevant sensorimotor areas. During the visual processing of a chair, for example, some neurons fire for edges, vertices, and planar surfaces, whereas others fire for color, orientation, and direction of movement. The overall pattern of activation across this hierarchically organized distributed system represents the entity in visual perception (e.g., Palmer, 1999; Zeki, 1993). Analogous distributions of activation on other modalities represent how the entity might sound and feel, and also actions performed on it. A similar

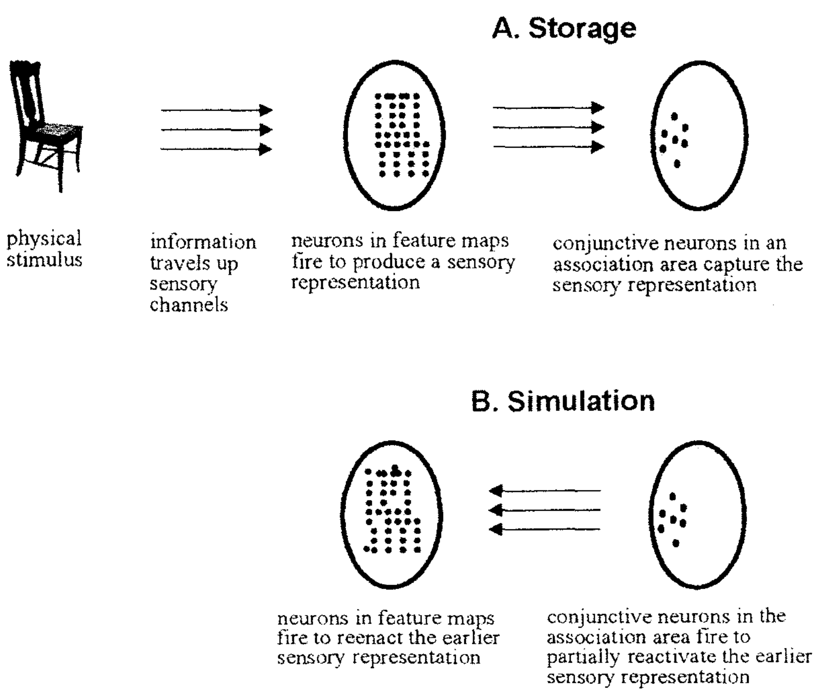

FIG. 15.1. Illustration of the storage (A) and simulation (B) of sensorimotor information in Damasio (1989) and Barsalou (1999).

account can be given for introspective states that arise while interacting with an entity. For example, patterns of activation in the amygdala and orbitofrontal areas represent emotional reactions to perceived entities. A tremendous amount of neuroscience research has documented the structure of feature maps across modalities and the states that arise in them.

One might be concerned that the neural activation of a feature map in Fig. 15.1 looks like a chair. Notably, however, feature maps in vision are often organized topographically. Indeed, many topographically mapped areas reside in the visual system alone, with others residing in the motor, somatosensory, and auditory modalities. Thus it is not unreasonable to assume that modality-specific representations take a somewhat topographic form. Most importantly, however, *nothing* in perceptual symbol systems, nor in the account to follow, depends on topographically mapped representations! If these representations were completely nontopographic, the theory would work identically. The important assumption is that sensorimotor representations exist, and that high-level cognitive processes reenact them to represent knowledge.

As Fig. 15.1 further illustrates, when a pattern becomes active in a feature map, conjunctive neurons in an association area capture the pattern's features for later use. Damasio (1989) referred to these association areas as "convergence zones," and assumed that they exist at multiple hierarchical levels, ranging from posterior to anterior in the brain. Most locally, posterior convergence zones near a particular modality capture patterns of activation within it. Thus association areas near the visual system capture patterns of activation there, whereas association areas near the motor system capture patterns of activation there. Further downstream in more anterior areas, higher level association areas, such as the temporal and frontal lobes, conjoin patterns of activation *across* modalities.

This architecture of feature maps and convergence zones has the functional capability to produce sensorimotor reenactment: Once a subset of conjunctive neurons in a convergence zone captures an activation pattern in a feature map, the conjunctive neurons can later reactivate the pattern in the absence of bottom-up sensory stimulation. While remembering a perceived object, for example, conjunctive neurons reenact the sensorimotor states that were active while encoding it. Similarly, when representing a concept, conjunctive neurons reenact the sensorimotor states characteristic of its instances. A given reenactment is never complete, and biases may enter into its reactivation, but at least some semblance of the original state is partially activated.

Although this basic mechanism is viewed widely as underlying mental imagery (e.g., Farah, 2000; Jeannerod, 1995; Kosslyn, 1994), the reenactments it produces need not be conscious mental images. As Barsalou (1999, 2003) suggested, *unconscious* reenactments may often underlie memory, conceptualization, comprehension, and reasoning. Whereas explicit attempts to construct mental imagery may typically create relatively vivid reenactments, other cognitive processes may typically rely on less conscious reenactments. In the account of abstraction to follow, the neural reenactment of sensorimotor mechanisms is the critical mechanism—*not* conscious mental images.

Simulators and Simulations. Barsalou (1999) developed a theory of concepts based on the neural reenactment of sensorimotor states. Figure 15.2 illustrates the basic constructs in this theory: *simulators* and *simulations*. As multiple instances of the same concept are encountered, they tend to activate similar neural states in feature maps (i.e., categories tend to have statistically correlated features; Rosch & Mervis, 1975). As a result, similar populations of conjunctive neurons tend to capture these states (Simmons & Barsalou, 2003, argued that these populations are localized topographically). Over time, conjunctive neurons integrate sensorimotor features across diverse category instances and across diverse settings, establishing a multimodal representation of a category. For the category *CAR*, visual information about how cars look is integrated, along with auditory information about how they sound, olfactory information about how they smell, motor information about driving them, somatosensory information about the feel of riding in them, and emotional information associated with acceleration, collisions, and so on. The result is a distributed system throughout the brain's association and modality-specific areas that establishes conceptual content for *CAR*. Barsalou (1999) referred to this distributed system as a *simulator*.

Once a simulator exists, it can reenact small subsets of its content as specific *simulations*. The entire content of a simulator is never activated at once; only a small subset becomes active on a given occasion (cf. Barsalou, 1987, 1989, 1993). As Barsalou (2003) proposed, the active subset is tailored to the agent's current context of action, providing goal-relevant inferences about objects, actions, mental states, and the background setting. Thus, on one occasion, the *CAR* simulator might produce a simulation of driving a car, whereas on others it might produce a simulation of fueling a car, of seeing a car drive by, and so forth. Although all the experienced content for a category resides implicitly in a simulator, only specific subsets are reenacted on a given occasion.

Once a simulation has been constructed, it can serve a wide variety of cognitive functions (Barsalou, 1999). For example, simulations can be used to draw inferences about physical instances of a category currently present in the environment. Alternatively, simulations can represent instances in their absence during memory, language, and thought.

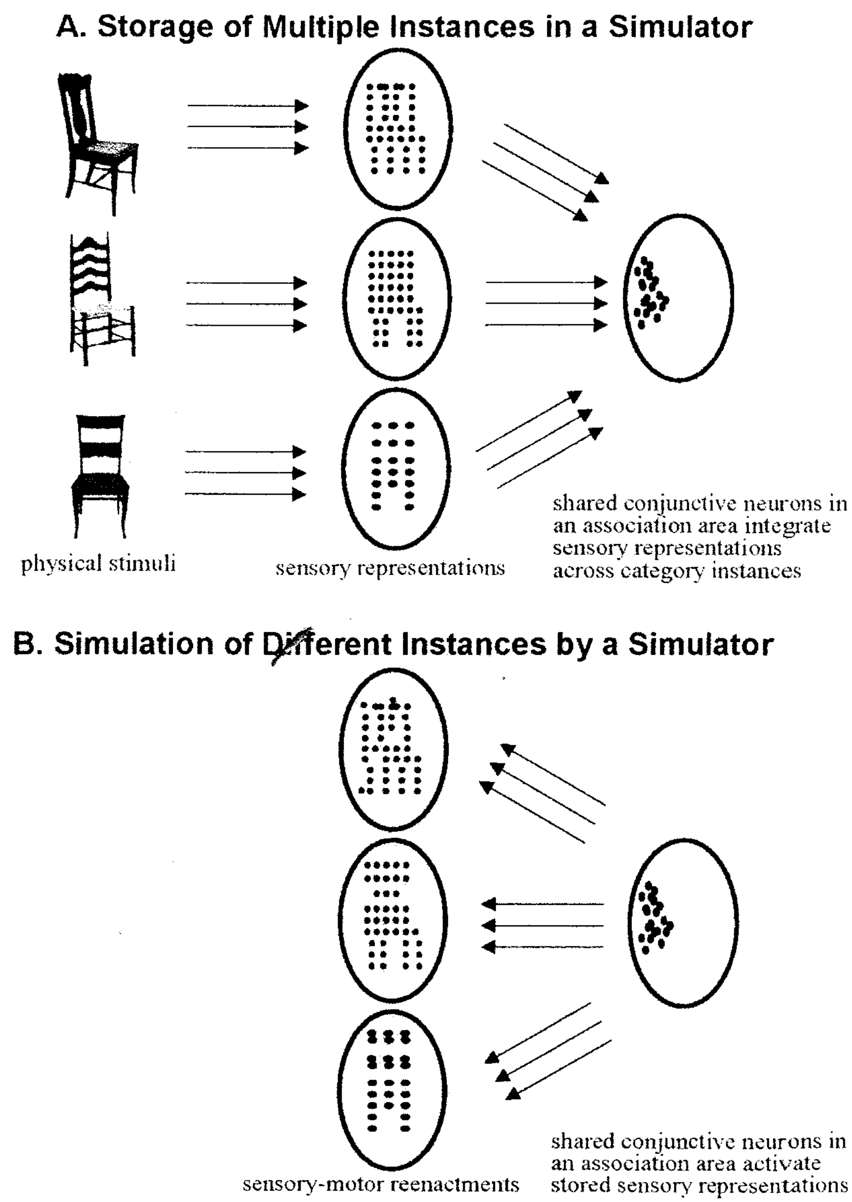

FIG. 15.2. Illustration of simulators (A) and simulations (B) in perceptual symbol systems (Barsalou, 1999).

Property Simulators

In principle, an infinite number of simulators can be established in the brain, reflecting the considerable flexibility of selective attention. Barsalou (1999) argued that a simulator develops for any component of experience that attention selects repeatedly. Thus, if attention focuses repeatedly on chairs, a simulator develops for them. Simulators don't just develop for physical objects, however, they also develop for locations, events, actions, mental states, and so forth. Such flexibility is consistent with Schyns, Goldstone, and Thibaut's (1998) argument that new features can be learned creatively as they become relevant for categorization.

The theory of abstraction to follow rests on simulators for properties (this section) and simulators for relations (next section). Most basically, a property simulator constructs specific simulations of the forms that a property takes across different categories. As Fig. 15.3 illustrates, the simulator for *nose* stimulates the noses of *HUMANS, DOGS, FISH, JETS,* and so forth. The next seven subsections develop the construct of a property simulator in greater detail.

Multimodal Property Simulations. Property simulators are multimodal. Although Fig. 15.3 only illustrates visual information for *nose,* a property simulator reenacts whatever information across modalities is relevant. Thus when the *nose* simulator simulates a human nose, the simulation might reenact blowing one's nose, how it feels, and how it sounds—not just how it looks.

Local Property Simulations. Property simulators represent properties locally (Solomon & Barsalou, 2001). Rather than there being a single global representation of a property that represents it in different objects, many lo-

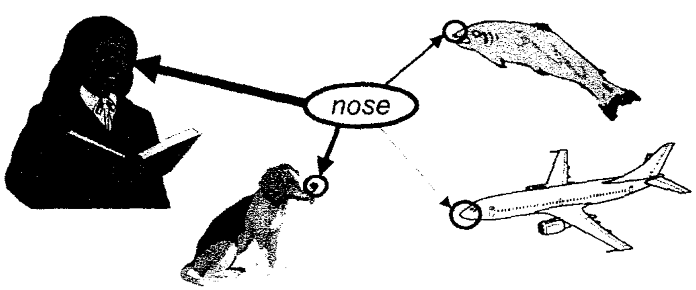

FIG. 15.3. Illustration of the local simulations for different property forms that a property simulator produces, each framed in a background simulation of the respective object (Solomon & Barsalou, 2001). Increasing link thickness represents a simulation's dominance in the simulator.

cal representations of a property represent it collectively. Thus, *nose* may be represented by local representations of *human noses, dog noses, fish noses,* and so forth. Although Fig. 15.3 only shows one nose for each category, many local representations may exist for each (e.g., many different *human noses*).

Simulations of Background Contexts. Local property representations are simulated in the background contexts of their respective objects (or events). As Fig. 15.3 illustrates, the property of *human nose* is not simulated as an isolated detached nose. Instead it is simulated as a focal region in a background face (or possibly a background body). Attention plays a central role in highlighting properties against background simulations (Barsalou, 1993, 1999; Langacker, 1986; Talmy, 1983).

Dominance Orders of Local Simulations. Within a property simulator, local property simulations are ordered implicitly by dominance (decreasing link thickness in Fig. 15.3). On activating a property simulator, dominant simulations are more likely to run than less-dominant ones. Although many factors influence dominance, frequency is likely to be particularly important. The more often a particular property form is experienced, the more dominant it becomes. Because people experience *human nose* more often than other noses, this local form becomes particularly dominant. As a result, when people think of *nose,* they are most likely to simulate *human nose* first. As they continue thinking about *nose,* however, other less dominant senses may be simulated.

Interpreting Simulated Objects and Events. Not only can a property simulator simulate local properties in their absence, it can interpret regions of simulated entities (and of perceived entities) as containing the property. Consider Fig. 15.4. At the top, the simulator for *JET* has run one particular simulation. At the bottom, the simulator for *nose* has run four different simulations, with one of them corresponding to a region in the *JET* simulation at the top. As a result of this mapping, the simulated object is interpreted as having the property of a *nose.*

Solomon and Barsalou (2001) discussed factors that affect this interpretive mapping. First, the likelihood of a successful mapping increases as a local property form increasingly matches some region of the simulated object (e.g., if a simulated human nose matches some region of a simulated human). Second, the likelihood of a successful mapping increases as the position of the simulated property in its background simulation increasingly corresponds to the position of the matching region in the simulated object (e.g., if both are in the center of a face on top of a body). To the extent that the two simulations are alignable, the mapping is more likely to be correct. Third, the likelihood of a successful mapping increases as the simulated

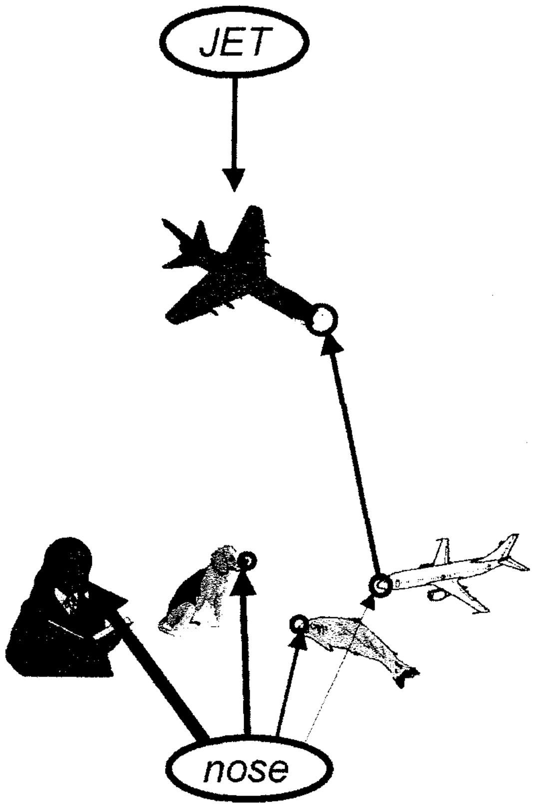

FIG. 15.4. Illustration of a property simulator, with one of its local simulations construing a region of an object simulation, thereby creating a type–token property mapping (Solomon & Barsalou, 2001).

function of the property corresponds to the function of the matching region in the object (e.g., if both sneeze, breathe, etc.). Assessments of function are likely to rely on the multimodal character of simulations (Barsalou, Sloman, Chaigneau, in press).

Type–Token Interpretation. Once a property simulator interprets a region of a simulated object as a property instance, an implicit type–token relation exists. The region of the simulated object is established as a token of the type that the property simulator represents. Mapping the *nose* simulator into a region of a simulated object types the region as a *nose.* The result is an implicit proposition that could be either true or false, and that carries inferences from the type to the token. For example, if an object region is interpreted as a *nose,* inferences about the region sneezing and breathing may

follow as further simulations from the *nose* simulator. In this manner, simulators produce the standard categorical inferences associated with type–token propositions (Barsalou, 1999).

Infinite Property Interpretations. In principle, an infinite number of property interpretations can be made of an entity. Given the continuous nature of a simulation, an infinite number of its regions (and groups of its regions) can be interpreted as properties. Furthermore, for a given region, an infinite number of simulators could interpret it truly or falsely. This open-ended character provides property interpretation with a dynamical character important in the later account of abstraction. This open-endedness also explains why enumerating a concept's properties exhaustively is impossible (Barsalou, 1993), as well as why people construct properties prolifically during learning (Schyns et al., 1998).

In summary, property simulators produce local simulations of properties that are multimodal, framed within the context of larger entities, and organized implicitly by dominance. When a local property simulation becomes active, it can interpret a region of a simulated or perceived entity, thereby establishing a type–token proposition that carries categorical inferences. An infinite number of such interpretations is possible.

Empirical Evidence for Property Simulators

Increasing evidence supports this account of property representation. Evidence for local property representations and dominance is reviewed first, followed by evidence for the modal character of these representations.[3]

Local Property Representations and Dominance. In Solomon and Barsalou (2001), participants verified a property first for one concept (e.g., *mane* for *HORSE*) and then, 15 to 25 trials later, verified the same property again for a different concept (e.g., *PONY-mane*). The key manipulation was the perceptual similarity of the first property sense to the second. Sometimes the two property forms were similar (e.g., *HORSE-mane* then *PONY-mane*) and sometimes they were not (e.g., *LION-mane* then *PONY-mane*). Of interest was whether only *HORSE-mane* would facilitate verifying *PONY-mane,* or whether *LION-mane* would facilitate it as well. If a single global representation underlies a property's meaning, then verifying *mane* for any concept earlier (e.g., *HORSE* or *LION*) should facilitate verifying it later for *PONY.* If local representations underlie a property's meaning instead, then only veri-

[3]This section only reviews evidence for property simulators. Barsalou (2003) provided a broader review of empirical results that support the presence of simulators across a variety of concepts and tasks.

fying *mane* for *HORSE* should produce a benefit; verifying *mane* for *LION* should not. Across several experiments, priming only occurred for similar local forms, ranging from 37 to 80 ms.[4]

This result indicates that local representations underlie properties. If a single representation underlies a property, the first property should facilitate it, regardless of the first property's similarity to the second. Limited facilitation indicates that a local form becomes active initially, and only facilitates similar forms later.

This finding also indicates that dominance organizes local property representations. Because a priming trial increases the accessibility of a local representation, an underlying dominance order is implied. Rather than being rigid, the dominance order of the local property representations within a simulator is malleable. A single trial can boost a local representation considerably.

An additional result also indicates the presence of dominance orders. When a property was verified for the first time, the dominant forms of properties were verified much more easily than less-dominant forms. For example, the dominant form of *mane* for *HORSE* was verified more easily than the less-dominant form for *LION*. Similarly, *HOUSE-roof* was easier than *CAR-roof*, and *HUMAN-nose* was easier than *AIRPLANE-nose*. These dominance effects were substantial. Nondominant senses were verified 173 ms slower than dominant ones (948 vs. 775 ms), and they exhibited many more errors (22% vs. 3%). Indeed, participants spontaneously noted that images of dominant forms came to mind while verifying nondominant forms, causing mistakes (imaging a human nose while verifying *AIRPLANE-nose*). These results indicate that local representations underlie the meaning of property words, and that dominance organizes them (also see Halff, Ortony, & Anderson, 1976; Wisniewski, 1998).

Modal Property Representations. Increasing evidence suggests that sensorimotor simulations underlie property representations. In Solomon and Barsalou (2004), some participants were asked to use imagery while verifying properties, whereas others received neutral instructions. If neutral participants spontaneously adopt amodal representations, their performance should differ from imagery participants. If, however, neutral participants spontaneously run simulations to verify properties, their performance should mirror the performance of imagery participants. Regression analyses showed strong similarities in the detailed performance of both groups. Perceptual factors were most important for both, followed by expectancy factors, and then linguistic factors. This predicted equivalence is consistent with neutral participants adopting simulations spontaneously.

[4]A potential concern is that concept similarity—not property similarity—underlies these effects. Limited facilitation may occur because the overall similarity of *HORSE* to *PONY* exceeds the overall similarity of *LION* to *PONY*, not because the local forms of *mane* are more similar for the first pair than for the second. To assess this possibility, Solomon and Barsalou (2001) included properties that were equally similar for all three concepts (e.g., *belly* was found to be equally similar for *HORSE*, *PONY*, and *LION*). However, property similarity—not concept similarity—continued to be the critical factor, given that *LION* facilitated *belly* as much as did *HORSE* when verifying it for *PONY*.

Furthermore, a perceptual variable, a property's size, was central in the performance of both neutral and imagery participants. The larger a property, the longer it took to verify. This finding is consistent with the interpretation that participants had to interpret a region of a simulation to verify the property in it. The larger the region to be processed, the longer the verification.

Kan, Barsalou, Solomon, Minor, and Thompson-Schill (2003) performed the Solomon and Barsalou (2004) experiment in an fMRI scanner. Under neutral conditions, activation occurred in the fusiform gyrus, an area that underlies visual imagery and high-level vision. These results corroborate Solomon and Barsalou's behavioral results, further implicating visual simulation in the verification of visual properties.

Pecher, Zeelenberg, and Barsalou (2003) found further evidence for modality-specific property representations. Perception research has shown that detecting a signal on a modality is faster when the previous signal was on the same modality than on a different one (e.g., Spence, Nicholls, & Driver, 2000). For example, verifying the presence of a tone is faster when the previous signal was a tone than when it was a light flash. Using linguistic materials and no imagery instructions, Pecher et al. demonstrated a similar phenomenon in the property verification task across six modalities (vision, audition, action, touch, taste, smell). When participants verified a property on the same modality as the previous trial, processing was 20 to 41 ms faster across experiments. For example, verifying *bland* (taste) for *CUCUMBER* was faster when *sour* (taste) had just been verified for *BUTTERMILK* than when *speckled* (vision) had just been verified for *BIRD EGG*. Further findings indicated that associative strength was not responsible for these effects. For a review of findings in this paradigm, see Barsalou, Pecher, Zeelenberg, Simmons, and Hamann (in press).

Finally, Martin and his colleagues performed an extensive program of fMRI research to localize property representations in the brain (for reviews, see Martin, 2001; Martin & Chao, 2001; Martin, Ungerleider, & Haxby, 2000). Across many experiments, they consistently found that properties are represented in modality-specific areas. Color properties reside in brain regions that process color; visual form properties reside in regions that process visual form; visual motion properties reside in regions that process visual motion; agentive action properties reside in regions that execute movements. Many reviews of the lesion literature reach similar conclusions (e.g., Damasio, 1989; Humphreys & Forde, 2001; McRae & Cree, 2002; Simmons & Barsalou, 2003; Warrington & Shallice, 1984).

Summary. The foregoing evidence is consistent with the account of property simulators presented earlier. A variety of local property representations, organized by dominance, underlies a property's meaning. Furthermore, these local representations are modality-specific, not amodal. Later we show how this account of properties lends itself to an account of abstraction.

Relation Simulators

Relation simulators will also be central to the later account of abstraction. Analogous to how a property simulator interprets a region of an object, a relation simulator interprets multiple regions and their configuration. For accounts that inspired this one, see Talmy (1983) and Langacker (1986).

Consider the *above* relation in Fig. 15.5a. Imagine that a child is told, "The jet is above the bird." While understanding this utterance, the child attends at some point to the spatial regions that contain the two objects. As a result, a memory of the regions, relative to spatial dimensions, is stored in memory, with the objects largely filtered out (i.e., in the dorsal stream). Further imagine that the upper region is of more interest and receives more attention (the bold regions in the middle of Fig. 15.5a), such that the memory represents the distribution of attention, as well as the spatial regions. Imagine that the child later hears "above" refer to other pairs of objects, as Fig. 15.5a illustrates. Over time, analogous to Fig. 15.3, multiple instances of the same spatial relation become stored together, establishing a simulator for *above*. As a result, it becomes possible to simulate different instances of *above*, each having a slightly different configuration of spatial regions (Fig. 15.5b). Furthermore, these simulations can be used to interpret spatial regions in perceptions and simulations. Just as a property simulator can be used to verify that a *HORSE* has a *mane*, a relation simulator can be used to verify that a *NOSE* is *above* a *MOUTH* in a face. As the bottom of Fig. 15.5b illustrates, mapping an *above* simulation into the regions of a perceived face can verify that the nose is above the mouth.

Empirical Evidence. Research in the attention and comprehension literatures supports the view that the meaning of a spatial relation is a simulation of spatial regions. In the attention literature, researchers present participants with a reference point, R, and tell them that another object, O, stands in some spatial relation to it (e.g., O is *above* R). O is then shown in many different positions around R, and the participant indicates how well each relation exemplifies *O is above R*. Sometimes the measure is simply a goodness rating; sometimes it's the time to find and process O.

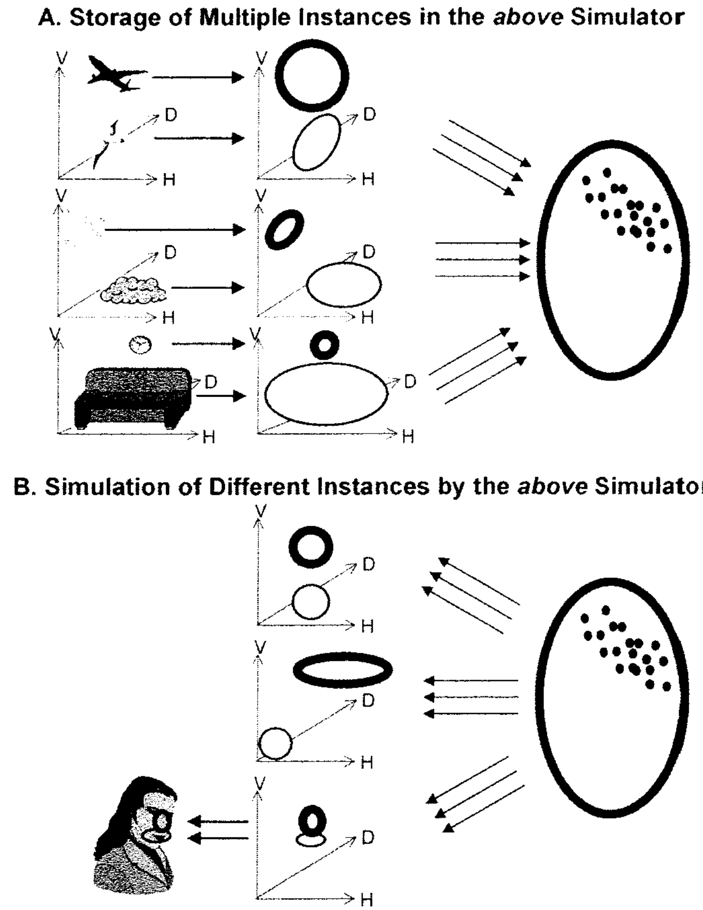

FIG. 15.5. Illustration of simulators (A) and simulations (B) for the *above* relation. V, H, and D represent the vertical, horizontal, and depth dimensions, respectively.

Much work has shown that a prototypical configuration of spatial regions underlies the meaning of a spatial preposition (e.g., Carlson-Radvansky & Logan, 1997; Hayward & Tarr, 1995; Logan & Compton, 1996). For *above*, the ideal configuration is for the center of O to be aligned geometrically above the center of R, not too far away. On hearing "above," participants appear to construct a perceptual simulation of the ideal configuration. When the subsequent display matches this simulation, processing is optimal. As the display departs increasingly from the ideal simulation, processing efficiency falls off in a graded manner. Importantly, however, even non-ideal displays are categorized as instances of the spatial relation, as long as they satisfy the qualitative criteria that define it. This graceful degradation is consistent with the amount of work that a simulator must do to adjust its simulation. The greater the departure from ideal, the more transformation necessary to simulate the configuration. Analogous to property simulators, a family of simulations—not just one simulation—underlies a spatial relation. Although some simulations are preferred, a wide variety exists.

Additional research shows that function modifies ideal geometric simulations to optimize situated action (e.g., Carlson-Radvansky, Conventry, 1998; Covey, & Lattanzi, 1999). Consider the statement, "The toothpaste tube is above the toothbrush." If spatial geometry were the only factor affecting the ideal simulation of a spatial relation, then a picture of a toothpaste tube centered over a toothbrush should be verified fastest as instantiating *above*. Verification is fastest, however, when the toothpaste tube is positioned over the end of the toothbrush with the bristles. This shows that hearing "above" does not always trigger a single idealized geometric simulation. Instead the arguments of "above" help select a simulation that is currently most appropriate. Thus, "the toothpaste tube is above the toothbrush" activates a configuration of *above* regions that differs from "The mercury in the thermometer is above 90" and "The moon is above your face."

Summary. Together, all of these findings suggest that simulators represent spatial relations. A given relation simulator has a dominant simulation that may be a geometric ideal. However it contains other simulations as well, each tailored to a specific context. Once one of these simulations becomes active, it creates a perceptual representation that directs attention to relevant regions of space. If the simulation becomes bound to a perceived situation, it provides an interpretation, specifying that a particular spatial relationship holds between the attended regions. In all these ways, relation simulators parallel property simulators.

This discussion of relation simulators has only addressed spatial relations. Clearly, though, other types of relations exist, too, including temporal, causal, and intentional relations. Similar analyses can be applied to them. According to perceptual symbol systems, any type of relation focuses on multiple space–time regions, and attempts to establish a particular configuration between them. For example, temporal relations represent configurations of space–time regions that vary in time. Thus, *before* highlights two nonsimultaneous events in a simulation, focusing attention on the first. Similarly, causal relations focus attention on causal components of entities and on the event sequences they produce as effects. Thus focusing on *gasoline* and *spark plug firing* in a *CAR* simulation activates subsequent simulated events that follow from their joint presence (e.g., *combustion, engine operation, driving*). The causal potency of *gasoline* and *spark plug firing* is established by assessing the counterfactual simulation in which *gasoline* and *spark plug firing* are absent, and finding that the event sequence isn't simulated (cf. Pearl, 2000). Thus assessing a complex configuration of regions across multiple simulations is necessary for establishing a causal relation.

Interpretation and Structured Representation

We have seen how perceptual symbol systems implement interpretation. When a property simulator becomes bound to a region of a perception or a simulation, it interprets the region as an instance of the property. Similarly, when a relation simulator becomes bound to multiple regions, it interprets them as an instance of the relation. In each case, interpretation carries inferential capability. Because a simulator accrues information about many instances across situations, it captures the broad content of the property or relation. This broad content then provides a wealth of inferences about any new instance bound to the simulator. For example, if a perceived nose is partially occluded by a scarf, the *nose* simulator can infer the rest of the nose during the property interpretation. Similarly, the *nose* simulator could infer nasal passages, breathing, and sneezing, with each inference being carried in a simulation that goes beyond the information perceived. In this manner, perceptual symbol systems achieve the basic functionality of type–token interpretation.

Structured representation is essentially a more complex form of type–token interpretation, where structure is accomplished by embedding simulations in one another (Barsalou, 1999). To see this, consider Fig. 15.6A. The left side depicts a visually perceived scene. The right side depicts simulators for *BALLOON, CLOUD*, and *above*. The middle depicts an embedded set of simulations that, first, form a structured representation, and second, interpret complex structure in the scene. Specifically, the *BALLOON* and *CLOUD* simulators run simulations that become bound to the relevant regions of the perceived scene (as indicated by the lines running from these simulators to their simulations to their referents in the scene). Simultaneously, the *above* simulator runs a simulation whose regions become bound to the regions containing the balloon and the cloud. Finally, the *BALLOON* and *CLOUD* simulations are embedded in the respective regions of the *above* simulation. This embedded simulation is formally equivalent to the standard amodal proposition:

above (*upper-region* = balloon-1, *lower-region* = cloud-1)

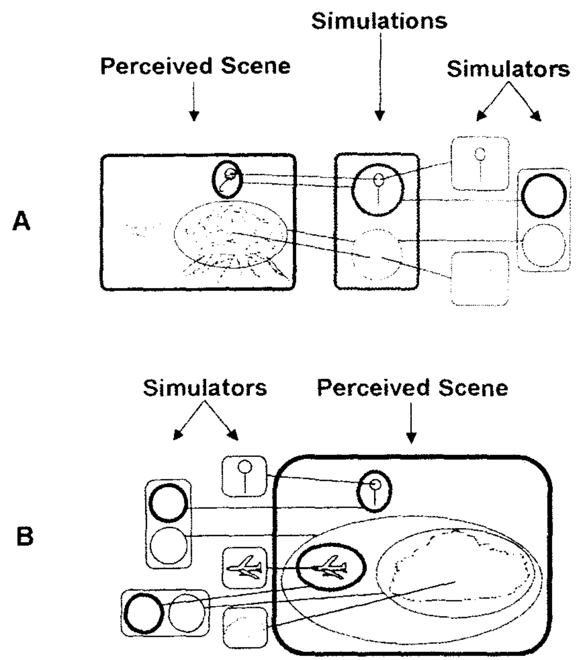

FIG. 15.6. Illustrations of structured interpretation (A) and recursive interpretation (B) in perceptual symbol systems. In (B), only simulators and the perceived scene are shown; simulations are assumed to exist but are omitted for simplicity.

Regions of the *above* simulation constitute arguments in a relation that become bound to embedded *BALLOON* and *CLOUD* simulations. Whereas classic propositions capture this conceptual structure explicitly, perceptual symbol systems capture it implicitly in the relations between simulations.

As Fig. 15.6B illustrates, this approach further implements the recursive embedding found in amodal propositions, where an argument of a conceptual relation takes another conceptual relation as a value. Thus the recursive proposition,

above (*upper-region* = balloon-1, *lower-region* = *left-of* [*left-region* = jet-1, *right-region* = cloud-1])

is implemented by having a *left-of* simulator map its two regions onto the jet and the cloud in the perceived scene, with this simulation then being embedded in *above's* lower region.

As these examples illustrate, perceptual symbol systems implement structured representations naturally and powerfully. By embedding simulations

in one another, binding arguments to values is accomplished readily—the hallmark of structured representations.

Holistic Simulations

One more piece of groundwork must be laid before returning to the issue of abstraction. Consider again the account of property verification depicted in Fig. 15.4. To verify that *JET* has a *nose*, a participant simulates *JET* and then attempts to map possible simulations of *nose* into the *JET* simulation. An unaddressed issue to this point is how the *JET* simulation arises. The possibility pursued here is that object simulators initially construct holistic simulations of instances from pre-attentive sensory representations.[5]

Holistic representations could include blob-like representations of an entity's global shape, extracted by low spatial frequency filters during visual processing (e.g., De Valois & De Valois, 1988; L. Smith, 1989). High spatial frequency information could further be captured at points in the image where such information exists (e.g., Morrison & Schyns, 2001). Holistic representations could also include primary axes, parsed subregions, and distributed configural features that capture direction and distance relations between subregions (e.g., Tanaka & Farah, 1993). Thus, the initial simulation of a *JET* could be a holistic representation of its basic shape, major axes, parsed subregions, and spatial frequency spectra.

Notably, no analytic properties of the sort described earlier exist in these early representations. These holistic simulations do not explicitly represent properties at the conceptual level—they only contain perceptual information. Instead such properties exist only after property simulators become explicitly bound to the holistic simulation (Fig. 15.4).

An additional possibility, though, is that some property and relation simulators become highly associated to particular category over time (e.g., the *wings* simulator becomes highly associated to the *JET* simulator). As a result, when holistic simulations of the category are produced, they quickly activate highly associated property and relation simulators, which fuse with the holistic simulation. Thus, the *wings* simulator might become active and run a simulation that enhances corresponding regions of the holistic *JET* simu-

[5]Holistic simulations are *not* bit-mapped recordings (cf. Hochberg, 1998). Much research on early vision demonstrates that pre-attentive sensorimotor representations are collections of features, not simply pixel-like representations. In early vision, for example, information about lines, surfaces, planes, orientations, and so on, are all extracted and coded as features. Thus early visual representations are interpretations themselves in the sense that detectors interpret subregions of the visual field as containing particular features. The difference is that these interpretations are pre-attentive and guide the formation of perceptual representations and experiences. In contrast, the interpretation in abstraction is attention driven and guides the learning of conceptual structures (simulators) in memory.

lation that is developing. As a result, the *JET* simulation becomes a mixture of holistic and analytic representations.

THE DIPSS THEORY OF ABSTRACTION

The previous sections have laid the groundwork for the DIPSS theory of abstraction (Dynamic Interpretation in Perceptual Symbol Systems). Again the goal is to account for the abstractions that represent categories such as *JET, BIRD*, and so forth.[6]

Loose Collections of Property and Relation Simulators

There are no static summary representations in DIPSS. Nor is there an attempt to construct a summary representation that perfectly describes all of a category's instances, or that provides a structurally coherent background theory. Instead the structural component of DIPSS is simply a loose collection of property and relation simulators. This collection is loose in two senses. First it's loose in the sense of being relatively unprincipled and open-ended. As Fig. 15.7A illustrates, a variety of property simulators develops to process the regions of a category's instances (and so do relation simulators, which are not shown). Thus simulators for *nose, wing, engine*, and *tail* develop to interpret the regions of *JETS* (and also to interpret other categories having similar properties). Typically, the property simulators that develop may be relatively unprincipled. A child learns those pointed out by speakers, and those that are functionally relevant in everyday activities. As attention is drawn to object regions via language and goal pursuit, simulators for them begin to develop. Although existing property words in a language may guide and constrain the properties learned, there is nothing particularly logical or systematic about the process.

Property simulators are also loose in the sense of the local simulations they contain. Although two people may have a simulator for the same property, its content may differ because of exposure to different category instances (Figs. 15.2 and 15.5). When different instances of a property become integrated in a simulator, they later produce different simulations, expectations, and inferences about the property. As a result, property information is far from a neat and tidy system across the speakers of a language. What remains in common is a shared awareness that certain regions of category instances have names and are important for goal-directed activity.

[6]Although the focus here is on the interpretation of objects, the same basic approach applies to the interpretation of physical events, mental events, settings, and so on. One notable difference is that the interpretation of physical events and settings often utilizes category simulators—not just property and relation simulators. To interpret instances of *EATING*, for example, categories such as *MONKEY, BANANA*, and *FORK* are required. Thus the interpretive system for physical events and settings requires the presence of object simulators, along with property and relation simulators.

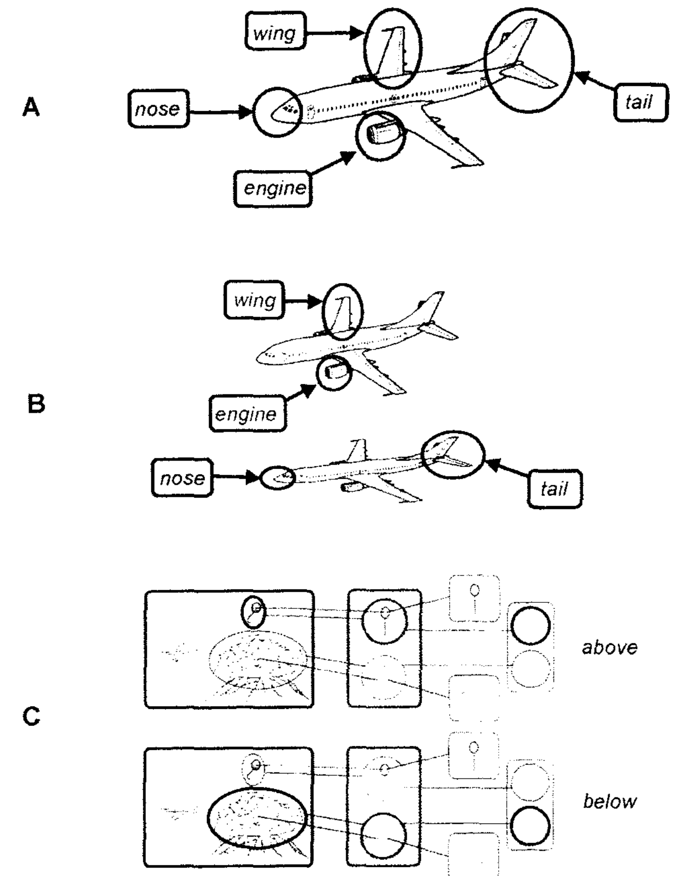

FIG. 15.7. (A) Illustration of the loose collection of property simulators that can interpret a category's instances. (B) Illustration of property simulators being applied dynamically to a category's instances. (C) Illustration of relation simulators being applied dynamically to a scene.

Dynamic Application of Property and Relation Simulators

As we just saw, DIPSS's first assumption is that people possess loose collections of property and relation simulators used to interpret category instances. The second assumption is that these simulators are applied dynamically—they are not applied identically across instances or occasions. As Fig. 15.7B illustrates, the property simulators used to interpret an instance vary widely. Whereas *wing* and *engine* might interpret a jet on one occasion, *nose* and *tail* might interpret a jet on another. Similarly, as Fig. 15.7C illustrates, two people could use different relation simulators—*above* and *below*—to interpret the same situation. Although a relatively fixed stock of simulators may exist for a person, the particular ones used to interpret a perception or simulation may vary considerably.

Abstractions. The subset of property and relation simulators applied to an instance on a given occasion can be viewed as an abstraction. This subset types, interprets, and structures aspects of an instance, and classifies it implicitly as something that the abstraction covers. However, such abstractions are not the classic sort of summary representation, because once an instance drops from attention, its abstraction on that occasion becomes largely irrelevant. The next time this instance or another is processed, a different abstraction may be constructed dynamically to interpret it. Thus abstractions are temporary online constructions, derived from an underlying set of property and relation simulators used to interpret many category instances.

Interpretative Attractors. Although the abstractions constructed to interpret instances vary widely, they are not constructed randomly from the available set of property and relation simulators. Due to frequency and recency, some simulators may be more likely to be applied than others. Simulators used frequently in the past will have an advantage, as will simulators applied recently. Furthermore, associations between simulators will produce correlations in the simulators applied. Thus if *steering wheel* is used for *CAR*, it may bring in other simulators associated with driving (e.g., *gear shift* becomes more likely, and *trunk* becomes less so). Furthermore, particular simulators may be associated with particular situations (e.g., *engine* and *oil* are associated with auto maintenance, whereas *trunk* and *stereo* are associated with travel). As a result, being in a particular situation activates relevant property simulators that bias interpretation.

The presence of attractors does not imply rigid interpretation. Even though attractors exist, a wide variety of factors may inhibit them and facilitate other interpretive strategies. As a result, the interpretation process retains a highly dynamic quality.

Category Learning as Skilled Online Abstraction

Again, no static abstractions reside in DIPSS. Instead, a loose collection of property and relation simulators is available for interpreting the instances of a category. On a given occasion, a temporary online abstraction is constructed dynamically to interpret the current instance. Across occasions, both statistical attractors and dynamic variability characterize the abstractions formed. Thus abstraction is more of a skill than a structure. As people learn about a category, they learn to interpret the properties and relations of its instances. With increasing skill, a person can effectively process more regions of instances, and know the most appropriate regions to process in a particular context. What develops permanently is not a fixed summary representation, but a skill for interpreting instances effectively and efficiently.

Explaining Abstraction

DIPSS naturally explains the three properties of abstraction presented earlier: type–token interpretation, structured representation, and dynamic realization. Type–token interpretation results from the application of property and relation simulators to the regions of perceived and simulated entities. As a result, implicit type–token propositions are established that carry a wealth of inferences from simulators. Structured representation results from embedding simulations in one another. The classic mechanisms of argument binding and recursion follow naturally from this process. Once a complex simulation exists, it can be bound to relevant regions of the instance to establish a complex structured proposition. Finally, dynamic realization results from the online application of a loose collection of property and relation simulators to category instances. On a given occasion, a subset of simulators interprets an instance, producing a temporary online abstraction. Across occasions, the abstractions constructed vary widely. The diversity of the resulting abstractions explains the problems associated with classic theories. It is impossible to identify and motivate any single abstraction as *the* summary representation of a category, because infinitely many are possible. Furthermore, none of them need to provide a complete account of the category—instead each simply interprets those aspects of an instance that are relevant in the current situation.

APPLICATIONS

DIPSS can be applied to a wide variety of abstraction phenomena in categorization, inference, background knowledge, and learning.

Categorization

Holistic Versus Analytic Processing. Much work has found that category learning begins holistically and then becomes increasingly analytic (e.g., Kemler, 1983; L. Smith, 1989). Early in learning, relatively undifferentiated representations of instances are stored, with no particular dimensions dominating. Later in learning, however, attention focuses on the most diagnostic dimensions for categorization, and they come to control learning.

According to DIPSS, simulators for properties and relations do not exist initially to interpret a category's instances. As a result, holistic representations established in early sensory processing dominate categorization. With increasing experience, however, attention focuses on the diagnostic regions of instances, and simulators come to represent the content of these regions. During later categorization, these simulators draw attention to the corresponding regions of new instances, thereby causing their content to dominate categorization.

Dimension Weights in Exemplar Models. Applications of exemplar models to learning data consistently find that some dimensions are weighted more than others (e.g., Lamberts, 1995; Medin & Schaffer, 1978; Nosofsky, 1984). These effects can be viewed as the result of simulators that represent dimensional values becoming differentially associated with categories. Thus if *shape* tends to take the form of *circle* in Category A, and *square* in Category B, then simulators for these values become associated with the categories.

What may be somewhat misleading about these learning paradigms is that they focus learners on a constant set of dimensions and values over the course of learning. As a result, learners apply simulators for them rigidly to instances. As DIPSS suggests, however, the construal of instances in everyday activity is typically much more dynamic. Rather than interpretation taking a rigid form across category instances, it varies widely as a function of experience and goals. Building such structure into category learning paradigms might yield more variable dimension weights than in previous studies. Instances are unlikely to be interpreted rigidly in everyday activity—their interpretation is probably much more dynamic.

Descriptive Inadequacy. Many theorists have noted the difficulty of specifying the properties that define a category (e.g., Wittgenstein, 1953). In classic work, Putnam argued that whatever description a person has for a category, it will never be sufficient to fix the category's reference (e.g., Putnam, 1973, 1975; also see Fodor, 1998; Margolis & Laurence, 1999). If the description turns out to be inadequate, the reference for the category often doesn't change, suggesting that something besides the description establishes membership. For example, if *WATER* turns out not to have the property, H_2O, but to have some other property instead, the physical things classified as *WATER* nevertheless tends to remain the same. The property, H_2O, never fixed reference and was therefore an inadequate description of category members.

DIPSS provides a natural account of descriptive inadequacy. Descriptions of a category are abstractions that arise from applying property and relation simulators in the available pool. Because this pool develops haphazardly, and because descriptions are constructed dynamically, it's no surprise that these descriptions never fix the category adequately. DIPSS embraces descriptive inadequacy. Because abstraction is a skill that supports goal achievement in particular situations, its purpose is not to construct summary representations that fix category membership.

The remaining question, though, is what does fix a category's reference as descriptions about it vary? As changes in scientific theories produce changes in the lay understanding of *WATER*, why does the reference of *WATER* remain basically the same? DIPSS explains this as the result of pre-attentive holistic representations. Low-level sensory representations of *WATER* are likely to remain relatively constant, as analytic properties and relations about it change. As beliefs about *WATER* come and go, the perception of *WATER* remains basically the same. Because these perceptions tend to be highly accurate in fixing category membership, they play the central role in everyday categorization, regardless of the analytic properties that currently reign.

Analytic properties certainly influence categorization. For example, Biederman and Shiffrar (1987) taught people analytic properties for chicken genitalia that facilitated chick sexing. Similarly, Lin and Murphy (1997) found that learning the functions of artificial objects influenced visual categorization. As such findings illustrate, categorization is not determined solely by holistic representations, but can be influenced by property and relation simulators as well. Nevertheless, it doesn't follow that these simulators completely fix reference; they only contribute to it partially, working together with holistic representations.

Inference

Feature Listing. The feature listing task has often been assumed to access and describe the underlying summary representation of a category. On this view, participants access a feature list, semantic network, or schema for a category, and then read out the information as verbal features.

According to DIPSS, however, no such underlying abstractions exist. Instead, participants run one of many holistic simulations of a category, and then attempt to interpret it using property and relation simulators. Rather than reading out a summary representation for the category, feature listing simply reflects one of many possible temporary abstractions that can be constructed online for a particular instance. Measuring these abstractions can be informative and useful (e.g., Wu & Barsalou, 2004), but they should not be viewed as accessing anything like a summary representation that covers the category descriptively or that fixes its reference.

Conceptual Instability. Barsalou (1987, 1989, 1993) reported that participants exhibit tremendous variability in accessing category information. In the feature listing task, different people produced very different features for the same category (an average overlap of only 44%). When the same person produced features for the same category on two occasions, the average overlap was only 66%. DIPSS explains this variability naturally. When different people produce features for the same category, they construct different holistic simulations, which leads to considerable diversity in the properties and relations used to interpret them. Even when two people construct similar holistic simulations, they may interpret it with different property and relation simulators. Analogously, when the same person produces features on two occasions, different features result for the same two reasons. Thus conceptual instability is a natural outcome of the dynamic simulation and interpretation process.

Script Tracks. As described earlier, a classic problem for Schank and Abelson's (1977) script construct was that are infinitely many script tracks. It didn't seem possible to construct a single summary representation for them all. According to DIPSS, such variability is to be expected, and the interpretative system should be geared to handling it. Because restaurant visits take infinitely many forms, a dynamical interpretive system—such as a loose collection of property and relation simulators—is needed to interpret them. As long as none of the properties or relations in a restaurant visit is new, a novel configuration of existing property and relation simulators can be configured to interpret it. Thus script tracks are a natural and desirable—not problematic—outcome in DIPSS.

Verbal Overshadowing. Schooler and his colleagues have demonstrated that verbally describing a perceptual stimulus interferes with remembering it later, relative to not describing it (see Schooler, Fiore, & Brandimonte, 1997, for a review). For example, describing a perceived face makes it less memorable than not describing it. According to DIPSS, the words in a description activate property and relation simulators, which in turn activate prototypical simulations that become linked to the perceptual stimulus (as in Figs. 15.5B bottom, 15.6, and 15.7C). For example, describing a face as having "big eyes, a long nose, and a full mouth" causes simulations of these properties' prototypical forms to be linked with the corresponding regions of the face's holistic representation. Later, at retrieval, these simulations become active, fuse with the holistic representation, and distort it. When no description is made, the holistic representation alone remains and provides superior information about the presented face. The dynamic application of property and relation simulators to holistic representations naturally explains verbal overshadowing phenomena. More generally, the interpretive process in DIPSS can be viewed as the general basis of encoding effects in episodic memory.

Background Knowledge

Intuitive Theories. Many researchers agree with Murphy and Medin's (1985) claim that intuitive theories of some sort provide background knowledge for categories. The problem has been little success in formulating these theories, and little agreement on the form they should take. DIPSS both explains this quandary and provides a solution to it. Just as there is no single script for an event, there is no single intuitive theory for a category. According to DIPSS, the background knowledge for a category is the loose collection of property and relation simulators used to interpret its instances, along with the skill to apply them appropriately in different contexts.

On some occasions, this interpretive system might produce an online abstraction along the lines of an intuitive theory. For example, property and relation simulators could be configured to explain how biological mechanisms keep an organism alive. Notably, however, a different abstraction might be constructed on another occasion for another purpose (e.g., to explain the reproductive origins of an organism). Over time, a diverse number of explanatory accounts of a category's members may be constructed. Analogous to script tracks, there is no single intuitive theory. Instead there is simply a system that can produce theory-like abstractions (among others) dynamically during the interpretation of category instances. Thus, the same person might construct vitalist, mechanistic, and psychological theories of *ANIMALS* on different occasions (Gutheil, Vera, & Keil, 1998).

Dimensions and Multidimensional Spaces. Theorists have often noted that a category's instances can be arranged in a multidimensional space (e.g., Gärdenfors, 2000; Rips, Shoben, & Smith, 1973). Various problems, though, have confronted this approach (e.g., Tversky, 1977). First, a given multidimensional space never seems to capture all of a category's properties. Second, the multidimensional space for a category is malleable, changing as the set of judged instances changes, and as the task changes.

DIPSS explains both the ability to construct multidimensional spaces and the problems they encounter. Multidimensional spaces are possible when subsets of property simulators have a higher order organization. For example, the simulators for *round, square, triangular,* and so on, can be organized into the higher order property, *shape,* which generally focuses attention on the exterior form of an object. Similarly, the simulators for *red, blue, yellow,* and so on, can be organized into the higher order property, *color,* which generally focuses attention on the surface appearance of an object.

When the simulator for a dimensional value becomes bound to a category instance, it activates the dimension. When another category member is encountered, the active dimension biases interpretation toward one of its values. As a result, the dimension comes to organize the set of instances.

This account explains the problems for multidimensional spaces. As dimensions come to guide the interpretation of instances, other properties and relations remain idle, and therefore don't show up in multidimensional solutions. The interpretive bias toward a few dimensions inhibits the use of other property simulators. As the instances and the task change, the dimensions activated change, such that the resulting multidimensional space reflects the system's current interpretive bias. Again DIPSS views these phenomena as natural products of dynamic interpretation—not as problems. They simply reflect the facts that not all potential property and relation simulators are used at once to interpret a category, and that the subset of simulators applied varies widely.

Analogy. When the same configuration of property and relation simulators can be applied to different categories, analogy becomes possible (cf. Gentner & Markman, 1997; Holyoak & Thagard, 1997). For example, a common configuration of simulators can be applied to holistic simulations of faces for *HUMAN, DOG,* and *FISH* to draw an analogy between them. Interpreting a given *HUMAN* face as having a *mouth*, a *nose*, and *eyes* results from applying the simulators for *mouth, nose,* and *eyes*, along with relation simulators, to a holistic simulation. Once this temporary abstraction has been constructed, the same configuration of simulators can be tried out on another holistic simulation, say a *DOG* face. If the configuration can be made to fit, an analogy is achieved.

The dynamic property of a simulator to produce different local simulations is central to analogy. The realization of a property or relation across two domains is rarely the same (e.g., different senses of *mouth* and *nose*). This variability is captured naturally by the idea that different simulations of a property reside in its simulator, one for each category. Furthermore, as an analogy is extended to new categories, the nature of the properties and relations that underlie it change, because the respective simulators acquire new local simulations (cf. Dietrich, 2000).

Learning

Novice Knowledge and Shallow Explanation. Keil and his colleagues have shown that people are over-confident about their understandings of how things work (e.g., Rozenblit & Keil, 2002; Wilson & Keil, 1998). For example, people believe that they understand how a zipper works, how a flush toilet works, and how a jet engine works. When asked for explanations, however, people can only produce partial, superficial ones. DIPSS explains overconfidence as resulting from people's ability to run holistic simulations. The ability to run a complete simulation of an object functioning from start to finish creates the illusion of understanding (e.g., zipping up a zipper). Conversely, DIPSS explains the shallowness of explanations as the result of having an insufficient set of property and relation simulators to construct adequate abstractions. As people began producing an explanation, they use the available property and relation simulators to construct a temporary online abstraction that interprets relevant regions of the holistic simulation. To the extent that a coherent configuration of simulators can't be assembled that covers the holistic simulation, the explanation fails.

More generally, DIPSS assumes that, early in learning, novices have a relatively limited set of property and relation simulators at their disposal. Furthermore, such simulators may mostly interpret "exterior" properties of perceived and simulated events. Novices may often fail to have simulators for key 'internal' properties whose causal properties give entities their functionality. Novices may also fail to have simulators for key relations that link components and events in causal chains.[7]

Expertise. With experience in a domain, learners acquire a much larger set of property and relation simulators. As a result, the depth and completeness of their explanations increases (cf. Chi, Feltovich, & Glaser, 1981). Experts can identify and interpret more critical regions in perceived events and simulations, and they can structure them in more sophisticated abstractions to form causal chains. By having greater ability to interpret and organize the regions of instances, experts also become better categorizers, moving their basic level down to the subordinate level (e.g., Gauthier, Skudlarski, Gore, & Anderson, 1999; Johnson & Eilers, 1998; Johnson & Mervis, 1997, 1998).

Theories of expertise generally assume that increased storage of exemplars, chunks, or rules speeds performance (e.g., Anderson, 1987; Logan, 1988; Newell, 1990). In DIPSS, the corresponding units are attractors for configurations of property and relation simulators. As an expert encounters instances in a domain, a configuration of simulators interprets each. Over time, the wide variety of configurations used to interpret most instances become well-established attractors and therefore highly accessible. As a result, relatively effortless performance becomes possible.

[7]Keil and Batterman's (1984) characteristic-to-defining shift may be a similar case. Initially during category learning, holistic simulations and simulators for exterior surface properties dominate categorization. Later, as simulators for internal causal properties are learned, they come to dominate.

Conceptual Change. DIPSS explains conceptual change as the result of an evolving set of property and relation simulators. As new simulators are acquired, the ability to interpret instances changes. Sometimes this change may appear gradual, as old-style interpretations are fleshed out with additional properties and relations. Change may be abrupt, however, when a new property or relation simulator is added that affects the use of other simulators. For example, when children add the property *gasoline* and *combustion* to the interpretive system for *CARS*, their interpretations of *CARS*'s behavior may change qualitatively. New online interpretations may be constructed that vary considerably from previous ones. To the extent that configurations of property and relation simulators become linked to form attractors, a large set of simulators may rise in dominance, while another large set falls. The result is a mini-revolution in interpreting the category.

ABSTRACTION REVISITED

We began with six senses of abstraction. The DIPSS account of Sense 3—abstraction as summary representations—can now be brought to bear on the remaining five senses.

Sense 1: Abstraction as categorical knowledge. According to this sense, knowledge of a specific category is abstracted from experience. In DIPSS, this amounts to establishing property and relation simulators that can interpret regions of perceived instances and holistic simulations.

Sense 2: Abstraction as the behavioral ability to generalize across instances. According to DIPSS, when people behaviorally state a generic, such as "Birds have wings," they have simulated a variety of *BIRD* instances, used the *wing* simulator to interpret these simulations, and then used language to describe the temporary online abstraction.

Sense 3: Abstraction as summary representation. Once a temporary abstraction exists for a category, a record of it becomes established in memory, increasing the likelihood of constructing the abstraction again in the future. Nevertheless the abstraction far from dominates the interpretive system that created it—it does not become part of a single summary representation for the category. It simply changes the dynamic qualities of the interpretive system, with the system remaining flexible and unsettled, such that future abstractions vary widely, each tailored to the current situation.

Sense 4: Abstraction as schematic representation. According to this sense, summary representations are sparser than exemplars, abstracting

critical properties and discarding irrelevant ones. DIPSS accomplishes this in three ways. First, the property and relation simulators that develop for a category never exhaust the simulators possible but instead constitute a relatively limited set. As a result, the interpretive system is schematic, only representing certain aspects of category instances. Second, the simulations that property and relation simulators construct typically contain far less information than the sensorimotor perceptions that produced them. Thus they are schematic in the sense of reenacting partial information and discarding details. Third, property and relation simulators are capable of producing idealized or caricatured simulations, thus being schematic in the sense of producing prototypical or diagnostic representations. Such simulations could result from the passive integration or averaging of information in a simulator, such that the most prototypical category information emerges as a dominant simulation (similar to Hintzman's, 1986, echo; also see Palmeri & Nosofsky, 2001).

Sense 5: Abstraction as flexible representation. According to this sense, summary representations can be applied flexibly to a wide variety of tasks. In DIPSS, this flexibility is not the result of a single abstracted representation, but of a dynamic interpretive system. As learning evolves, the set of property and relations simulators increases, as does skill in applying them to instances. The result is increased flexibility of interpretation, although attractors may produce ruts that work against flexibility to some extent.

Sense 6: Abstraction as abstract concepts. According to this sense, some concepts become increasingly detached from physical entities and increasingly associated with mental events (e.g., *truth*). In an extensive analysis, Wiemer-Hastings (2004) found that most abstract concepts refer to properties and relations—not to objects and events—suggesting that abstract concepts belong to interpretive systems. This fits well with Barsalou's (1999) proposal that abstract concepts pick out complex relational configurations of physical and mental states in background events.

For example, one sense of *truth* picks out a complex relation where one person makes a claim about the world to another person, who assesses whether the claim is accurate. For *truth* to apply in such situations, a speaker must make a claim, a listener must represent the claim, the listener must compare this representation to the world, and the representation must be accurate. When this complex relation exists, *truth* is a valid interpretation of the speaker's claim (also see Barsalou & Wiemer-Hastings, in press).

In general, abstract concepts often appear to capture complex configurations of physical and mental events in this manner. Analogous to relation simulators, abstract concepts interpret multiple regions of simulated and perceived events, and can thus be viewed as belonging to the loose collections of simulators that constitute interpretive systems. What distinguishes abstract concepts, perhaps, is the complexity of the relational information they capture, along with their frequent inclusion of mental states.

CONCLUSION

Although abstraction has gone out of fashion, it will not go away. Interpretation and structured representations are hallmarks of human cognition. The problem has been explaining these phenomena with mechanisms that are psychologically plausible and well suited for the job. Dynamic interpretation in perceptual symbol systems appears to offer promise in these regards. Although empirical evidence and computational models are necessary to realize this promise, the first step is to sketch its solution to the problem. It is hoped that this chapter has accomplished this goal.

REFERENCES

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem situations. *Psychological Review, 94,* 192–210.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 629–654.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101–140). Cambridge, England: Cambridge University Press.

Barsalou, L. W. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76–121). Cambridge, England: Cambridge University Press.

Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In E. Kittay & A. Lehrer (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization* (pp. 21–74). Hillsdale, NJ: Lawrence Erlbaum Associates.

Barsalou, L. W. (1993). Structure, flexibility, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A. C. Collins, S. E. Gathercole, & M. A. Conway (Eds.), *Theories of memory* (pp. 29–101). London: Lawrence Erlbaum Associates.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–660.

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. Conceptual representation [Special Issue]. *Language & Cognitive Processes, 18*(5–6), 513–562.

Barsalou, L. W., & Hale, C. R. (1993). Components of conceptual representation: From feature lists to recursive frames. In I. Van Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 97–144). San Diego, CA: Academic Press.

Barsalou, L. W., Pecher, D., Zeelenberg, R., Simmons, W. K., & Hamann, S. B. (in press). Multi-modal simulation in conceptual processing. In W. Ahn, R. Goldstone, B. Love, A. Markman, & P. Wolff (Eds.), *Categorization inside and outside the lab: Festschrift in honor of Douglas L. Medin.* Washington, DC: American Psychological Association.

Barsalou, L. W., Sloman, S. A, & Chaigneau, S. E. (in press). The HIPE theory of function. In L. Carlson & E. van der Zee (Eds.), *Representing functional features for language and space: Insights from perception, categorization and development.* Oxford, England: Oxford University Press.

Barsalou, L. W., & Wiemer-Hastings, K. (in press). Situating abstract concepts. In D. Pecher & G. R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought.* New York: Cambridge.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 13,* 640–645.

Bobrow, D. G., & Collins, A. M. (Eds.). (1975). *Representation and understanding: Studies in cognitive science.* New York: Academic Press.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes, 12,* 177–210.

Carlson-Radvansky, L. A., Covey, E. S., & Lattanzi, K. M. (1999). "What" effects on "where": Functional influences on spatial relations. *Psychological Science, 10,* 516–521.

Carlson-Radvansky, L. A., & Logan, G. D. (1997). The influence of reference frame selection on spatial template construction. *Journal of Memory and Language, 37,* 411–437.

Charniak, E., & McDermott, D. (1985). *Introduction to artificial intelligence.* Reading, MA: Addison-Wesley.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

Church, A. (1956). *The problem of universals.* Notre Dame, IN: University of Notre Dame Press.

Coventry, K. R. (1998). Spatial prepositions, functional relations, and lexical specification. In P. Oliver & K. P. Gapp (Eds.), *Representation and processing of spatial expressions* (pp. 247–262). Mahwah, NJ: Lawrence Erlbaum Associates.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition, 33,* 25–62.

De Valois, R., & De Valois, K. (1988). *Spatial vision.* New York: Oxford University Press.

Dietrich, E. (2000). Analogy and conceptual change, or you can't step into the same mind twice. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 51–77). Cambridge, MA: MIT Press.

Dretske, F. (1995). *Naturalizing the mind.* Cambridge, MA: MIT Press.

Farah, M. J. (2000). The neural bases of mental imagery. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (2nd ed., pp. 965–974). Cambridge, MA: MIT Press.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong.* New York: Oxford University Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28,* 3–71.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought.* Cambridge, MA: MIT Press.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3,* 191–197.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52,* 45–56.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20,* 1–55.

Goldstone, R. L., & Medin, D. L. (1994). The course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 29–50.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371–395.

Gutheil, G., Vera, A., & Keil, F. C. (1998). Do houseflies think? Patterns of induction and biological beliefs in development. *Cognition, 66,* 33–49.

Halff, H. M., Ortony, A., & Anderson, R. C. (1976). A context-sensitive representation of word meanings. *Memory & Cognition, 4,* 378–383.

Hampton, J. A. (1997). Conceptual combination. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 133–159). Cambridge, MA: MIT Press.

Haugeland, J. (1985). *Artificial intelligence: The very idea of it.* Cambridge, MA: MIT Press.

Haugeland, J. (1991). Representational genera. In W. Ramsey, S. P. Stitch, & D. E. Rumelhart (Eds.), *Philosophy and connectionist theory* (pp. 61–89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hayward, W. G., & Tarr, M. J. (1995). Spatial language and spatial representation. *Cognition, 55,* 39.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hochberg, J. (1998). Gestalt theory and its legacy: Organization in eye and brain, in attention and mental representation. In J. Hochberg (Ed.), *Perception and cognition at century's end: Handbook of perception and cognition* (2nd ed., pp. 253–306). San Diego, CA: Academic Press.

Holyoak, K. J., & Thagard, P. (1997). The analogical mind. *American Psychologist, 52,* 35–44.

Humphreys, G. W., & Forde, E. M. E. (2001). Hierarchies, similarity, and interactivity in object recognition: "Category-specific" neuropsychological deficits. *Behavioral & Brain Sciences, 24*(3), 453–509.

Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology, 82,* 472–481.

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia, 33,* 1419–1432.

Johnson, K. E., & Eilers, A. T. (1998). Effects of knowledge and development on subordinate level categorization. *Cognitive Development, 13,* 515–545.

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General, 126,* 248–277.

Johnson, K. E., & Mervis, C. B. (1998). Impact of intuitive theories on feature recruitment throughout the continuum of expertise. *Memory & Cognition, 26,* 382–401.

Kan, I. P., Barsalou, L. W., Solomon, K. O., Minor, J. K., & Thompson-Schill, S. L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. Invited article under review at *Cognitive Neuropsychology, 20,* 525–540.

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning & Verbal Behavior, 23,* 221–236.

Kemler, D. G. (1983). Exploring and reexploring issues of integrality, perceptual sensitivity, and dimensional salience. *Journal of Experimental Child Psychology, 36,* 365–379.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–394.

Kosslyn, S. M. (1994). *Image and brain.* Cambridge, MA: MIT Press.

Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General, 124,* 161–180.

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.

Langacker, R. W. (1986). An introduction to cognitive grammar. *Cognitive Science, 10,* 1–40.

Lin, E. L., & Murphy, G. L. (1997). Effects of background knowledge on object categorization and part detection. *Journal of Experimental Psychology: Human Perception & Performance, 23,* 1153–1169.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Logan, G. D., & Compton, B. J. (1996). Distance and distraction effects in the apprehension of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 159–172.

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review, 99,* 587–604.

Margolis, E., & Laurence, S. (1999). Concepts and cognitive science. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 3–81). Cambridge, MA: MIT Press.

Markman, A. B., & Gentner, D. (1997). The effects of alignability on memory. *Psychological Science, 8,* 363–367.

Martin, A. (2001). Functional neuroimaging of semantic memory. In. R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (pp. 153–186). Cambridge, MA: MIT Press.

Martin, A., & Chao, L. (2001). Semantic memory and the brain: Structure and process. *Current Opinion in Neurobiology, 11,* 194–201.

Martin, A., Ungerleider, L. G., & Haxby, J. V. (2000). Category-specificity and the brain: The sensory-motor model of semantic representations of objects. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 1023–1036). Cambridge, MA: MIT Press.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models.* Cambridge, MA: MIT Press.

McRae, K., & Cree, G. S. (2002). Factors underlying category-specific semantic deficits. In E. M. E. Forde & G. Humphreys (Eds.), *Category-specificity in mind and brain* (pp. 211–249). East Sussex, England: Psychology Press.

Medin, D. L., & Schaffer, M. (1978). A context theory of classification learning. *Psychological Review, 85,* 207–238.

Miikkulainen, R., & Dyer, M. G. (1991). Natural language processing with modular PDP networks and distributed lexicon. *Cognitive Science, 15,* 343–399.

Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review, 8,* 454–469.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104–114.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79.

Paivio, A. (1986). *Mental representations: A dual coding approach.* New York: Oxford University Press.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology.* Cambridge, MA: MIT Press.

Palmeri, T. J., & Nosofsky, R. M. (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *Quarterly Journal of Experimental Psychology. Section A, Human Experimental Psychology, 54,* 197–235.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, England: Cambridge University Press.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychological Science, 14*(2), 119–124.

Pollack, J. (1990). Recursive distributed representations. *Artificial Intelligence, 46,* 77–105.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Putnam, H. (1973). Meaning and reference. *Journal of Philosophy, 70,* 699–711.

Putnam, H. (1975). The meaning of "meaning." In H. Putnam (Ed.), *Mind, language, and reality: Philosophical papers* (Vol. 2, pp. 215–271). New York: Cambridge University Press.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80,* 1–24.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representation of faces. *Cognitive Psychology, 19,* 473–497.

Rips, L. J. (1995). The current status of research on concept combination. *Mind & Language, 10,* 72–104.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12,* 1–20.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*(5), 521–562.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations.* Cambridge, MA: MIT Press.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 7–57). Cambridge, MA: MIT Press.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Schank, R. C., & Colby, K. M. (Eds.). (1973). *Computer models of thought and language.* San Francisco: Freeman.

Schooler, J. W., Fiore, S. M., & Brandimonte, M. A. (1997). At loss *from* words: Verbal overshadowing of perceptual memories. *The Psychology of Learning and Motivation, 37,* 291–340.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21,* 1–54.

Simmons, K., & Barsalou, L. W. (2003). The similarity-in-topography principle: Reconciling theories of conceptual deficits. The organisation of conceptual knowledge in the brain: Neuropsychological and neuroimaging perspectives [Special Issue]. *Cognitive Neuropsychology, 20*(3–6), 451–486.

Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 800–811.

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review, 96,* 125–144.

Smolensky, P. (1988). On the proper treatment of connectionism. *The Behavioral and Brain Sciences, 11,* 1–74.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence, 46,* 159–216.

Solomon, K. O., & Barsalou, L. W. (2001). Representing properties locally. *Cognitive Psychology, 43,* 129–169.

Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition, 32,* 244–259.

Spalding, T. L., & Ross, B. H. (1994). Comparison-based learning: Effects of comparing instances during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1251–1263.

Spence, C., Nicholls, M. E. R., & Driver, J. (2000). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics, 63,* 330–336.

Talmy, L. (1983). How language structures space. In H. Pick & L. Acredelo (Eds.), *Spatial orientation: Theory, research, and application* (pp. 225–282). New York: Plenum.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 42,* 225–245.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science, 14,* 355–384.

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain, 107,* 829–854.

Wiemer-Hastings, K. (2004). *Abstract concepts.* Manuscript in preparation.

Wiemer-Hastings, K., Krug, J., & Xu, X. (2001). Imagery, context availability, contextual constraint, and abstractness. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society,* 1134–1139.

Wilson, R. A., & Keil, F. (1998). The shadow and shallows of explanation. *Minds & Machines, 8,* 137–159.

Winograd, T. (1972). *Understanding natural language.* San Diego, CA: Academic Press.

Winograd, T. (1975). Frame representations and the declarative-procedural controversy. In D. G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press.

Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review, 4,* 167–183.

Wisniewski, E. J. (1998). Property instantiation in conceptual combination. *Memory & Cognition, 26,* 1330–1347.

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). New York: Macmillan.

Wu, L., & Barsalou, L. W. (2004). *Perceptual simulation in property generation.* Manuscript under review.

Zeki, S. (1993). *A vision of the brain.* Oxford, England: Blackwell Scientific.