Talk about motion: The semantic representation of verbs by motion dynamics

Erin N. Cannon[1]

Paul R. Cohen[2]

[1]Department of Psychology; University of Massachusetts
Amherst, MA 01003 USA
Email: ecannon@psych.umass.edu

[2]Information Sciences Institute; University of Southern California
Marina del Rey, CA 90292 USA
Email: cohen@isi.edu

Contents

Introduction

Humans are perceivers and cognizers in an ever-changing dynamic world. Every moment is unique and different. How are we able to make sense of our experiences, label them with words, and speak in a way that is meaningful to others? If we label every situation uniquely, then the number of words in the human vocabulary would be infinite, making the utterance of a word not only uncommunicative, but essentially meaningless. Therefore, for purposes of effective communication with others, we cannot view every situation as unique. There must be commonalities between situations that call for the same words to be uttered in slightly different situations, and conversely, for words with slightly different meanings to be appropriate in overlapping situations. A situation, which we will call **s**, must be an abstraction of some kind. This chapter will explore the role of motion dynamics in making these abstractions available through perception, based on the assumption that dynamic real world movement is a reliable cue providing meaning about the world. In the case of action words (i.e., verbs), we assert that the dynamical movement of objects through space provides the semantics for the words we choose to utter.

The focus of this chapter is abstractions of patterns of movement that give rise to the utterance of verbs. That is, we consider the possibility that situations, **s,** are representations containing abstractions of movement patterns. We begin by putting forth a theory of word meaning suggested by Oates (2001), which is based on the ideas of pattern extraction, and a new way for cognitive scientists to view the questions of semantic language learning. We then review literature that spans the fields of social, cognitive, and linguistic development, which demonstrates that humans are remarkably sensitive to patterns of movement in space. We survey what is known about neonatal abilities to discriminate patterns of movement. Then we look at how different languages may influence movement patterns attended to. Finally, we present in detail one account of **s**, Cohen's *maps for verbs* framework, and discuss empirical evidence for it.

Word Meaning

The choice of words is conditional: One is more likely to say "dog" than "Thursday" when a dog is present, even if "Thursday" has a higher unconditional probability of being uttered. Informally, the choice of words is conditioned on the situation – a dog is present, or someone asks what day it is. It is difficult to think of situations that *determine* particular utterances. In general, a word has a *probability* of being uttered given the situation, which includes the words that have been uttered. Following Oates (2001) we define the *meaning* of a word as this propensity to be uttered in a situation. What does "Thursday" mean in a given situation? It means something in the situation makes "Thursday" a likely word to be uttered. In general, the probability of uttering word **w** in situation **s**, $\Pr(\text{utter}(\mathbf{w}) \mid \mathbf{s})$, is not the same as the probability that **s** is true given that **w** has been uttered – $\Pr(\mathbf{s} \mid \text{utter}(\mathbf{w}))$ – but these probabilities are proportional to one another, as any intuitive account of word meaning requires.[1]

The general form of this theory of word meaning might be right, but lacks three specifics. First, the probability that a word will be uttered depends not only on the situation but also on the

---

[1] From Bayes' theorem we have

$$P(\text{utter}(w) \mid s) = P(s \mid \text{utter}(w)) * P(\text{utter}(w)) / P(s)$$
$$P(s \mid \text{utter}(w)) = P(\text{utter}(w) \mid s) * P(s) / P(\text{utter}(w))$$

These expressions correspond to language generation and understanding, respectively. The first governs the probability that one will say a word in a given situation, the second is used to infer which situation holds given that a word is spoken. These conditional probabilities are clearly proportional, each is a scaled version of the other, where the scaling is by a ratio of two prior probabilities, the unconditional probability of the situation and the unconditional probability of uttering the word. For a given $P(\text{utter}(\mathbf{w}) \mid \mathbf{s})$ the probability of **s** given **w** is proportional to the unconditional probability of **s** and inversely proportional to the probability of uttering **w**. This latter condition is another way of saying that the word **w** carries information about the situation **s**: The less likely one is to utter **w**, the more likely it makes **s** given **w**.

speaker. What we really need is Pr(utter(**p,w**) | **s**) for every person **p**. Of course, we cannot have this information, so we must approximate it. Oates (2001) describes how to make the approximation. Second, this simple theory of word meanings does not explain how compositions of words (e.g., sentences) have meanings. This chapter says nothing about syntax and the composition of words into sentences. Third, the theory does not specify the elements of situations that go into **s**, the propositions on which word choices are conditioned. However, by bridging the fields of cognitive development and language acquisition, we can hypothesize and test potential candidates for **s**. This is the goal we set forth in this paper, and for guiding future research.

We do not suppose that patterns of movement are the *only* elements of situations **s** on which word choices are conditioned. Presumably **s** contains other physical observables such as the number, shape, and classes of objects. Complicating the story, **s** might also contain *unobservable* elements, particularly attributions of beliefs and goals. Suppose one observes George walking down the street a few yards behind Fred. The word "follow" is ambiguous in this context. It might mean only that George is walking behind Fred, or it might mean George intends to walk behind Fred and go wherever Fred goes. Let us assume that nothing in Fred's or George's observable behavior indicates that George is following Fred in the second sense of the word, and yet a speaker, observing the scene, decides to use this sense of "following"; indeed, the speaker might even say, "tailing" or "stalking," or some other word that indicates George intends to stay close to Fred as he walks along. If the choice of words is conditioned on a representation of the situation, **s**, then **s** must contain an attribution of George's intention to remain close behind Fred. Of course, this attribution might be wrong (e.g., a false belief), but is an element of **s**, contributing to the word choice uttered.

Furthermore, intentional words complicate an otherwise straightforward theory of the acquisition of word meanings. If word choices are conditioned on *observable* aspects of the situation, **s**, then a child could learn word meanings by associating words with situations, that is, by learning conditional probabilities $\Pr(\text{utter}(\mathbf{w})|\mathbf{s})$. However, if word choices are conditioned on *unobservable* aspects of situations, then associative learning is more difficult. Suppose a child observes a dog running after a squirrel while her mother says, "The dog is chasing the squirrel." One can see how the child might learn to associate "chasing" with the observable, physical aspects of the scene – both animals are running, when the squirrel changes direction the dog does, too – but how can the child learn that "chasing" implies something about the intentional states of both the dog and the squirrel, when these states are not observable? Presumably, at some point in the child's development, she is able to supply these unobservable elements, herself. She *imagines* the intentional states of the animals and associates these states with the word "chasing." The problem with this theory is that it is difficult to prove, because it asserts that the child conditions her word choices on intentional states she imagines, and we cannot observe what she imagines. More concretely, we cannot be sure that, to a young child, "chasing" does *not* mean only the physical aspects of chasing, nor can we easily discover when, in the child's development, the meaning is extended to include intentional aspects of the situation.

In fact, it is difficult to interpret some of the literature that seems relevant to our claim that word choices might be conditioned on patterns of movement. The general problem has this schematic form: Infants or older children are shown to discriminate patterns of movement, say P1 and P2, which adults label with intentional terms, such as "avoid" or "pursue." Presented with P1 and P2, what discrimination is the infant, child, or adult *really* making? The adult might be comparing the raw movement data, P1 vs. P2, or she might be comparing her intentional interpretations of P1 and P2, or both. In one case we say the adult discriminates the dynamics of the displays, in another we say the adult discriminates "avoid" and "pursue." We do not know which is true, and both might be. The same goes for the infant and the child: We cannot say *when* or even *whether* intentional attributions inform discriminations of displays, particularly when displays might be discriminated based on (even subtle) differences in dynamical motion. We should not assume that, because adults make intentional attributions to displays, the child's ability to discriminate entails discriminating intentional states.

Review of the Literature

We begin with Heider and Simmel's (1944) classic demonstration that patterns of movement evoke rich linguistic descriptions. Evocation is a phenomenon, not an explanation. We cannot say *why* subjects find so much to say about Heider and Simmel's displays. However, the only information-carrying aspect of the display is the relative movement of a few shapes. The lengthy and imaginative stories about the displays must be cued somehow by these movements. Next, we review work based on point-light displays, which shows that humans can reliably extract movement information in the absence of shape cues. Having established humans' sensitivity to patterns of movement, we build a case that these patterns support *semantic* distinctions, including differences in word meanings. Infants can discriminate patterns of movement generated by different classes of things, and young children appear to discriminate causal from non-causal movement in launching events. The patterns available to neonates are candidates for elements of **s**, the situation descriptions on which probabilities of uttering words are conditioned. This literature gets us ready for linguistic theories in which word meanings are grounded in physical dynamics. We review these theories, including developmental arguments. We then discuss the ways in which a scene is parsed into meaningful motion-based components, which will inform **s**. In conclusion, further candidates for the semantic core are suggested in P. Cohen's Maps for Verbs framework.

Patterns of Movement Evoke Intentional Descriptions

In Heider and Simmel's (1944) classic study, adults were shown a film clip of three shapes in motion. The adult participants created elaborate storylines describing the interactions, even though the only information in the stimuli was object shape and motion. Human-like characteristics were easily attributed to the triangles and circles, including intentional states. Moreover, the often overlooked phenomenon discovered in this study is that the attributions given to each shape were highly similar across participants. All reports included common event features: a fight scene, a chase scene, and a scene in which one object became trapped in the house and tried to escape. Thus, not only did these simple motion patterns elicit detailed anthropomorphized descriptions and storylines, but the actual verbal reports were similar. Although Heider & Simmel did not test for similarities between particular utterances, their findings suggest that movement patterns may predict *which* intentional descriptions are attributed to them.

If adults have tendency to extract intentional attributes from patterns of movement or events, then so might children. Berry & Springer (1993) tested three- to five-year-olds to investigate the influence of motion dynamics on anthropomorphic attributions. Four groups of children were tested systematically. One group received the original Heider and Simmel movie, another received the movie with the object shapes obscured, preserving only the motions; the third group received static displays taken from the movie, with shapes and figure information preserved; and the last group received static displays where both shape and motion were obscured. The experimenters obscured the shapes of objects to rule out the possibility that object shape or size contributed to the characteristics attributed to the objects. While watching the film, children were asked, “What do you see?” Like adults, children attributed intentions to the objects in the movies, and were about five times more likely to use anthropomorphic language, including intentional attributions, than children who were shown static displays. Shape did not seem to be a relevant factor in the intentional attributions. Clearly then, by the age of three, motion is a sufficient cue to determine word choices whose meanings convey intention.

Two factors make these findings quite compelling. First, an understanding of intentionality is a prerequisite to children’s theory of mind (TOM; e.g., Leslie, 1984), yet three-year-olds have difficulty understanding that other people’s intentions may vary from their own (particularly about beliefs, it may be less difficult for desires; see Bartsch & Wellman, 1995; or Flavell, 1999, for review of TOM literature). It is curious, then, that young children so adamantly ascribed intentional states (as indicated by their word choice) to the moving shapes in the Heider and Simmel movie. Berry & Springer did find a trend toward increasingly anthropomorphic descriptions with age, but it did not reach significance. It might be fair to say this that some portion of the anthropomorphic descriptions then did come from three-year-olds. Second, the task was not forced-choice. Children gave open-ended descriptions of the films they watched. These children were young language learners, with a far more limited vocabulary than adults. Yet even by the age of three, their choice of words to describe the scene was remarkably adult-like with respect to intentional attributions. This suggests that the children were no less able than adults to extract the motion patterns that elicited their word choices.

More compelling is that even preverbal infants show an ability to extract intentional information from movement patterns (e.g., Golinkoff & Kerr, 1978; Legerstee, Barna, & DiAdamo, 2000; Leslie, 1984; Spelke, Phillips, & Woodward, 1995; Woodward, 1998).

Intentional attributes have been suggested in habituation and violation-of-expectation paradigms focused on the understandings of goal-directed actions and concepts of agency. Both goal-directedness and a concept of agency implies that intentionality is involved in a scene. One difficulty, however, is the confound of infants' familiarity with human actions. Humans are inherently agents, thus intentional beings, and are also often the subjects in these experiments. However, non-human and inanimate objects have been successfully utilized to serve as "agents" in motion events also (e.g., Cohen, Rundell, Spellman, & Cashon, 1999; Cohen & Oakes, 1993). In some cases, infants *may* perceive inanimate objects as intentional, based solely on particular motion characteristics such as self-propulsion and trajectory (Baron-Cohen, 1994; Premack, 1990), or by moving along a trajectory through space in a "rational" manner (Csibra et al., 1999)

As touched upon in the introduction, we cannot be sure that the discrimination of intentional states is the same in early childhood and infancy as it is in adulthood. The early discrimination of goal direction and agency early in life, based from motion dynamics is, however, suggestive that attributions of intentionality might begin prior to the first words being uttered. It could be that some unknown is present in the motion dynamics, or that something draws the infant to attend to particulars of the motion specifying intentionality. Children may learn to attach intention-loaded words to these motions, perhaps even before they fully understand the implications of that particular word. As vocabulary increases, so does the child's understanding of intentionality, which probably develops from a motion-based understanding, to more a psychologically-based and adult-like one. In experiments such as Heider and Simmel's, perhaps the motion-based elements in **s** are substantial enough to elicit the intentional words that were associated with them earliest in development.

Sensitivity to Patterns of Movement

The work of Johansson (1973) proposed that the visual system parsed biomechanical movement presented in a point-light displays, into two separate types of motion: *common* motion, from which the trajectory of the group of lights relative to the observer is perceived, and *relative* motion, the invariant relations between these lights, from which structure, or figure, is perceived. Indeed, using similar point-light displays, Bertenthal, Proffitt, & Cutting (1984) found infants as young as three months discriminated biological motion, specifically the relative motion patterns of human walkers. In a habituation (with partial lag) experiment, infants were able to discriminate upright human walkers from inverted human walkers, but they could not make this discrimination when tested with static light displays. The infants evidently extracted figural coherence from information in the moving displays. In a second experiment, absolute motion was held constant, and thus the only motion information available was the relative motion from the light points. In this experiment, infants were able to discriminate the real walkers from anomalous, scrambled points of light. Moreover, infants were not using *absolute* motion cues in the detection of biomechanical motion. These findings suggest that perception of patterns of relative motion is functioning early in life. It is not unreasonable to assume that this information is extracted and utilized to inform and create semantic representations about the world as the child experiences it.

Additionally, Bertenthal (1993) suggested there might be several other processing constraints responsible for biomechanical motion perception that are available to the perceptual system early on. For instance, a slight spatial discrimination seems not to affect infants' discriminations of biological motion (disruptions of local rigidity), but temporal disruptions in the movements of individual points of light do in fact disrupt this perception. Bertenthal & Pinto (1994) found similar results when testing adults; temporal disruptions made to the individual points of light impaired the perception of biological motion, more so than spatial disruptions, supporting the idea that motion is extremely important in information extraction. In addition, the influence of stimulus familiarity also constrains biomechanical motion perception (Bertenthal, 1993). When tested with non-human biological motion, in this case, spiders, three-month-olds discriminated inverted displays from upright ones but five-month-olds did not. Bertenthal attributes this discrepancy to a shift in perceptual processing by five months to a level based on "perceived meaning" (p. 209).

Sensitivity to specific patterns of motion containing meaning is not exclusive, however, to biological motion. As discussed earlier, the non-biological pattern of motions presented by Heider & Simmel (1944) elicited responses *as if* the objects themselves were "biological". Guyulai (2000) found that manipulating the kinetic patterns of movement between objects in a 2-D movie display influenced the attributed meanings (based on specific questions asked to participants about the event) more so than changes to object hue, size, shape, or luminance. Other perceptual cues did not change the overall impression. It appears that kinetic patterns amongst objects (or points) really do influence how we perceive the content or meanings of events.

Semantic core and patterns of movement

In the introduction to this article we suggested that word choices are conditioned in part on representations of the current scene, which we denoted **s**. Representations are constructed from elements, and we are particularly interested in the most *primitive* elements, the ones infants might have or learn. Several cognitive scientists think these elements may be learned through interaction with the physical world (Barsalou, 1999; Johnson, 1987; Mandler, 1992; 2000). In the following sections we will survey some candidates for these primitive representational elements, which we call the semantic core, and then show how these might serve to specify the meanings of words (P. Cohen et al., 2001; Oates, 2001). We are particularly interested in those primitive semantic distinctions that can be grounded in patterns of movement.

Motion and Causality

Michotte (1963) suggested that the perception of causality could be manipulated. His simple animations of two squares interacting suggested causality is perceived directly, without cognitive interpretation. Of particular interest here is the *launching event*. Perceived as a whole-body interaction, a launching event is one in which object A moves toward a static object B, stops at the point of contact, and then object B appears to be set into motion as a result. Adults report perceiving this sort of event as causal, in that object A caused the movement in object B. When Michotte manipulated temporal and/or relative velocity patterns, interactions were perceived as qualitatively different. For example, if object B began to move within 70 ms of contact, its movement was perceived as causally related to the interaction with object A. If object B moved after 160 ms, then its movement and A's movement were perceived as disconnected, not causally related. Similarly, manipulating the gap between the two objects just prior to the movement of the second one, or their velocities, affected whether the interactions were perceived as causal or separate autonomous movements. Thus highly specific spatio-temporal features of interactions affect whether events are perceived as causal or not.

The ability to detect spatio-temporal features of interactions is present early in life (Leslie, 1982, 1984). Young infants tested in a habituation paradigm were shown Michottian launching events, with manipulations of delays at contact and spatial gaps. Leslie (1984, 1988) suggested the ability to detect the internal structure of a launching event was present by six months of age. Six-and-a-half month olds habituated to a launching event then dishabituated to events involving a spatial gap plus a temporal delay. However, infants habituated to a delayed launch did not dishabituate to scenes involving a spatial gap, and vice versa (Leslie, 1984). These infants showed sensitivity to specific disruptions in spatio-temporal continuity. Leslie & Keeble (1987) supported this notion by reversing the direct and delayed launching events. Six-month-olds were habituated to a film clip of a red square directly launching a green square. Then the clip was played backwards. The reasoning goes that a causal event (the direct launch) involves an agent (the causer of an action) and a recipient of that action. Reversal of the causal event involves a reversal also, of the mechanical roles. A second group of infants was habituated to a delayed launch, then tested on the film played backwards. If the event was not perceived as causal, then there should be no change in role reversal either. The hypothesis was confirmed; infants dishabituated in the direct launching condition, but not to the action reversal in the delayed launching. Leslie and Keeble (1987) concluded that infants discriminated on the basis of causal relations.

Whereas Leslie wants to argue from a modularity perspective that causality is a primitive concept (e.g., Leslie, 1994), the more recent work of L. Cohen and colleagues (e.g., L. Cohen & Oakes, 1993; L. Cohen & Amsel, 1998; Oakes, 1994) suggests that the perception of causality is actually developmental and is built up from simpler percepts. In terms we introduced earlier, the semantic core would include these simpler percepts and the launching event itself would be what we have called **s**, the situation. Here we will briefly review the evidence that infants perceive components of a launching event.

Cohen & Amsel (1998) investigated the development of causal perception for infants slightly younger than those used in Leslie's (1984) experiment. They tested for changes in habituation from direct launching events to both types of non-causal events – those with a temporal delay and those with a spatial gap. Note that these discriminations are more finely tuned than Leslie's noncausal events involving both spatial gaps and delays. They found that four-month-olds did not dishabituate to non-causal events, but showed a general preference of looking to causal events. By five-and-a-half months, infants dishabituated to change in any feature, causal or non-causal. By six and a quarter months, the infants dishabituate on the basis of causality only. Oakes (1994) also found that by 7 months, infants discriminate based on causality only, and not in changes in independent features.

However, the ability at six and seven months of age to discriminate events on the basis of causality is not particularly strong. At this age, it is fairly situation-specific. For example, Oakes & L. Cohen (1990) tested the perception of causal events using complex stimuli, more like objects in the real world (as opposed to animated squares and such). Six month-olds did not dishabituate on the basis of causality in this case, but 10-month-olds did. Furthermore, Oakes (1994) found that 7-month-olds did not discriminate on the basis of causality when the paths or trajectories of the objects in the event varied. By 10 months, infants were not bothered by changes in path, but did discriminate on basis of causality. But even at this age, L. Cohen & Oakes (1993) argue that causality is still somewhat tied in with object perception. For example, 10-month-olds tended to respond differentially to changes in identity of the objects before generalizing the event in terms of causality.

Taken together, the literature on perception of physical causality suggests that, by the end of the first year, causal perception is nearly adult-like. Furthermore, it has a developmental trend: There is an initial preference for responding to causal events, perhaps making the infant pay attention to them. Then, early on, there is detection of subcomponents of the event. This is the time at which infants learn which features of scenarios make up causal versus noncausal events. These spatial and temporal features are perhaps components of the semantic core, as each component conveys meaning. Once the child can assemble them into representations of situations, **s**, responses tend to be no longer based on the individual features themselves, but rather on the basis of **s**. However, instances of **s** are initially situation-specific, then abstracted, as other developing elements of **s** (such as object and agency concepts, which happen to also draw upon spatiotemporal components of the semantic core) are also refined.

Motion and Classification

As described, a situation **s** can be parsed into elements of the semantic core. We have seen that elements of a situation are the basis for judgments of physical causality. Now we consider elements that might account for both object and action *classes*.

The categorical distinctions children (and perhaps infants) make are based on the different types of motion patterns that become associated with a particular class. We are certainly not the first to make this claim (see Lakoff, 1987; Mandler, 1992, 2000; Rakison & Poulin-Dubois, 2001). A central example is the animate-inanimate distinction. Mandler (1992) proposed a semantic core composed of primordial image-schemas to account for the animate-inanimate class. These schemas are based on motion properties, such as motion trajectory, in relation to ground and other objects, and self-propulsion. Rakison & Poulin-Dubois (2001) provide a different perceptually-based associationist explanation of how the distinction develops, which includes the properties Mandler asserts, in addition to properties such as goal-directedness and agency. Others have also considered animacy as derived from object motion in the absence of physical or mechanical causality (Leslie, 1994; Premack, 1990). For example, Premack (1990) suggested that if an object's change of movement is self-propelled, and not due to the movement of any other objects, then it is perceived as intentional. If both objects are self-propelled, then they might be perceived as one object being directed by the goal to affect the other object.

One issue is whether the animate/inanimate distinction is purely perceptual or whether it is knowledge-based. Perceptual categorization is based only on physical features of objects, and requires no knowledge of object function, or of what the object *is*. Mandler (1992, 2000) proposed that the behaviors demonstrated by young children are guided by conceptual knowledge about objects in the physical world, an understanding of what they are. Mandler suggested that conceptual knowledge is produced by perceptual redescriptions based on primordial *image-schemas*.

Much of the animate-inanimate distinction research has been based on discrimination between two domains: animals and vehicles. Objects in these domains can be perceptually similar (e.g., birds and airplanes) or perceptually dissimilar (horses and motorcycles). However, the motion patterns of the animal domain are different from the motion patterns of the vehicle domain. For instance, the pendular motion of animals is quite different from the rotary motion of vehicles. While much research favoring Mandler's conceptual knowledge has involved the extended imitation paradigm (e.g., Mandler & McDonough, 1996), and has found children to make distinctions toward the end of the first year, it is unclear that motion cues are the basis. The objects tested are not actually moving in the experiment. It is quite possible that the distinction is made early, but the nature of the paradigm makes this difficult to test. The image schemas however, are not necessarily "knowledge rich" in the sense that this paradigm tests for. Image schemas are dynamical – about movement and change. They are the semantic primitives that distinguish situations, **s**, thus organizing the knowledge acquired in these learning situations. An alternative approach, the use of point-light displays, has been an effective means of determining whether motion cues alone are a sufficient basis for the classification of animals and vehicles.

Arterberry & Bornstein (2001) tested 3-month-old infants in a multiple exemplar habituation paradigm to search for evidence of a possible categorical distinction between animals and vehicles made at this early age. Furthermore, they tested whether this distinction was based primarily on the dynamic motion features inherent in these domains (by using point-light displays of animals and vehicles in motion), or on static featural information (pictures of animals and vehicles). The infants in both conditions dishabituated to novel categories, suggesting that they are making the animal-vehicle distinction early. Because they dishabituated in both the static and dynamic conditions, an animate/inanimate distinction could not be claimed. The figural features in the static pictures, such as legs versus wheels, could not be ruled as a basis for classification in this study.

In a similar task, Arterberry & Bornstein (2002) tested six- and nine-month old in the same paradigm. Six-month-olds again showed the ability to categorize animals and vehicles based on either static or dynamic features. However, only nine-month-olds showed transfer between these display modalities. Nine-month-olds who were habituated on dynamic motion features were then able to transfer this knowledge to static displays at test. However, if the nine-month-olds were habituated to static displays of animals or vehicles, they did not transfer the categorical distinction when tested with dynamic motion displays of those animals or vehicles. This suggests that 1) there is a developmental aspect to this categorization, 2) dynamic motion conveys more transferable information than the figural features available in static displays, and 3) the transference of discriminations based on dynamic features to static displays suggests that the children somehow "connect" the figural information in the static displays with the dynamic information. The ability fits nicely into our theory that dynamic features represented in the semantic core are easily transferred into new instances of **s**.

Linguistic Research and Cognitive Semantics

Thus far, we have discussed possible elements of **s**, the situation description which is constructed from elements of a semantic core. We have focused on psychological evidence that the semantic core contains abstractions of patterns of movement. We have not discussed linguistic issues, particularly our characterization of word meaning as the conditional distribution of a word given situations **s**. In this section we review evidence that patterns of motion influence the choice of words, that is, the proposition that **s** contains representations of patterns of motion.

Talmy coined the term "force dynamics" (1975, 1988, 2000), to denote a semantic category that covers a full range of relations that any object or entity can have with respect to some force imposed on it. Force dynamics pertains to motion events involving two objects that are broken into linguistic primitives of causation, but further allows for other concepts such as letting or resisting. Talmy's framework includes such concepts as the exertion of force, amount of resistance, obstructing force, and overcoming resistance. Talmy (1975) claimed that there are universal structures in all languages, reflecting motion situations in which one object is moving or located with respect to another object. The motion situation is universally encoded by the following four components: 1) Figure, 2) Ground, 3) Path and 4) Motion. Of particular interest here to the issue of verb usage are Path and Motion. Figure and Ground are typically expressed as nouns. Talmy (1988, 2000) described *verb-framed* languages as those that conflate path with motion, meaning that verbs usually express or encode path. Spanish is an exemplar. In contrast, *satellite-framed* languages tend to conflate manner with motion, as in English. Work by Naigles et al. (1998) found these typological differences in verb usage demonstrated by English and Spanish adult speakers when presented dynamic motion events.

Developmental linguistics

The story so far is that motion is an important component of word meanings. It is one of the elements of situations **s** that influence the probabilities of uttering or hearing particular words. On this account, learning word meanings is just learning conditional probability distributions Pr(utter(**w**) | **s**). However, this account cannot be complete, because it does not explain why children in different language communities do not learn the particular kinds of words in roughly the same order. Let us assume that American (native English-speaking) and Korean (native Korean-speaking) children have roughly the same experiences: Both live in a world of surfaces, objects, movements, physical influences and control, animate and inanimate motion, and so on. Thus, the situations **s** to which the children are exposed are the same. The words to which they are exposed are different, but the *kinds* of words – nouns, verbs, and so on – are not. Let us modify our account of word meaning a little to include word classes: The meaning of a particular verb class, say, is just the probability distribution over uttering a verb in that class given the situation: Pr(utter(verb class) | **s**). If lexical learning is no more than learning these conditional distributions, then Korean and American children should learn identical distributions for identical word classes. After all, the children are exposed to the same situations, **s**, so if both learn a particular verb class **v**, they should learn the same conditional distributions Pr(utter(**w** in **v**) | **s**). However, American and Korean kids do not map elements of **s** to word classes in the same way, nor do they learn instances of word classes in the same order.

Choi and Bowerman (1991) found evidence that nouns are not always acquired before verbs, as previously thought (e.g., Gentner, 1978, 1982). Diary accounts of English and Korean learning children were examined, and differences in verb acquisition tended to reflect the language they learned. They suggested an interaction between young children's linguistic input (i.e., the language they are learning) and cognitive development. Korean is a verb-framed language, in which Path is typically expressed in the main verb and manner expressed separately. English, a noun-based, satellite-framed language, expresses Manner in the main verb and Path separately. Choi and Bowerman (1991) concluded that an initial sensitivity to the semantic structures of a language is responsible for differences in language acquisition. A simple mapping of learned words to semantic elements (e.g., Slobin, 1973) cannot fully account for the meanings of children's spatial words (in this study) being language-specific. Learning the lexicon might in fact mean learning conditional distributions $\Pr(\text{utter}(\mathbf{w} \text{ in } \mathbf{v}) \mid \mathbf{s})$, but we still must explain how a Korean word class is conditioned on the element of $\mathbf{s}$ we call Path while the same word class in English is conditioned on an element of $\mathbf{s}$ called Manner.

The work of Tardif and colleagues (Tardif, 1996; Tardif, Shatz & Naigles, 1997) suggested noun-verb differences in language acquisition between English and Mandarin learners could be explained by looking at the linguistic input (e.g., proportion of nouns and verbs spoken) from the caregiver. Mandarin-speaking caregivers tended to produce more verbs than nouns when speaking to their children. In turn, this bias was reflected in children's vocabulary development. Work by Hoff (2003) found that environmental input factors, other than language *type*, should also be considered. Within the English-speaking population, her work has found influences of maternal speech (i.e., linguistic input) on vocabulary development as a function of socioeconomic status (SES). Specifically, children with higher SES had vocabularies that were larger and faster growing than lower SES children. Differences were present by the age of two, and were linked to the frequency and length of mothers' utterances to (and with) the child.

Tomasello (1992, 1995) further emphasized the importance of the social context in verb learning, pointing out that children best learn from the observations of other people's actions and through their social interactions with others.

In addition to vocabulary development, Gopnik & Choi (1995) have shown a direct effect of language's influence on development of cognitive structures. Korean mothers tend to use more relational terms and action verbs when talking to their children, whereas English-speaking mothers tend to initially label objects most often with their young children. They noted that Korean children have a "verb spurt," analogous to the noun-spurt in learners of the English language. Consequently, these differences in vocabulary were reflected in children's cognitive development. Korean children showed means-ends skills earlier than English-learning children, but the English-learning children showed more advanced skills in an object categorization task.

In sum, the idea that lexical acquisition involves learning conditional probabilities Pr(utter(w) | **s**) is not necessarily wrong, but it does not explain how individual languages select particular elements of a situation **s** to serve as the features that condition word probabilities. We have already seen that Manner is a conditioning element of **s** for English verbs whereas Path is a conditioning element of **s** for Korean verbs. Nothing in our theory of word meanings yet explains this difference.

Parsing the Scene

The challenge is to explain how elements of the semantic core – the most primitive distinctions – are collected into situation descriptions **s**, and to explain why these elements are bundled in different ways in different languages. We assume that all humans have access to the same elements of the semantic core; for example, American and Korean children are equally able to detect the Path or Manner of motion. It might be that the apparent differences in how English and Korean bundle elements of the semantic core are all explained by simple associative learning. This is how it might work: An English-speaking child and a Korean child are both observing the same situation, and both hear verbs with essentially the same meanings, but the best account of the verb meaning for the English speaker is obtained by conditioning the probability of the verb on elements of the scene called Manner, while the best account for the Korean child is had by conditioning the probability on Path. In this context “best account” means, “maximizes discriminability.” Said differently, the English-speaking child will be more able to discriminate verbs by attending to Manner, while the Korean kid will prefer to attend to Path. If this happens often enough, then Manner will become an important element of s for English speakers and Path will serve the same purpose for Koreans.

If this account is correct, then it will appear as though the child has rules for parsing a scene into situation descriptions **s**, and these rules are related to the child’s native language. The rules are illusory, however. Students of each language simply search for those elements of the semantic core that best explain why words are used in particular scenes. Recent work is suggestive that certain motion cues and intention-based actions predict where a scene may be parsed (Baldwin et al., 2001; Zacks, 2004), but says nothing about the role of language.

Evidence that these linguistic elements are accessed in motion events has recently been studied in young children and infants. Golinkoff et al. (2002) used point-light displays to test for sensitivity to path and manner with an intermodal preferential looking paradigm. Three-year-olds were able to match a motion, stripped of any identifying information other than path and manner, with the target-verb spoken by an experimenter. A follow-up experiment indicated young children could also produce appropriate (action) verbs when prompted, using only point-light displays. The authors concluded that point-light displays are (and will be in future research) useful for detecting the components most useful to verb learning. Yet before one can learn a verb that encodes manner or path, it is conceivable that the infant should attend to such components in an event. Casasola, Hohenstein, & Naigles (2003), Zheng & Goldin-Meadow (2002), and Pulverman et al. (2003) have provided preliminary evidence that manner and path are attended to even with little to no previous exposure to language models.

The manipulation of parts of a motion event, involving an interaction between two objects such as using a Michottian-like manipulation with varied velocities and/or delays, in relation to verb usage and word choice has not been studied to-date. While the original elements described by Talmy to comprise a motion event, such as Path and Manner, should be addressed, they may only be determinants of verb meaning for 'simple' motion events (i.e. events involving only one agent, not involving an interaction with some recipient). More components may be involved in whole-body interactions that should not be overlooked. In P. Cohen's Maps for Verbs (1998) framework, elements such as velocity and energy transfer serve as candidates for other elements accessible in the semantic core.

Maps for Verbs

We tested the hypothesis that word choices are conditioned on patterns of motion in a study called "Maps for Verbs." We began with a dynamical representation of verbs that denote physical interactions between two agents or objects named A and B. Examples include bump, hit, push, overtake, chase, follow, harass, hammer, shove, meet, touch, propel, kick, bounce, and so on (P. Cohen, 1998).

The maps for verbs framework proposes that simple interactions between whole bodies can be characterized by the physical dynamics of the interaction. According to the framework, whole-body interactions are naturally divided into three phases: before, during and after contact.

Figure 1 depicts these three phases. A given interaction is then described as a trajectory through these phases. Maps enable identification of characteristic patterns present in the dynamics of classes of interactions.

Insert Figure 1 about here

P. Cohen (1998) proposes that the *before* and *after* phases should plot relative velocity against the distance between the two bodies. Relative velocity is the difference between the velocity of one body, A, and another, B. Many verbs (e.g., transitive verbs) predicate one body as the “actor” and the other as the “target” (or “subject” or “recipient”) of the action. For example, in a scenario involving a PUSH, the actor is the one doing the pushing, and the target is the body being pushed. By convention, the actor is designated as body A and the target is body B. Thus, when relative velocity is positive, the actor's velocity is greater than that of the target; and when relative velocity is negative, the target's velocity is greater than that of the actor. Distance, in turn, is the measure of the distance between the bodies.

The *during* phase plots perceived energy-transfer (from the actor to the target), against time or distance. If energy transfer is positive, then the actor is imparting to the target more energy than the target originally had; if energy transfer is negative, then the situation is reverse: the target is imparting more energy to the actor. To measure perceived energy transfer, we used the simplification of calculating the acceleration of the actor in the direction of the target while in contact.

Figure 1 depicts a set of labeled trajectories that characterize the component phases of seven interaction types as described by the verbs push, shove, hit, harass, bounce, counter-shove and chase. Using these labels, an interaction can be described as a triple of trajectory labels, indicating the before, during and after characteristic trajectories. For example, **[b,b,b]** describes a *shove*: The actor approaches the target at a greater velocity than the target, closing the distance between the two bodies. As it nears the target, the actor slows, decreasing its velocity to match that of the target. Trajectory **b** of the *before* phase in Figure 1 illustrates these dynamics. At contact, the relative velocity is near or equal to zero. During the contact phase, the actor rapidly imparts more energy to the target in a short amount of time, as illustrated by **b** of the *during* phase. And after breaking-off contact with the target, the agent rapidly decreases its velocity while the target moves at a greater velocity from the energy imparted it (trajectory **b** in the *after* phase).

Following this scheme, the remaining six interaction types are characterized by the following triples:

Push: b, a, a -- Begins like shove, but at contact relative velocity is near or equal to zero and the actor smoothly imparts more energy to the target; after breaking contact, the agent gradually decreases its velocity.

Hit: c or **d, c, c** -- May begin with the actor already at high velocity relative to the target or increasing in relative velocity, and thus is characterized by **c** or **d** in the before phase.

Harass: c or **d, c, d** -- Similar to a hit, except the after-phase involves the actor quickly recovering its speed and moving back toward the target, not allowing the distance between the two to get very large. Harass highlights that all interactions are not to be viewed only as single movement to contact, but may involve many such movements to contact, one after another, and may even switch between different kinds of contact interactions.

Bounce: c or **d, d, e** -- Along with counter-shove, bounce involves the target making a more reactive response to the actor's actions. Bounce begins like a hit or harass, but at contact, the target transfers a large amount of energy back to the actor.

Counter-shove: b or **c** or **d, e, e** -- A version of a shove where the target imparts energy to the actor.

Chase: a, -, - -- The agent moves toward the target, closing the distance between the two, but never quite making contact, so the during and after phases are not relevant. This is depicted as the circular trajectory **a** in the *before* phase.

Morrison, Cannon, & Cohen (2004) used these seven classes of interaction as the basis for a study in which we looked at the frequency of verb usage of adults asked to describe the interaction types after observing them. Forty-four undergraduates (M = 20.5 years old) at the University of Massachusetts participated in this study. We used *breve* 1.4, an environment for developing realistic multi-body simulations in a three dimensional world with physics (Klein, 2002), to implement a model of the seven interaction classes described in the previous section. The model is rendered as two generic objects (a blue ball for the actor and a red ball for the target) moving on a white background.

We generated a set of movies based on the rendered interactions. For several of the interaction classes we also varied the behavior of the target object, as follows: the target object, (a) did not move except when contacted (“stationary”), (b) moved independently in a random walk (“wander”), or (c) moved according to billiard ball ballistic physics, based on the force of the collision (“coast”). We generated a total of 17 unique movies. These were presented on a G3 iMac with 14 inch screen.

A total of 18 movies were presented to each participant, with “chase” being viewed twice. After watching a movie, participants were asked to write down answers to questions on a sheet of paper given to them by the experimenter. The questions were the same for every movie:

1. What are the balls doing in this movie? (Give your overall impression of what was happening between them, the ‘gist’)
2. What is the red ball doing?
3. What is the blue ball doing?
4. Can you think of any words to describe the tone or the mood of the movie? (e.g., the balls are friendly/ not friendly)

The experimenter encouraged participants to write as much as they could to describe the movies. All the action words and other content words for each trial were extracted and “canonicalized,” converting verbs in different tenses or forms (e.g., ending in -ed, -ing, etc.) to a unique form. Also, negation phrases, such as “it's not zooming” or “red didn't move,” were also transformed into a single token, e.g., not-zooming and not-moving.

After canonicalization, we kept only the verbs from the content words (a total of 155 verbs). The following 65 verbs are those that were each used by ten or more subjects to describe the movies: advancing, annoying, approaching, attaching, attacking, avoiding, backing, beating, bouncing, bullying, bumping, catching, charging, chasing, circling, coming, controlling, defending, dominating, escaping, fighting, floating, following, forcing, getting, giving, guiding, helping, hitting, kissing, knocking, leading, leaving, letting, looking, losing, nudging, pursuing, placing, playing, propelling, pushing, repeating, repelling, resisting, responding, rolling, running, shoving, slamming, slowing, sneaking, standing, standing-ones-ground, staying, stopping, striking, tagging, teasing, touching, traveling, trying, waiting, wanting, winning.

Recall that the maps-for-verbs framework hypothesizes that a representation based on the dynamics of before, during and after interactions are a foundation for the semantics of verbs describing physical interactions between objects. If this hypothesis is correct, we would expect the subjects in the preceding experiment to use particular verbs when describing the movies they observed. Furthermore, movies that share the same kind of dynamics in terms of before, during and after phases of interaction should elicit similar groups of verbs. To see whether this was the case, we clustered the 17 movies according the frequency of word usage, where frequency was according to the number of different subjects who used a given word to describe a movie (that is, if five different subjects used the word "approaching" to describe the harass-wander movie, then the frequency recorded was 5). We used hierarchical agglomerative clustering (Duda, Hart & Stork, 2001) to cluster the movies based on these word frequencies. Figure 2 shows the generated dendrogram tree depicting the results of clustering (ignore for the moment the additional labels and notation to the right).

Insert Figure 2 here

At first the dendrogram looks disappointing; while there is some structure, it is not clear how to interpret the groupings. However, recall that the movies were generated by behavioral programs, written in *breve*, that attempt to match the dynamics outlined in Figure 1. The program specifications do not guarantee that the salient perceptual features of before, during and after interaction dynamics will be perspicuous.

To explore this further, we independently observed each movie and chose what we believed to be features that help distinguish movies from one-another. We came up with a total of five very simple features: whether red (the target of the interaction) looked purposeful before or after contact (*purpose-before, purpose-after*) – “purposeful” was in terms of whether red appeared to change its heading on its own; whether red seemed to react to contact (*reactive-during*) – “react” was in terms of whether red appeared to change its behavior based on blue's contact; and whether the initial or final stages of the contact appeared gentle (*gentle-start, gentle-end*).

We then went through each movie and assigned a minus or plus, depending on whether each feature was present (“-“ = no; “+” = yes). Some cases were uncertain, so we assigned a “+?” or “-?”; and some cases were indeterminable, receiving a “?.” We have placed these feature vectors next to the corresponding leaves of the dendrogram in Figure 2. We can now see that there is significant structure to the clusters, based on the similar features that are grouped. The internal node labeled 1 in the dendrogram tree of Figure 2 distinguishes between the cluster of movies where red is not reactive to blue's contact while the contact begins gently from movies in which red is reactive and contact does not begin gently. The node labeled 2 in the dendrogram distinguishes between whether red looks purposeful before or after interaction (although the placement of harass-wander is problematic; it should be associated with hit-wander and bounce-wander). Finally, the node labeled 3 appears to separate groups of movies that involve gentle starts to interactions or red reactivity from movies that all involve abrupt starts and ends to the contact phase of interaction (except for bounce-wander).

These results indicate that the dynamical features present in the movies influence the choice of verbs used by the subjects to describe the movies. Although, to date, we have only tested a subset of the possible interaction types outlined in Figure 1, the data thus far seem to indicate that the distinctions in the maps for verbs framework, which led us to develop 17 distinct movies, do in fact influence word choices people make. We have demonstrated that words are selected preferentially in response to different dynamics, but we have not demonstrated that the distinctions in the maps for verbs framework (i.e., the different paths through the three phases) are systematically associated with different distributions of evoked words. Word use certainly seems to be associated with dynamics, but not necessarily exclusive to the ways described by the maps for verbs framework. More work is needed to show that this framework *predicts* distributions of word use for different movies. Preliminary work with preschool aged children is suggestive that even fairly new English language learners, word choice is associated with these motion components of an interaction. While there may be other elements and parameters contributing to **s**, this study suggests that we have a good starting place to begin looking seriously at the detail of dynamics involved, their development, and also the possibility of additional elements involved in giving these whole-body interactions meaning.

Future directions

There are two avenues of research within the existing Maps for Verbs framework which could make considerable advancements to our understanding of motion-based semantics and word choice. Given the evidence discussed throughout in this paper, both cross-cultural and developmental work in this area is warranted.

If we were to test a Korean population with manipulations set out in the Maps for Verbs framework, would we see the same distributions of verbs for the movies? At this point we can only speculate. Not only have differences been found in verb usage between Korean and English speakers (e.g., Choi & Bowerman, 1991), but also differences in spatial categorization (Choi et al., 1999), and (potentially) universal early sensitivities to these distinctions may disappear if the language does not lexicalize them (e.g., McDonough, Choi, & Mandler, 2003). It would be interesting to see along which parameters the Korean population categorizes whole-body interactions in comparison to an English-speaking population. Furthermore, we know nothing, at this point, about cross-cultural emphases on different phases of an interaction. It is plausible that dynamics within some phases of an interaction dictate word choice more than others. And maybe these phase differences vary across cultures. In other words, perhaps the sensitivity of one language is focused on the before phase- that the behavior of an agent, just prior to contact with another, has more influence over the semantics, and therefore word choice, than whatever happens in the during or after contact phases. In another language, events within the "during contact" phase might be most informative. Comparing the phases would not only be informative in discovering something more about cross-cultural ontological distinctions (and similarities) but it also may suggest other contributing elements to **s**, present in the semantic core.

As we have also discussed in this chapter, infants are remarkably capable of extracting meaning from motion dynamics. The work described earlier in this chapter on the perception of physical causality in infancy suggests infants may make categorical distinctions along the dimensions of the Maps for Verbs framework within the first year of life. Perhaps, as they learn the interaction categories most relevant to the language being learned, we will see a loss of some distinctions and the refinement of others. Perhaps infants' early sensitivities to motion dynamics also contribute new elements to the semantic core from which **s** is formed.

Concluding Remarks

We began this chapter with the question of how we can effectively communicate through language in an ever-changing world. We suggested that, in a world that is constantly in motion, movement must be a powerful cue for extracting meaningful information from our environment. A general theory of word meaning was offered, stating that, in all language, words uttered are conditioned on the representation of a situation, which is largely made up of these situational motion elements we perceive. We reviewed the literature that even unobservable elements in **s** can be inferred through motion. Moreover, we provided a review of cognitive and linguistic evidence suggesting that infants are initially sensitive to far more motion cues than what are later represented in **s**, and whether or not the sensitivity remains will depend on how the perceptual system and spoken language bundles these elements. The context provides meaning, and while we claim motion as a central component, we have never claimed it as the sole contributor. We have reviewed several proposed representational semantic frameworks for investigating motion elements and discussed one in detail, Maps for Verbs. However, there may be other additional motion elements not presented here, that have not yet been discovered.

We interact with objects in the world from birth, so it seems fitting to study the dynamics of interactions when making claims about semantics and language development. But other potential contributors to **s** should also be examined, such as syntax, other words uttered, intention, number, etc. While these domains are studied extensively on their own, a comprehensive associative learning theory would have to consider the influences of all of the elements that may contribute to **s** in order to have a complete model for the acquisition of word meaning.

References

Arterberry, M. E., & Bornstein, M. H. (2002). Infant perceptual and conceptual categorization: The roles of static and dynamic attributes. Cognition, 86, 1-24.

Arterberry, M. E., & Bornstein, M. H. (2001). Three-month-old infants' categorization of animals and vehicles based on static and dynamic attributes, Journal of Experimental Child Psychology, 80, 333-346.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic actions. Child Development, 72, 708-718.

Baron-Cohen, S. (1994). How to build a baby that can read minds: Cognitive mechanisms in mind-reading. Cahiers de Psychologie Cognitive, 13, 513-552.

Barsalou, L. W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22, 577–609.

Bartsch, K., & Wellman, H. M. (1995). Children Talk about the Mind. New York: Oxford University Press.

Bertenthal, B. I. (1993). Infants’ perception of biomechanical motions: Intrinsic image and knowledge-based constraints. In C. E. Granrud (Ed.), Visual Perception and Cognition in Infancy, (175-214). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motions. Psychological Science, 5, 221-225.

Bertenthal, B. I., Proffitt, D. R., & Cutting, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motions. Journal of Experimental Child Psychology, 37, 213-230.

Berry, D. S., & Springer, K. (1993). Structure, motion, and preschoolers perception of social causality. Ecological Psychology, 5, 273-283.

Casasola, M., Hohenstein, J., & Naigles, L. R. (2003, April). Ten-month-old infants' discrimination of manner and path in motion events. In M. Casasola (Chair), From infancy to adulthood: Exploring the effect of linguistic input on the discrimination of manner and path in motion events. Symposium conducted at the Society for Research in Child Development Biennial Meeting, Tampa, FL.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. Cognition, 41, 83-121.

Choi, S., McDonough, L., Bowerman, M., & Mandler, J. M. (1999). Early sensitivity to language-specific spatial categories in English and Korean. Cognitive Development, 14, 241-268).

Cohen, L. B., & Amsel, G. (1998). Precursors to infants' perception of the causality of a simple event. Infant Behavior and Development, 21, 713-732.

Cohen, L. B., & Oakes, L. M. (1993). How infants perceive a simple causal event. Developmental Psychology, 29, 421-433.

Cohen, L. B., Rundell, L. J., Spellman, B. A., & Cashon, C. H. (1999). Infants' perception of causal chains. Psychological Science, 10, 412-418.

Cohen, P. R. (1998). Maps for Verbs. In Proceedings of the Information and Technology Systems Conference, (pp. 21–33). Fifteenth IFIP World Computer Congress, Vienna.

Cohen, P. R., Oates, T., Beal, C. R., & Adams, N. (2002). Contentful mental states for RobotBaby. Paper presented at the meeting of the American Association for Artificial Intelligence.

Csibra, G., Gergely, G., Biro, S., Koos, O., & Brockbank, M. (1999). Goal-attribution without agency cues: the perception of 'pure reason' in infancy. Cognition, 72, 237-267.

Duda, R. O.; Hart, P. E.; & Stork, D. G. (2001). Pattern Classification. New York, NY: Wiley.

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. Annual Review of Psychology, 50, 21-45.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S.A. Kuczaj, II (ed.), Language development Vol. 2: Language, thought, and culture (pp. 301-334). Hillside, NJ: Lawrence Erlbaum Associates.

Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. Child Development, 49, 988-998.

Golinkoff, R. M., & Kerr, J. L. (1978). Infants' perception of semantically defined action role changes in filmed events. Merrill Palmer Quarterly, 24, 53-61.

Golinkoff, R. M., Chung, H. L., Hirsh-Pasek, K., Liu, J., Bertenthal, B. L., Brand, R., Maguire, M. J., & Hennon, E. (2002). Young children can extend motion verbs to point light displays. Developmental Psychology, 38, 604-614.

Gopnik, A., & Choi, S. (1995). Names, relational words, and cognitive development in English and Korean speakers: Nouns are not always learned before verbs. In M. Tomasello & W.E. Merriman (Eds.), Beyond names for things: Young children's acquisition of verbs (pp. 63-80), Hillsdale, NJ: Lawrence Erlbaum Associates.

Guylai, E. (2000). Attribution of meaning from movement of simple objects. Perceptual and Motor Skills, 90, 27-35.

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

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. Child Development, 74, 1368-1378.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. Perception and Psychophysics, 14, 201-211.

Johnson, M. (1987). The Body in the Mind. Chicago, IL: University of Chicago Press.

Klein, J. (2002). Breve: a 3D simulation environment for the simulation of decentralized systems and artificial life. In Proceedings of Artificial Life VIII, the 8th International Conference on the Simulation and Synthesis of Living Systems. http://www.spiderland.org/breve/

Lakoff, G. (1987). Women, fire, and dangerous things: What categories reveal about the mind. Chicago, IL: University of Chicago Press.

Legerstee, M., Barna, J., DiAdamo, C. (2000). Precursors to the development of intention at 6 months: Understanding people and their actions. Developmental Psychology, 36, 627-634.

Leslie, A. M. (1994). ToMM, ToBY, and Agency: Core architecture and domain specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), Mapping the mind: Domain specificity in cognition and culture, (pp. 119-148). New York, NY: Cambridge University Press.

Leslie, A. M. (1988). The necessity of illusion: Perception and thought in infancy. In L. Weiskrantz (Ed.), Thought without language (pp. 185-210). Oxford, England: Clarendon.

Leslie, A. M. (1984). Spatiotemporal continuity and the perception of causality in infants. Perception, 13, 287-305.

Leslie, A. M. (1982). Perception of causality in infants. Perception, 11, 173-186.

Leslie, A. M. & Keeble, S. (1987). Do six-month-olds perceive causality? Cognition, 25, 265-288.

Mandler, J. (2000). Perceptual and conceptual processes in infancy. Journal of

Cognition and Development, 1, 3-36.

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

Mandler, J. M., & McDonough, L. (1996). Drinking and driving don't mix: Inductive generalization in infancy. Cognition, 59, 307-335.

McDonough, L., Choi, S., & Mandler, J. (2003). Understanding spatial relations: Flexible infants, lexical adults. Cognitive Psychology, 46, 229-259.

Michotte, A. (1963). The perception of causality. London: Methuen.

Morrison, C. T., Cannon, E. N., & Cohen, P. R. (2004). When Push Comes to Shove: A Study of the Relation Between Interaction Dynamics and Verb Use. In Working Notes of AAAI Spring Symposium Workshop: Language Learning: An Interdisciplinary Perspective. Menlo Park, CA: AAAI.

Naigles, L. R., Eisenberg, A. R., Kako, E. T., Highter, M., & McGraw, N. (1998). Speaking of motion: verb use in English and Spanish. Language and Cognitive Processes, 13, 521-549.

Oakes, L. M. (1994). Development of infants' use of continuity cues in their perception of causality. Developmental Psychology, 30, 869-879.

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. Cognitive Development, 5, 193-207.

Oates, T. (2001). Grounding Knowledge in Sensors. Ph.D. Dissertation, University of Massachusetts, Amherst.

Premack, D. (1990). The infant's theory of self-propelled objects. Cognition, 36, 1-16.

Pulverman, R., Sootsman, J. L., Golinkoff, R. M., & Hirsh-Pasek, (2003a, April).

Attention to manner and path in nonlinguistic event processing: English-speaking infants learn to mind their manners. In M. Casasola (Chair), From infancy to adulthood: Exploring the effect of linguistic input on the discrimination of manner and path in motion events. Symposium conducted at the Society for Research in Child Development Biennial Meeting, Tampa, FL.

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate-inanimate distinction. Psychological Bulletin, 127, 209-228.

Slobin, D.I. (1973). Cognitive prerequisites for the development of grammar. In C.A. Ferguson & D.I. Slobin (Eds.), Studies in child language development (pp. 175–208). New York: Holt, Rinehart & Winston.

Spelke, E. S., Phillips, A. T., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action. In D. Sperber, D. Premack, & A. Premack (Eds.), Causal Cognition: A multidisciplinary debate (pp. 44-78). New York: Oxford University Press.

Talmy, L. (2000). Toward a Cognitive Semantics: Vol I. Cambridge, MA: MIT Press.

Talmy, L. (1988). Force dynamics in language and cognition. Cognitive Science, 12, 49-100.

Talmy, L. (1975). Semantics and syntax of motion. In J. Kimball (Ed.), Syntax and semantics. New York: Academic Press.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. Developmental Psychology, 32, 492-504.

Tardif, T., Shatz, M. & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. Journal of Child Language, 24, 535-565.

Tomasello, M. (1995). Pragmatic contexts for early verb learning, In M. Tomasello & W.E. Merriman (Eds.), Beyond names for things: Young children's acquisition of verbs (pp. 115-146), Hillsdale, NJ: Lawrence Erlbaum Associates.

Tomasello, M. (1992). First verbs: A case study of early grammatical development. Cambridge, UK: Cambridge University Press.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. Cognition, 69, 1-34.

Zacks, J. M. (2004). Using movement and intention to explain simple events. Cognitive Science, 28, 979-1008.

Zheng, M., & Goldin-Meadow, S. (2002). Thought before language: how deaf and hearing children express motion events across cultures. Cognition, 85, 145-175.

Figure Captions.

1. Maps for Verbs model of the three phases of interaction

2. Dendrogram representing clustering of movies based on word usage frequencies, where word usage is based on the number of different subjects who used a given word. The complete set of 155 verbs were used to characterize word usage. The labels inside the leaves of the dendrogram correspond to movie names; the numbers are unique identifiers assigned by the clustering procedure and should be ignored

Figure 1.

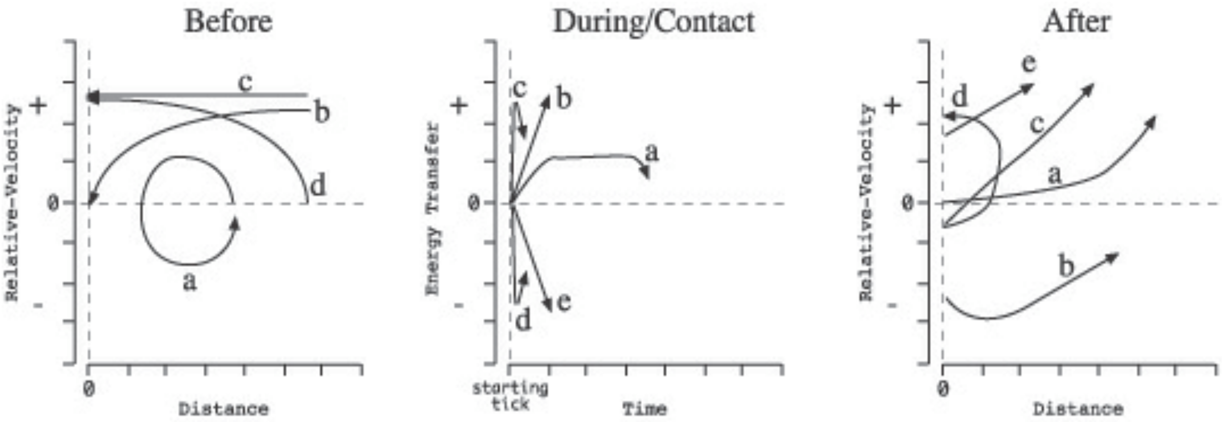
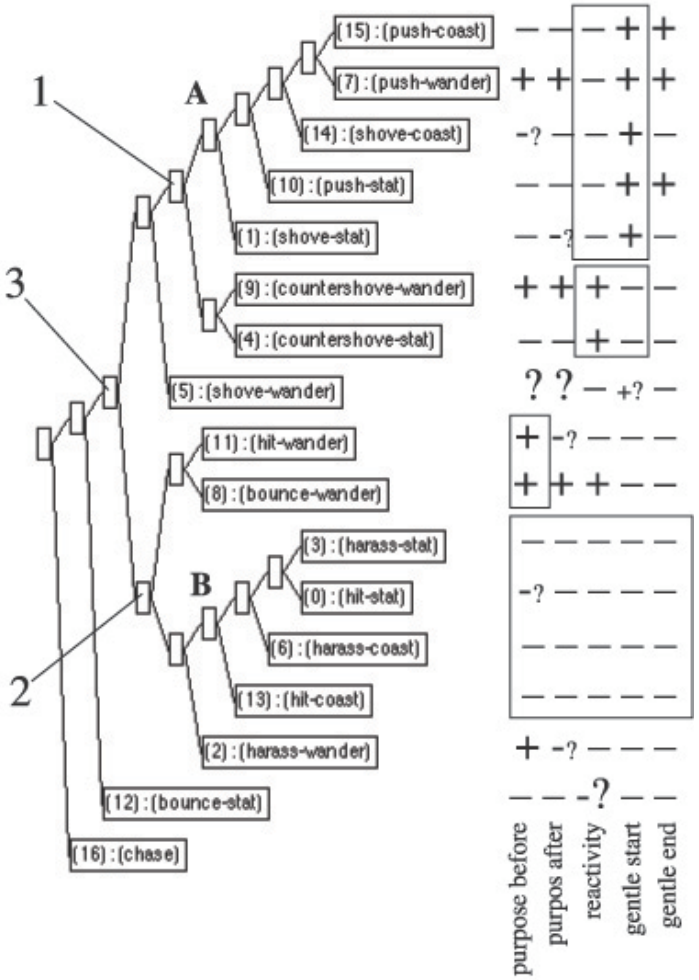

Before
During/Contact
After
Relative-Velocity
Energy Transfer
Relative-Velocity
+
0
-
0
Distance
starting tick
Time
0
Distance
a
b
c
d
a
b
c
d
e
a
b
c
d
e

Figure 2.


1
2
3
A
B
(15) : (push-coast)
(7) : (push-wander)
(14) : (shove-coast)
(10) : (push-stat)
(1) : (shove-stat)
(9) : (countershove-wander)
(4) : (countershove-stat)
(5) : (shove-wander)
(11) : (hit-wander)
(8) : (bounce-wander)
(3) : (harass-stat)
(0) : (hit-stat)
(6) : (harass-coast)
(13) : (hit-coast)
(2) : (harass-wander)
(12) : (bounce-stat)
(16) : (chase)
– – – + +
+ + – + +
-? – – + –
– – – + +
– -? – + –
+ + + – –
– – + – –
? ? – +? –
+ -? – – –
+ + + – –
– – – – –
-? – – – –
– – – – –
– – – – –
+ -? – – –
– – -? – –
purpose before
purpos after
reactivity
gentle start
gentle end