12

Connecting Concepts to Each Other and the World

Robert L. Goldstone, Ying Feng, and Brian J. Rogosky

Consider two individuals, John and Mary, who each possess a number of concepts. How can we determine that John and Mary both have a concept of, say, **Horse**? John and Mary may not have exactly the same knowledge of horses, but it is important to be able to place their horse concepts into correspondence with one another, if only so that we can say things like, "Mary's concept of **horse** is much more sophisticated than John's." Concepts should be public in the sense that they can be possessed by more than one person (Fodor, 1998; Fodor & Lepore, 1992), and for this to be the possible, we must be able to determine correspondences, or translations, between two individuals' concepts.

There have been two major approaches in cognitive science to conceptual meaning that could potentially provide a solution to finding translations between conceptual systems. According to an "external grounding" account, concepts' meanings depend on their connection to the external world (this account is more thoroughly defined in the next section). By this account, the concept **Horse** means what it does because our perceptual apparatus can identify features that characterize horses. According to what we will call a "Conceptual web" account, concepts' meanings depend on their connections to each other. By this account, **Horse'**s meaning depends on **Gallop**, **Domesticated**, and **Quadruped**, and in turn, these concepts depend on other concepts, including **Horse** (Quine & Ullian, 1970).

In this chapter, we will first present a brief tour of some of the main proponents of conceptual web and external grounding accounts of conceptual meaning. Then, we will describe a computer algorithm that translates between conceptual systems. The initial goal of this computational work is to show how translating across systems is possible using only within-system relations, as is predicted by a conceptual web account. However, the subsequent goal is to show how the synthesis of external and internal information can dramatically improve translation. This work suggests that the external grounding and conceptual web accounts should not be viewed as competitors, but rather, that these two sources of information strengthen one another. In the final section of the chapter, we will present applications of the developed ABSURDIST algorithm to object recognition, large corpora translation, analogical reasoning, and statistical scaling.

In this chapter, we will be primarily interested in translating between conceptual systems, but many of our examples of translation will involve words. Concepts are not always equivalent to word meanings. For one thing, we can have concepts of things for which we do not have words, such as the cylindrical plastic sheath at the tips of shoelaces. Although there may not be a word for every concept we possess, behind every word there is a conceptual structure. Accordingly, when we talk about a concept of **Horse**, we are referring to the conceptual structure that supports people's use of the word "Horse" as well as their ability to recognize horses, predict the behavior of horses, and interact appropriately with horses.

GROUNDED CONCEPTS

For a concept to be externally grounded means that, in one way or another, its meaning is based on its connection to the world. There are several ways for this to occur. First, aspects of the concept may come to us via a perceptual system. Our concept of **Red**, **Fast**, and **Loud** all have clear perceptual components, but most, if not all (Barsalou, 1999), other concepts do as well. Second, a concept may be tied to objects in the world by deixis – by linguistically or physically pointing in a context. When a parent teaches a child the concept **Horse** by pointing out examples, this provides contextualized grounding for the child's emerging concept. A final third way, which we will not be addressing in our work, is that meaning may be tied directly to the external world without being mediated through the senses. Putnam's (1973) famous "twin earth" thought experiment is designed to show how the same internal, mental content can be associated with two different external referents. Putnam has us imagine a world, twin earth, that is exactly like our earth except that the compound we call water (H_2O) has a different atomic structure (xyz), while still looking, feeling, and acting like water as we on real earth know it. Two molecule-for-molecule identical individuals, one on earth and one on twin earth, would presumably have the same internal mental state when thinking "water is wet," and yet, Putnam argues, they *mean* something different. One means stuff that is actually, whether they know it or not, made up of H_2O, while the other means stuff that is made up of xyz. Putnam concludes that what is meant by a term is not determined solely by mental states, but rather depends upon the external world as well.

The rest of the chapters in this book give excellent grounds for believing that our concepts are not amodal and abstract symbolic representations, but rather are grounded in the external world via our perceptual systems. Lawrence Barsalou has presented a particularly influential and well-developed version of this account in the form of Perceptual Symbols Theory (Barsalou, 1999). By this account, conceptual knowledge involves activating brain areas dedicated for perceptual processing. When a concept is brought to mind, sensory-motor areas are reactivated to implement perceptual symbols. Even abstract concepts, such as truth and negation, are grounded in complex perceptual simulations of combined physical and introspective events. Several lines of empirical evidence are consistent with a perceptually grounded conceptual system. Detailed perceptual information is represented in concepts and this information is used when reasoning about those concepts (Barsalou et al., 2003). Concepts that are similar to one another give rise to similar patterns of brain activity, and a considerable amount of this activity is found in regions associated with perceptual processing (Simmons & Barsalou, 2003). When words are heard or seen, they spontaneously give rise to eye movements and perceptual images that would normally be evoked by the physical event designated by the words (Richardson, Spivey, Barsalou, & McRae, 2003; Stanfield & Zwaan, 2001). Switching from one modality to another during perceptual processing incurs a processing cost. The same cost is exhibited during the verification of concept properties, consistent with the notion that perceptual simulation underlies even verbal conceptual processing (Pecher, Zeelenberg, & Barsalou, 2003).

Much of the recent work on perceptual and embodied accounts of concepts has involved verbal stimuli such as words, sentences, and stories. The success of grounded accounts of language is noteworthy and surprising because of its opposition to the standard conception of language as purely arbitrary and symbolic. An acknowledgment of the perceptual grounding of language has lead to empirically validated computational models of language (Regier, 1996; Regier & Carlson, 2001). It has also provided insightful accounts of metaphors for understanding abstract notions such as time (Boroditsky, 2000; Boroditsky & Ramscar, 2002) and mathematics (Lakoff & Nunez, 2000). There has been a recent torrent of empirical results that are inconsistent with the idea that language comprehension is based on concepts that are symbols connected only to each other. Instead, the data support an embodied theory of meaning that relates the meaning of sentences to human perception and action (Glenberg & Kaschak, 2002; Glenberg & Robertson, 2000; Zwaan, Stanfield, & Yaxley, 2002).

Consistent with Barsalou's Perceptual Symbols Theory, other research has tried to unify the typically disconnected literatures on low-level perceptual processing and high-level cognition. Goldstone and Barsalou (1998) argue for strong parallels between processes traditionally considered to be perceptual on the one hand and conceptual on the other, and that perceptual processes are co-opted by abstract conceptual thought. Other research indicates bidirectional influences between our concepts and perceptions

(Goldstone, 2003; Goldstone, Lippa, & Shiffrin, 2001; Schyns, Goldstone, & Thibaut, 1998). Like a good pair of Birkenstock sandals that provide support by flexibly conforming to the foot, perception supports our concepts by conforming to these concepts. Perceptual learning results in perceptual and conceptual systems that are highly related. Taken together, this work suggests that apparently high-level conceptual knowledge and low-level perception may be more closely related than traditionally thought/perceived.

The case for grounding conceptual understanding in perception has a long philosophical history. As part of the British empiricist movement, David Hume (1740/1973) argued that our conceptual ideas originate in recombinations of sensory impressions. John Locke (1690) believed that our concepts ("ideas") have their origin either by our sense organs or by an internal sense of reflection. He argued further that our original ideas are derived from sensations (e.g., yellow, white, heat, cold, soft, and hard), and that the remaining ideas are derived from or depend upon these original ideas. The philosophical empiricist movement has been reinvigorated by Jesse Prinz (2002), who argues for sensory information as the ultimate ground for our concepts and beliefs. Stevan Harnad has similarly argued that concepts must be somehow connected to the external world, and this external connection establishes at least part of the meaning of the concept. In his article "The symbol grounding problem," Stevan Harnad (1990) considers the following thought experiment: "Suppose you had to learn Chinese as a *first* language and the only source of information you had was a Chinese/Chinese dictionary. [. . .]. How can you ever get off the symbol/ symbol merry-go-round? How is symbol meaning to be grounded in something other than just more meaningless symbols? This is the symbol grounding problem" (pp. 339–340).

CONCEPTUAL WEBS

In stark antithesis to Harnad's thought experiment, artificial intelligence researchers have argued that conceptual meaning *can* come from dense patterns of relations between symbols even if the symbols have no causal connections to the external world. Lenat and Feigenbaum (1991) claim that "The problem of 'genuine semantics' . . . gets easier, not harder, as the K[nowledge] B[ase] grows. In the case of an enormous KB, such as CYC's, for example, we could rename all of the frames and predicates as G001, G002, . . . , and – using our knowledge of the world – reconstruct what each of their names must be" (p. 236). This claim is in direct opposition to Harnad's image of the symbol-symbol merry-go-round, and may seem ungrounded in several senses of the term. Still, depending on the power of intrasystem relations has been a mainstay of artificial intelligence, linguistics, and psychology for decades.

In semantic networks, concepts are represented by nodes in a network, and gain their functionality by their links to other concept nodes (Collins & Loftus, 1975; Quillian, 1967). Often times, these links are labeled, in which case different links refer to different kinds of relations between nodes. **Dog** would be connected to **Animal** by an **Is-a** link, to **Bone** by an **Eats** link, and to **Paw** by a **Has-a** link. These networks assume a conceptual web account of meaning because the networks' nodes are typically only connected to each other, rather than to an external world or perceptual systems.

A computational approach to word meaning that has received considerable recent attention has been to base word meanings solely on the patterns of co-occurrence between a large number of words in an extremely large text corpus (Burgess, Livesay, & Lund, 1998; Burgess & Lund, 2000; Landauer & Dumais, 1997). Mathematical techniques are used to create vector encodings of words that efficiently capture their co-occurrences. If two words, such as "cocoon" and "butterfly" frequently co-occur in an encyclopedia or enter into similar patterns of co-occurrence with other words, then their vector representations will be highly similar. The meaning of a word, its vector in a high dimensional space, is completely based on the contextual similarity of words to other words.

The traditional notion of concepts in linguistic theories is based upon conceptual webs. Ferdinand de Saussure (1915/1959) argued that all concepts are completely "negatively defined," that is, defined solely in terms of other concepts. He contended that "language is a system of interdependent terms in which the value of each term results solely from the simultaneous presence of the others" (p. 114) and that "concepts are purely differential and defined not in terms of their positive content but negatively by their relations with other terms in the system" (p. 117). By this account, the meaning of **Mutton** is defined in terms of other neighboring concepts. **Mutton**'s use does not extend to cover sheep that are living because there is another lexicalized concept to cover living sheep (**Sheep**), and **Mutton** does not extend to cover cooked pig because of the presence of **Pork**. Under this notion of interrelated concepts, concepts compete for the right to control particular regions of a conceptual space (see also Goldstone, 1996; Goldstone, Steyvers, & Rogosky, 2003). If the word **Mutton** did not exist, then "all its content would go to its competitors" (Saussure, 1915/1959, p. 116).

According to the conceptual role semantics theory in philosophy, the meaning of a concept is given by its role within its containing system (Block, 1986, 1999; Field, 1977; Rapaport, 2002). A conceptual belief, for example, that dogs bark, is identified by its unique causal role in the mental economy of the organism in which it is contained. A system containing only a single concept is not possible (Stich, 1983). A common inference from this view is that concepts that belong to substantially different systems must have different meanings. This inference, called "translation holism" by Fodor and Lepore (1992), entails that a person cannot have the same concept as another person unless the rest of their conceptual systems are at least highly similar. This view has had perhaps the most impact in the philosophy of science, where Kuhn's incommensurability thesis states that there can be no translation between scientific concepts across scientists that are committed to fundamentally different ontologies (Kuhn, 1962). A chemist indoctrinated into Lavoisier's theory of oxygen cannot translate any of their concepts to earlier chemists' concept of phlogiston. A more recent chemist can only entertain the earlier phlogiston concept by absorbing the entire Pre-Lavoisier theory, not by trying to insert the single phlogiston concept into their more recent theory or by finding an equivalent concept in their theory. A concept can only be understood if an entire system of interrelated concepts is also acquired.

translating between conceptual systems

We will not directly tackle the general question of whether concepts gain their meaning from their connections to each other, or from their connection to the external world. In fact, our eventual claim will be that this is a false dichotomy, and that concepts gain their meaning from both sources. Our destination will be a synergistic integration of conceptual web and external grounding accounts of conceptual meaning. On the road to this destination, we will first argue for the sufficiency of the conceptual web account for conceptual translation. Then, we will show how the conceptual web account can be supplemented by external grounding to establish meanings more successfully than either method could by itself.

Our point of departure for exploring conceptual meaning will be a highly idealized and purposefully simplified version of a conceptual translation task. The existence of translation across different people's conceptual systems, for example between John and Mary's **Horse** concepts, has been taken as a challenge to conceptual web accounts of meaning. Fodor and Lepore (1992) have argued that if a concept's meaning depends on its role within its larger conceptual system, and if there are some differences between Mary's and John's systems, then the meanings of Mary's and John's concepts would necessarily be different. A natural way to try to salvage the conceptual web account is to argue that determining corresponding concepts across systems does not require the systems to be identical, but only similar. However, Fodor (1998) insists that the notion of similarity is not adequate to establish that Mary and John both possess a concept of **Horse**. Fodor argues that "saying what it is for concepts to have similar, but not identical contents presupposes a prior notion of beliefs with similar but not identical concepts" [p. 32]. In opposition to this, we will argue that conceptual translation can proceed using only the notion of similarity, not identity, between concepts. Furthermore, the similarities between Mary's and John's concepts can be determined using only relations between concepts *within* each person's head.

We will present a simple neural network called ABSURDIST (Aligning Between Systems Using Relations Derived Inside Systems Themselves) that finds conceptual correspondences across two systems (two people, two time slices of one person, two scientific theories, two cultures, two developmental age groups, two language communities, etc.) using only interconceptual similarities, not conceptual identities, as input. Laakso and Cottrell (1998, 2000) describe another neural network model that uses similarity relations within two systems to compare the similarity of the systems, and Larkey and Love (2003) describe a connectionist algorithm for aligning between graphs that is highly related. ABSURDIST belongs to the general class of computer algorithms that solve graph matching problems. It takes as input two systems of concepts in which every concept of a system is defined exclusively in terms of its dissimilarities to other concepts in the same system. ABSURDIST produces as output a set of correspondences indicating which concepts from System A correspond to which concepts from System B. These correspondences serve as the basis for understanding how the systems can communicate with each other without the assumption made by Fodor (1998) that the two systems have exactly the same concepts. Fodor argues that any account of concepts should explain their "publicity" – the notion that the same concept can be possessed by more than one person. Instead, we will advocate a notion of "correspondence." An account of concepts should explain how concepts possessed by different people can correspond to one another, even if the concepts do not have exactly the same content. The notion of corresponding concepts is less restrictive than the notion of identical concepts, but is still sufficient to explain how people can share a conversational ground, and how a single person's concepts can persist across time despite changes in the person's knowledge. While less restrictive than the notion of concept identity, the notion of correspondence is stronger than the notion of concept similarity. John's **Horse** concept may be similar to Mary's **Donkey** concept, but the two do not correspond because John's **Horse** concept is even more similar in terms of its role within the conceptual system. Two concepts correspond to each other if they play equivalent roles within their systems, and ABSURDIST provides a formal method for determining equivalence of roles.

A few disclaimers are in order before we describe the algorithm. First, ABSURDIST finds corresponding concepts across individuals, but does not connect these concepts to the external world. The algorithm can reveal that Mary's **Horse** concept corresponds to John's **Horse** concept, but the basic algorithm does not reveal what in the external world corresponds to these concepts. However, an interesting extension of ABSURDIST would be to find correspondences between concepts within an internal system and physically measurable elements of an external system. Still, as it stands

ABSURDIST falls significantly short of an account of conceptual meanings. The intention of the model is simply to show how one task related to conceptual meaning, finding corresponding concepts across two systems, can be solved using only within-system similarities between concepts. It is relevant to the general issue of conceptual meaning given the arguments in the literature (e.g. Fodor, 1998) that this kind of within-system similarity is insufficient to identify cross-system matching concepts.

Second, our initial intention is not to create a rich or realistic model of translation across systems. In fact, our intention is to explore the simplest, most impoverished representation of concepts and their interrelations that is possible. If such a representation suffices to determine cross-system translations, then richer representations would presumably fare even better. To this end, we will not represent concepts as structured lists of dimension values, features or attribute/value frames, and we will not consider different kinds of relations between concepts such as **Is-a**, **Has-a**, **Part-of**, **Used-for**, or **Causes**. Concepts are simply elements that are related to other concepts within their system by a single, generic similarity relation. The specific input that ABSURDIST takes will be two two-dimensional proximity matrices, one for each system. Each matrix indicates the similarity of every concept within a system to every other concept in the system. While an individual's concepts certainly relate to each other in many ways (Medin, Goldstone, and Gentner, 1993), our present point is that even if the only relation between concepts in a system were generic similarity, this would suffice to find translations of the concept in different systems. In the final section, we will describe an extension of ABSURDIST to more structured conceptual systems.

A third disclaimer is that ABSURDIST is not primarily being put forward as a model of how people actually communicate and understand one another. ABSURDIST finds correspondences between concepts across systems, and would not typically be housed in any one of the systems. Unless Mary knows all of the distances between John's concepts, then she could not apply ABSURDIST to find translations between John and Mary. If the primary interpretation of ABSUDIST is not as a computational model of a single human's cognition, then what is it? It is an algorithm that demonstrates the available information that could be used to find translations between systems. It is an example of a hitherto underrepresented class of algorithms in cognitive science – *computational ecological psychology*. The ecological movement in perception (Gibson, 1979) is concerned with identifying external properties of things in the world that are available to be picked up by people. Although it is an approach in psychology, it is just as concerned with examining physical properties as it is with minds. Similarly, ABSURDIST is concerned with the sufficiency of information that is available across systems for translating between the systems. Traditional ecological psychology proceeds by expressing mathematical relations between physical properties. However, in the present case, a computational algorithm is necessary to determine the information that is available in the observed systems. Thus, the argument will be that even systems with strictly internal relations among their parts possess the information necessary for an observer to translate between them. However, unlike a standard interpretation of claims for "direct perception," an observer using ABSURDIST would perform a time-extended computation in order to successfully recover these translations.

ABSURDIST

ABSURDIST is a constraint satisfaction neural network for translating between conceptual systems. Unlike many neural networks, it does not learn, but rather only passes activation between units. Each of the units in the network represents an hypothesis that two concepts from different systems correspond to one another. With processing, a single set of units will tend to become highly active and all other units will become completely deactivated. The set of units that eventually becomes active will typically represent a consistent translation from one system to the other.

Elements $A_{1..m}$ belong to System A, while elements $B_{1..n}$ belong to System B. $C_t(A_q, B_x)$ is the activation, at time t, of the unit that represents the correspondence between the qth element of A and the xth element of B. There will be $m{\cdot}n$ correspondence units, one for each possible pair of corresponding elements between A and B. In the current example, every element represents one concept in a system. The activation of a correspondence unit is bound between 0 and 1, with a value of 1 indicating a strong correspondence between the associated elements, and a value of 0 indicating strong evidence that the elements do not correspond. Correspondence units dynamically evolve over time by the equations:

$$C_{t+1}(A_q, B_x) = \begin{cases} C_t(A_q, B_x) + N_t(A_q, B_x)(1 - C_t(A_q, B_x)) \\ \qquad\qquad\qquad\qquad \text{if } N_t(A_q, B_x) > 0 \\ C_t(A_q, B_x) + N_t(A_q, B_x)C_t(A_q, B_x) \ \ \text{otherwise} \end{cases} \qquad (1)$$

If $N_t(A_q, B_x)$, the net input to a unit that links the qth element of A and the xth element of B, is positive, then the unit's activation will increase as a function of the net input, passed through a squashing function that limits activation to an upper bound of 1. If the net input is negative, then activations are limited by a lower bound of 0. The net input is defined as

$$N_t(A_q, B_x) = \alpha E_t(A_q, B_x) + \beta R_t(A_q, B_x) - (1 - \alpha - \beta) I_t(A_q, B_x), \qquad (2)$$

where the E term is the external similarity between A_q and B_x, R is their internal similarity, and I is the inhibition to placing A_q and B_x into correspondence that is supplied by other developing correspondence units.

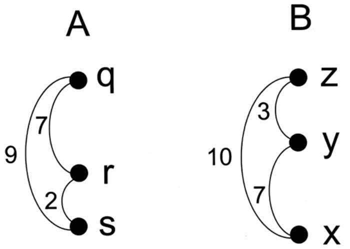

FIGURE 12.1. An example of the input to ABSURDIST. Two systems, A and B, are each represented solely in terms of the distances/dissimilarities between elements within a system. The correct output from ABSURDIST would be a cross-system translation in which element q was placed in correspondence with *x*, *r* with *y*, and *s* with *z*. Arcs are labeled with the distances between the elements connected by the arcs.

When $\alpha = 0$, then correspondences between A and B will be based solely on the similarities among the elements within a system, as proposed by a conceptual web account.

The amount of excitation to a unit based on within-domain relations is given by

$$R_t(A_q, B_x) = \frac{\sum_{\substack{r=1 \\ r\neq q}}^{m} \sum_{\substack{y=1 \\ y\neq x}}^{n} S(D(A_q, A_r)D(B_x, B_y))C_t(A_r, B_y)}{Min(m, n) - 1},$$

where $D(A_q, A_r)$ is the psychological distance between elements A_q and B_x in System A, and $S(F, G)$ is the similarity between distances F and G, and is defined as $S(F, G) = e^{-|F-G|}$. The amount of inhibition is given by

$$I_t(A_q, B_x) = \frac{\sum_{\substack{r=1 \\ r\neq q}}^{m} C_t(A_r, B_x) + \sum_{\substack{y=1 \\ y\neq x}}^{n} C_t(A_q, B_y)}{m + n - 2}.$$

These equations instantiate a fairly standard constraint satisfaction network, with one twist. According to the equation for R, Elements A_q and B_x will tend to be placed into correspondence to the extent that they enter into similar similarity relations with other elements. For example, in Figure 12.1, A_q has a distance of 7 to one element (A_r) and a distance of 9 to another element (A_s) within its System A. These are similar to the distances that B_x has to the other elements in System B, and accordingly there should be a tendency to place A_q in correspondence with B_x. Some similarity relations should count much more than others. The similarity between $D(A_q, A_r)$ and $D(B_x, B_y)$ should matter more than the similarity between $D(A_q, A_r)$ and $D(B_x, B_z)$ in terms of strengthening the correspondence between A_q and B_x, because A_r corresponds to B_y not to B_z. This is achieved by weighting the similarity between two distances by the strength of the units that align elements that are placed in correspondence by the distances. As the network begins to place A_r into correspondence with B_y, the similarity between $D(A_q, A_r)$ and $D(B_x, B_y)$ becomes emphasized as a basis for placing A_q into correspondence with B_x. As such, the equation for R represents the sum of the supporting evidence (the consistent correspondences), with each piece of support weighted by its relevance (given by the similarity term). This sum is normalized by dividing it by the minimum of $(m - 1)$ and $(n - 1)$. This minimum is the number of terms that will contribute to the R term if only 1-to-1 correspondences exist between systems.

The inhibitory Term I is based on a one-to-one mapping constraint (Falkenhainer et al., 1989; Holyoak & Thagard, 1989). The unit that places A_q into correspondence with B_x will tend to become deactivated if other strongly activated units place A_q into correspondence with other elements from B, or B_x into correspondence with other elements from A.

One problem with the original ABSURDIST algorithm described by Goldstone and Rogosky (2002) is that many iterations of activation passing between correspondence units is required before a single set of units converges. An analysis of the network dynamics often reveals that all correspondence units initially decrease their activation value, and then very gradually a set of consistent correspondence units becomes more activated. One strategy that has proven helpful in both speeding convergence in ABSURDIST and improving alignment accuracy has been to define a measure of the total amount of activation across all correspondences units,

$$T = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{t+1}(A_i, B_j).$$

Next, if T is less than the intended sum if there were a complete set of one-to-one mappings, then each correspondence unit is adjusted so that it is more active. The adjustment is the difference between the ideal sum and the actual sum of activations, weighted by the ratio of the current activation to the total activation. Hence, the boost in activation for a correspondence unit should increase as the activation of the unit relative to the total activation of the network increases. These requirements are met by the following equation for dynamically adjusting correspondence units:

$$\text{if } T < \min(m, n) \text{ then } C'_{t+1}(A_q, B_x) = C_{t+1}(A_q, B_x) + \frac{C_{t+1}(A_q, B_x)}{S}(\min(m, n) - T),$$

which would be applied after Equation (1).

Activations that would fall outside of the 0–1 range are assigned the closest value in this range. Correspondence unit activations are initialized to random values selected from a normal distribution with a mean of 0.5 and a standard deviation of 0.05. In our simulations, Equation (1) is iterated for a fixed number of cycles. It is assumed that ABSURDIST places two elements into correspondences if the activation of their correspondence unit is greater than 0.55 after a fixed number of iterations have been completed. Thus, the network gives as output a complete set of proposed correspondences/translations between Systems *A* and *B*.

ASSESSING ABSURDIST

Our general method for evaluating ABSURDIST will be to generate a number of elements in an *N*-dimensional space, with each element identified by its value on each of the *N* dimensions. These will be the elements of System *A*, and each is represented as a point in space. Then, System *B*'s elements are created by copying the points from System *A* and adding Gaussian noise with a mean of 0 to each of the dimension values of each of the points. The motivation for distorting *A*'s points to generate *B*'s points is to model the common phenomenon that people's concepts are not identical, and are not identically related to one another. The Euclidean distance between every pair of elements within a system is calculated. The correspondences computed by ABSURDIST after Equation (1) is iterated are then compared to the correct correspondences. Two elements correctly correspond to each other if the element in System *B* was originally copied from the element in System *A*.

Tolerance to Distortion

An initial set of simulations was conducted to determine how robust the ABSURDIST algorithm was to noise and how well the algorithm scaled to different sized systems. As such, we ran a 11 × 6 factorial combination of simulations, with 11 levels of added noise and 6 different numbers of elements per system. Noise was infused into the algorithm by varying the displacement between corresponding points across systems. The points in System A were set by randomly selecting dimension values from a uniform random distribution with a range from 0 to 1000. System *B* points were copied from System *A*, and Gaussian noise with standard deviations of 0–1% was added to the points of *B*. The number of points per system was 3, 4, 5, 6, 10, or 15. Correspondences were computed after 1,000 iterations of equation (1). α was set to 0 (no external information was used to determine correspondences), β was set to 0.4. For each combination of noise and number of items, 1,000 separate randomized starting configurations were tested. The results from this simulation are shown in Figure 12.2, which plots the percentage of simulations in which each of the proper correspondences between systems is recovered. For example, for 15-item systems, the figure plots the percentage of time that all 15 correspondences are recovered. The graph shows that performance gradually deteriorates with added noise, but that the algorithm is robust to modest amounts of noise. Relative to the ABSURDIST algorithm described by Goldstone and

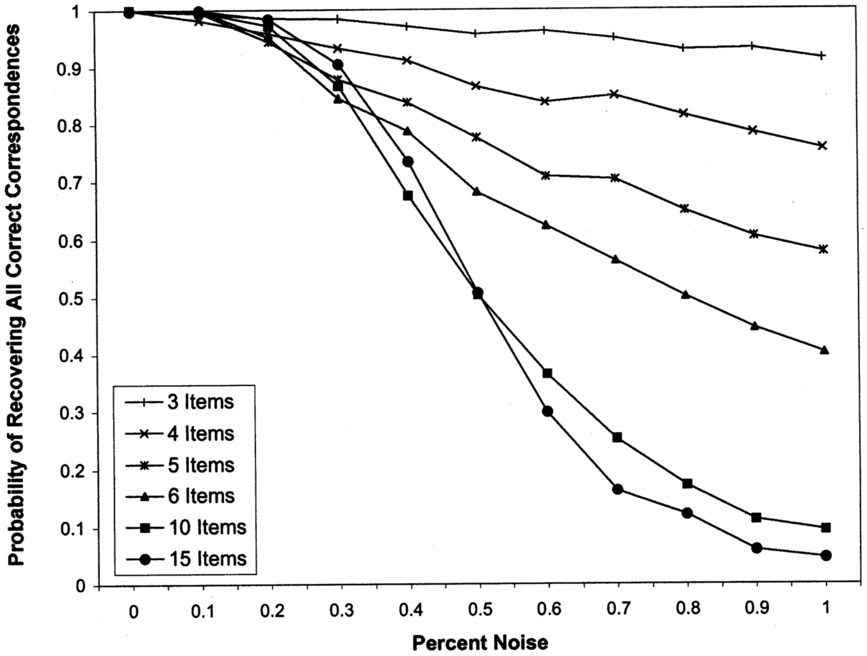

FIGURE 12.2. Probability of correctly translating every element in one system to every element in a second system, as a function of the number of items per system, and the amount of noise with which the elements of the second system are displaced relative to their positions in the first system.

Rogosky (2002) which lacked adaptive tuning of summed correspondence unit activation, the results in Figure 12.2 show about twice the noise tolerance and only one fifth of the iterations needed.

More surprisingly, Figure 12.2 also shows that for small levels of noise the algorithm's ability to recover true correspondences increases as a function of the number of elements in each system. Up to 0.2% noise, the highest probability of recovering all correct mappings is achieved for the largest, 15-item system. The reason for this is that as the number of elements in a system increases, the similarity relations between those elements provide increasingly strong constraints that serve to uniquely identify each element. The advantage of finding translations as the number of points in a system increases is all the more impressive when one considers chance performance. If one generated random translations that were constrained to allow only one-to-one correspondences, then the probability of generating a completely correct translation would be 1/n! when aligning systems that each have n items. Thus, with 0.7% noise, the 92% rate of recovering all 3 correspondences for a 3-item system is about 5.5 times above chance performance of 16.67%. However, with the same amount of noise, the 16% rate of recovering all of the correspondences for a 15-item system is remarkably higher than the chance rate of 7.6×10^{-13}. Thus, at least in our highly simplified domain, we have support for Lenat and Feigenbaum's (1991) argument that establishing meanings on the basis of within-system relations becomes more efficient, not harder, as the size of the system increases.

Integrating Internal and External Determinants of Conceptual Correspondences

The simulations indicate that within-system relations are sufficient for discovering between-system translations, but this should not be interpreted as suggesting that the meaning of an element is not also dependent on relations extrinsic to the system. In almost all realistic situations, translations between systems depends upon cues that are external to each system. For example, the alignment between John and Mary's **Horse** concepts is enormously facilitated by considerations other than within-system connections between concepts. Namely, strong evidence for conceptual alignment comes from the use of the same verbal label, pointing to the same objects when prompted, and being taught to use the concepts within similar cultures. Moreover, even though we have used within-system relations to represent conceptual webs, much of the basis for these within-system relations will come from external sources. It is possible that the only thing that somebody knows about flotsam is that it is similar to jetsam, but in most typical cases, the reason why two concepts are related to each other within a system is because of their perceptual-motor resemblances.

ABSURDIST offers a useful, idealized system for examining interactions between intrinsic (within-system) and extrinsic (external to the system) aspects of meaning. One way to incorporate extrinsic biases into the system is by initially seeding correspondence units with values. Thus far, all correspondence units have been seeded with initial activation values tightly clustered around 0.5. However, in many situations, there may be external reasons to think that two elements correspond to each other: they may receive the same label, they may have perceptual attributes in common, they may be associated with a common event, or a teacher may have provided a hint that the two elements correspond. In these cases, the initial seed-value may be significantly greater than 0.5.

Figure 12.3 shows the results of a simulation of ABSURDIST with different amounts of extrinsic support for a selected correspondence between two elements. Two systems are generated by randomly creating a set of points in two dimensions for System 1, and copying the points' coordinates to System 2 while introducing noise to their positions. When Seed = 0.5, then no correspondence is given an extrinsically supplied bias. When Seed = 0.75, then one of the true correspondences between the systems

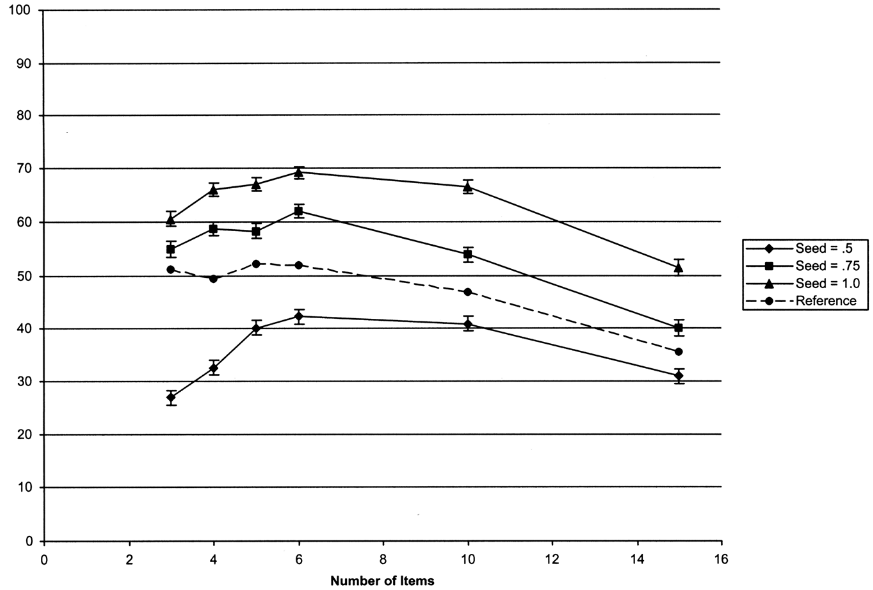

FIGURE 12.3. Percentage of correct alignments found by ABSURDIST, as a function of the number of items per system, and the amount of external bias that seeds a single correct alignment between two elements. As the strength of external bias increases, the percentage of correct correspondences increases, and this increase exceeds the increase predicted if seeding one alignment only affected the alignment itself (the "Reference" line). As such, the influence of extrinsic information is accentuated by within-system relations.

is given a larger initial activation than the other correspondences. When Seed = 1.0, this single correspondence is given an even larger initial activation. Somewhat unsurprisingly, when a true correspondence is given a relatively large initial activation, then ABSURDIST recovers a higher percentage of correct correspondences. The extent of this improvement is more surprising. For example, for a system made up of 15 elements, a mapping accuracy of 31% is obtained without any extrinsic assistance (Seed = 0.5). If seeding a single correct correspondence with a value of 1 rather than 0.5 allowed ABSURDIST to recover just that one correspondence with 100% probability, then accuracy would increase at most to 35.6% $(((0.31 * 14) + 1)/15)$. The reference line in Figure 12.1 shows these predicted increases in accuracy. For all systems tested, the observed increment in accuracy far outstretches the increase in accuracy predicted if seeding a correspondence only helped that correspondence. Moreover, the amount by which translation accuracy improves beyond the amount predicted generally increases as a function of system size. Thus, externally seeding a correspondence does more than just fix that correspondence. In a system where correspondences all mutually depend upon each other, seeding one correspondence has a ripple-effect through which other correspondences are improved. Although external and role-based accounts of meaning have typically been pitted against each other, it turns out that the effectiveness of externally grounded correspondences is radically improved by the presence of role-based correspondences.

Equation 2 provides a second way of incorporating extrinsic influences on correspondences between systems. This equation defines the net input to a correspondence unit as an additive function of the extrinsic support for the correspondence, the intrinsic support, and the competition against it. Thus far, the extrinsic support has been set to 0. The extrinsic support term can be viewed as any perceptual, linguistic, or top-down information that suggests that two objects correspond. For example, two people using the same verbal label to describe a concept could constitute a strong extrinsic bias to place the concepts in correspondence. To study interactions between extrinsic and intrinsic support for correspondences, we conducted 1,000 simulations that started with 10 randomly placed points in a two-dimensional space for System A, and then copied these points over to System B with Gaussian-distributed noise. The intrinsic, role-based support is determined by the previously described equations. The extrinsic support term of Equation 2 is given by

$$E(A_q, B_x) = e^{-D(A_q, B_x)},$$

where D is the Euclidean distance function between point q of System A and point x of System B. This equation mirrors the exponential similarity function used to determine intrinsic similarities, but now compares absolute coordinate values. Thus, the correspondence unit connecting A_q and B_x will tend to be strengthened if A_q and B_x have similar coordinates. This is extrinsic support because the similarity of A_q and B_x's coordinates can be determined without any reference to other elements. If the two dimensions reflect size and brightness for example, then for A_q and B_x to have similar coordinates would mean that they have similar physical appearances along these perceptual dimensions.

In conducting the present simulation, we assigned three different sets of weights to the extrinsic and intrinsic support terms. For the "Extrinsic only" results of Figure 12.4, we set $\alpha = 0.4$ and $\beta = 0$. For this group, correspondences are only based on the extrinsic similarity between elements. For the "Intrinsic only" results, we set $\alpha = 0$, and $\beta = 0.4$. This group is comparable to the previous simulations in that it uses only a role-based measure of similarity to establish correspondences. Finally, for "Intrinsic and Extrinsic," we set $\alpha = 0.2$ and $\beta = 0.2$. For this group, correspondences are based on both absolute coordinate similarity, and on elements taking part in similar relations to other elements. Note, that both the intrinsic and

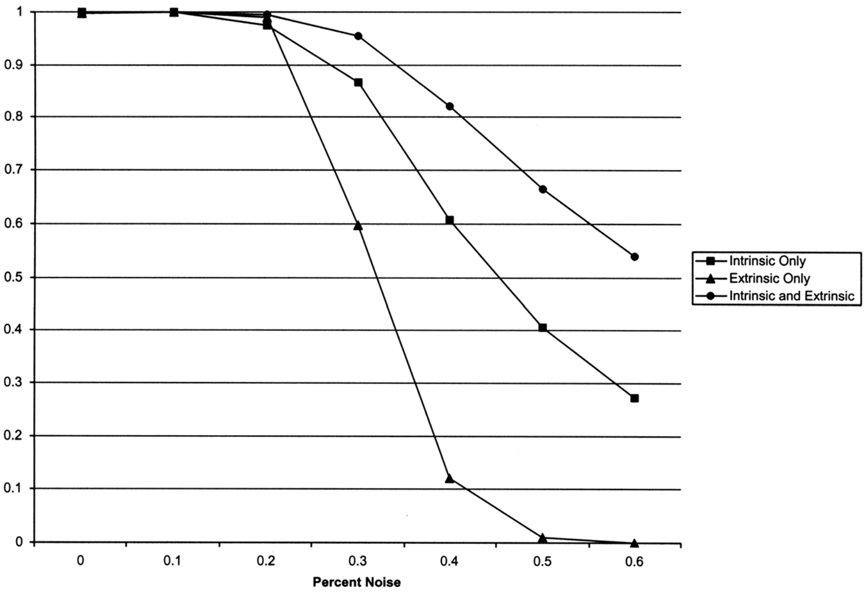

FIGURE 12.4. Probability of ABSURDIST achieving a perfect translation between two systems, as a function of noise, and the weighting of extrinsic and intrinsic information. Better performance is achieved when all weight is given to intrinsic information than when only extrinsic information is used. However, the best performance is achieved when both sources of information are weighted equally.

extrinsic terms are based on the same coordinate representations for elements. The difference between these terms centers on whether absolute or relative coordinate values are used.

Figure 12.4 shows that using only information intrinsic to a system results in better correspondences than using only extrinsic information. This is because corresponding elements that have considerably different positions in their systems can often still be properly connected with intrinsic information if other proper correspondences can be recovered. The intrinsic support term is more robust than the extrinsic term because it depends on the entire system of emerging correspondences. For this reason, it is surprising that the best translation performance is found when intrinsic and extrinsic information are both incorporated into Equation 2. The excellent performance of the network that uses both intrinsic and extrinsic information derives from its robustness in the face of noise. Some distortions to points of System *B* adversely affect the intrinsic system more than the extrinsic system, whereas other distortions have the opposite effect. A set of distortions may have a particularly disruptive influence on either absolute coordinates or relative positions. A system that incorporates both sources of information will tend to recover well from either disruption if the other source of information is reasonably intact.

discussion of initial simulations

The first theoretical contribution of ABSURDIST is to show that translations *between* elements of two systems can be found using only information about the relations between elements *within* a system. ABSURDIST demonstrates how a conceptual web account of meaning is compatible with the goal of determining correspondences between concepts across individuals. Two people need not have exactly the same systems to create proper conceptual correspondences. Contra Fodor (1998; Fodor & Lepore, 1992), information in the form of interconceptual similarities suffices to find intersystem equivalences between concepts. Furthermore, it is sometimes easier to find translations for large systems than small systems. This is despite two large disadvantages for systems comprising many elements: there are relatively many opportunities to get the cross-system alignments wrong, and the elements tend to be close together and hence confusable. The compensating advantage of many-element systems is that the roles that an object plays within a system are more elaborated as the number of elements in the system increases.

The second theoretical contribution of ABSURDIST is to formalize some of the ways that intrinsic, within-system relations and extrinsic, perceptual information synergistically interact in determining conceptual alignments. Intrinsic relations suffice to determine cross-concept translations, but if extrinsic information is available, more robust, noise-resistant translations can be found. Moreover, extrinsic information, when available, can actually increase the power of intrinsic information.

The synergistic benefit of combining intrinsic and extrinsic information sheds new light on the debate on accounts of conceptual meaning. It is common to think of intrinsic and extrinsic accounts of meaning as being mutually exclusive, or at least being zero-sum. Seemingly, either a concept's meaning depends on information within its conceptual system or outside of its conceptual system, and to the extent that one dependency is strengthened, the other dependency is weakened.

In contrast to this conception of competition between intrinsic and extrinsic information, meaning in ABSURIDST is both intrinsically and extrinsically determined, and the external grounding makes intrinsic information more, not less, powerful. An advantage of this approach to conceptual meaning is that it avoids an infelicitous choice between reducing conceptual meanings to sense data and leaving conceptual systems completely ungrounded. Having concepts in a system all depend upon each other is perfectly compatible with these concepts having a perceptual basis.

other cognitive science applications of absurdist

We have focused on the application of ABSURDIST to the problem of translating between different people's conceptual systems. However, the algorithm is applicable to a variety of situations in which elements from two system must be placed in correspondence in an efficient and reasonable (though not provably optimal) manner. A combination of properties makes ABSURDIST particularly useful for applications in cognitive science: (1) the algorithm can operate solely on relations within a system; (2) the within-system relations can be as simple as generic similarity relations; (3) the algorithm can combine within-system and between-systems information when each is available; (4) the algorithm has a strong bias to establish one-to-one correspondences; and (5) the algorithm does not require larger numbers of iterations for convergence as the number of elements per system increases.

Aligning Subsystems

In the simulations thus far considered, the systems to be aligned have had the same number of elements. This is not required for the algorithm. When different-sized systems are compared, the correspondences tend to be one-to-one, but many elements of the larger system are not placed into correspondence at all. One useful application of aligning systems of unequal size is finding subsystems that match a pattern. An example of this from conceptual translation might be the comparison of people with very different levels of knowledge about a domain. If Mary is a large animal vet, then she will have many more horse-related concepts than John. Some of Mary's concepts may not have correspondences in John, but it would still be useful to identify the equivalent of John's concepts in Mary's system. Figure 12.5A presents a simplified example of this scenario, in which a three-element pattern can be identified within a larger seven-element pattern by aligning the target three-element pattern with the larger pattern. In this fashion, ABSURDIST provides an algorithm for finding patterns concealed within larger contexts.

A second use of finding alignments between sub-systems is as a method of making consistent forced-choice similarity judgments. To decide whether Pattern A is more similar to Pattern Y or Z, ABSURDIST can be given Pattern A as System 1, and Patterns Y and Z as System 2. In Figure 12.5B, Patterns A, Y, and Z each consist of three elements. In this example, Y and Z correspond equally well to Pattern A. Even in this case, ABSURDIST creates a consistent alignment in which all of the elements of A are placed into correspondence with either elements from Y or Z, and each alignment occurs on roughly half of the simulations. The mutually supportive constraint satisfaction virtually assures that a consistent alignment will be found, similar to the consistent perception of ambiguous forms found

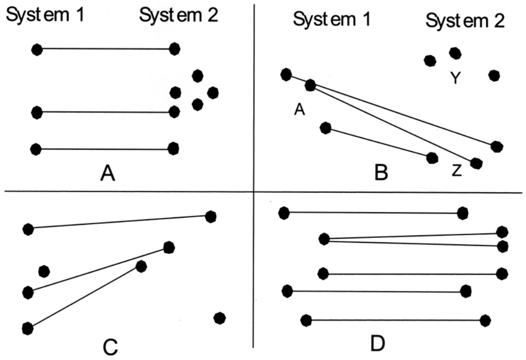

FIGURE 12.5. Four examples of ABSURDIST translations, with typical alignments shown by solid line connecting elements across the systems. In Panel A, the pattern represented by System 1 is aligned with the subsystem of System 2 that optimally matches this pattern. In Panel B, even though two subsystems of System 2 can correspond to the elements of System 1, one of the coherent subsystems will become strongly aligned with System 1 as the other loses its alignment to System 1 completely. In Panel C, the three pairs of elements that have mostly comparable similarity relations within their systems are placed into correspondence, but a fourth element of each system is not placed into any correspondence because it is too dissimilar to other elements in terms of its similarity relations. In Panel D, a two-to-one mapping is constructed because two elements from the System 2 have very similar relations to a single element from the System 1.

by Rumelhart, Smolensky, McClelland, and Hinton (1986). If Z aligns better with A than does Y, then it would be selected reliably more often.

A particularly challenging situation for ABSURDIST occurs if two systems have the same number of elements, but only a subset of them properly matches. For example, Mary and John both have concepts of **Horse**, **Stallion**, and **Pony**, but only Mary has a concept of **Palomino** and only John has a concept of **Pinto**. This situation is implemented in Figure 12.5C by having four elements per system, with three of the elements matching well across the systems, but one element from each system having no strong correspondence in the other system. This is challenging because ABSURDIST's one-to-one mapping constraint will tend to match two elements if neither participates in any other strong correspondences. Despite this tendency, given the situation shown in Figure 12.5C, ABSURDIST, will draw correspondences between the three pairs of elements that share the majority of their roles in common, but not between the fourth, mismatching elements. The unit that places the mismatching elements into correspondence does receive excitation from the three units that place properly matching elements into correspondence due to one-to-one mapping consistency. However, the lack of similarity between the mismatching elements' similarity relations to other elements overshadows this excitation.

Finally, unlike most computer science graph matching algorithms, ABSURDIST can violate a one-to-one mapping constraint under certain circumstances. Figure 12.5D presents one such configuration, in which two objects from a larger system enter into very similar similarity relations to a single object from the other system. This typically happens when the two objects from the larger system are themselves very similar. These particular circumstances allow ABSURDIST to aptly model conceptual change over development. For many domains, adults have more highly differentiated concepts than do children (Smith, Carey, & Wiser, 1985). To take examples from Carey (1999), the adult concepts of heat and temperature are not differentiated for children, nor are weight and density. In Figure 12.5D, System 2 might correspond to a part of an adults' conceptual system, with separate but close concepts for heat and temperature. Both of these concepts correspond to a single concept in the younger and less differentiated System 1. In this manner, ABSURDIST shows how it is possible to translate between conceptual systems even if the systems are differentially articulated. A second example of the situation shown in Figure 12.5D occurs when languages carve up the semantic universe in different manners (Gentner & Goldin-Meadow, 2003). Dutch, for example, uses the same word "Schaduw" to refer to things that, in English, would be called "Shadow" or "Shade." Conversely, English uses the word "Leg" to refer to things that, in Dutch, would be called "Been" (the legs of people and horses) or "Poot" (the legs of other animals and furniture) (R. Zwaan, personal communication, 2004). In these cases, ABSURDIST is able to align "Schaduw" from the Dutch system with two concepts from the English system.

Object Recognition and Shape Analysis

The ABSURDIST algorithm can be applied to the problem of object recognition that is invariant to rotation and reflection. For this application, a pictorial object is the system, and points on the object are elements of the system. Unlike many approaches to object recognition (Ullman, 1996), ABSURDIST's robustness under rotation is achieved automatically rather than being explicitly computed. Figure 12.5 shows the alignments obtained when one object is placed into correspondence with a rotated version of itself. These alignments are useful for object recognition. For example, if the form on the left side of Figure 12.6 were memorized, then the form on the right can be identified as an exact match to this memorized form without needing to rotate either form. Rotation is not required because ABSURDIST uses relative distances between points to determine correspondences. Distances between points are not affected by rotation or reflection.

A standard solution to recognizing rotated objects is to find critical landmark points that are identifiable on a stored object and a presented input

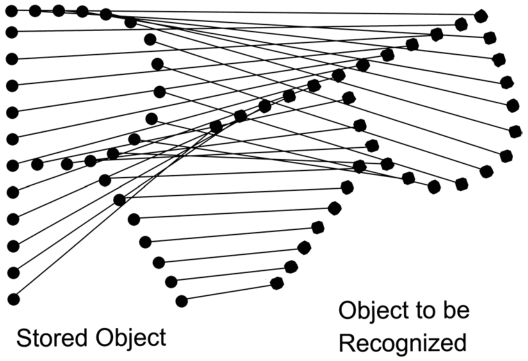

FIGURE 12.6. ABSURDIST can be applied to orientation- and translation-invariant object recognition. An object to be recognized can be aligned to a stored version of the object without requiring either object to be rotated, and without requiring predetermined matching points to be available. Within-object relations suffice to unambiguously align any asymmetric object to a stored version of itself.

object. For example, a distinctive angle or colored point may be unambiguously identified on stored and probed objects. Ullman (1989) has shown that if three such landmarks can be identified, then two views of the same 3-D object can be perfectly aligned. ABSURDIST shows that even if zero extrinsically aligned landmarks can be identified, it is still possible to align the objects. ABSURDIST does so by taking advantage of the wealth of information in the form of within-object relations (see also Edelman, 1999). Object recognition algorithms have typically used extrinsic information without considering the useful constraints provided by solely intrinsic information.

ABSURDIST gets rotational and reflection invariance for "free" because inter-point distances are used and these are inherently relational. Size invariance is easily achieved by normalizing distances between points of an object so as to fall in a 0-to-1 range. In situations where rotation or size *does* make a difference ("b" is not the same object as "d"), extrinsic, absolute information can be added to Equation 2 to supplement the intrinsic information. Alternatively, a point can be added in a specific location, such as the upper left hand corner, to both the memorized and probed objects. If these anchor points are added to the system, then ABSURDIST correctly predicts that, up to 180 degrees, as the angular disparity between two objects increases, so does the time required to align them (Shepard & Cooper, 1986). In particular, if we measure ABSURDIST's response time by the number of cycles of activation passing required to reach a threshold level of summed unit activation, then rotating one object relative to the other slows response times. If the anchor points are not included, then ABSURDIST is completely unaffected by rotation, which improves the algorithm's object recognition speed, but at the cost of making it less psychologically plausible.

ABSURDIST can also be used to analyze the symmetries of an object for shape analysis. A symmetry is a distance-preserving transformation that copies an object exactly onto itself. For example, the capital letter "A" has a reflection symmetry because every point of the A matches another point on the A if they are reflected across a vertical line running down the center of the "A." A square has eight symmetries (four rotational symmetries multiplied by two reflectional symmetries). ABSURDIST determines the symmetries of an object by finding translations *between* the object and itself. For a random set of points, the only translation that ABSURDIST is likely to find is the one that aligns each point with itself. However, for four points arranged in a square, eight different sets of correspondences are equally probable, corresponding to the eight symmetries of a square. An equilateral triangle produces six different translations; points A, B, C (ABC) correspond with equal likelihood to ABC, ACB, BCA, BAC, CAB, or CBA. An isosceles triangle has only two different translations, corresponding to an identity and reflection transformation. If some randomness is added to Equation (1), then the distribution of alignments reveals not only the symmetries of a shape, but also the near-symmetries of a shape, with the probability of an alignment denoting how close the associated symmetry is to perfection.

Pre- and Postprocessing for Multidimensional Scaling

The ABSURDIST algorithm bears some similarities to multidimensional scaling (MDS), but its input and output are fundamentally different. MDS takes as input a matrix of proximities by which each object in a set is compared to every other object, and the matrix is ordered such that each matrix entry is identified with a particular pair (e.g., "Entry {2,3} is the similarity of **Bear** to **Dog**"). The output to MDS is a geometric representation of the objects, with each object represented as a point and the distances between points approximating their dissimilarity. MDS also provides methods, such as individual Differences MDS (Carroll & Chang, 1970) that can be used to compare two people's MDS solutions and derive a common geometric solution for both people, as long as the same objects are given to the two people, and they are ordered in an identical manner in the two persons' proximity matrices. ABSURDIST supplements MDS by applying in situations where the two people are comparing similar set of objects, but the people do not use the same labels for the objects (e.g., one person is comparing objects in Chinese, and the other in English), or the matrices are not in the same order. The output of ABSURDIST is a suggested way to reorder the matrices (a translation).

Using a pre-processor to MDS, ABSURDIST can reorder proximity matrices so that they are in the best alignment possible. For examples, a set of common concepts or objects can be evaluated by different cultures, language groups, scientific fields, or philosophical movements. A single, grand MDS can be derived for all of these groups once ABSURDIST has aligned the different groups' concepts as well as possible. The stress in ABSURDIST's alignments can be used to gauge how appropriate it is to generate a single geometric solution to represent several groups' assessments of similarity. Stress would be measured by finding the cumulative dissimilarity between the systems across of all the aligned distances.

ABSURDIST can also be applied as a post-processor once an MDS solution has been obtained. The distance measures between points in an MDS space are ideal inputs to ABSURDIST because they provide exactly the kind of relationally defined similarity matrix that ABSURDIST requires. We can use ABSURDIST to determine whether two different MDS solutions are in fact similar to each other, unaffected by reflections, translations, and rotations.

Interpretation of Neural Networks

ABSURDIST can be used to determine whether two neural networks have developed similar solutions to a problem. This application builds on a recent proposal by Laakso and Cottrell (1998, 2000) to compare neural networks by comparing the distances between their neural activities. By their approach, two neural networks offer a similar solution to a problem if input patterns produce similar, *relative* patterns of activation across the two networks. For example, if an input representing **Dog** produces similar hidden-unit activations to a **Cat** input in one network, and if they produce similar hidden-unit activations in another network, this would be some reason to believe that the networks are functionally similar despite any differences they might have in terms of their architecture. More formally, they compute all distances between activation patterns produced by a set of inputs, for each of two networks to be compared. They then compute the correlation between the distances for the two networks to determine their similarity.

Similar to ABSURDIST, Laakso and Cottrell's approach (2000) emphasizes the role-based, relative nature of similarity, and how it suffices to compare architecturally different systems. ABSURDIST significantly extends their approach by using even more relational, less absolute information to compare systems. Their technique requires that one know what stimulus is presented to the two networks (e.g., **Dog** in the example above). With ABSURDIST, two networks can be compared even if one cannot match up inputs to the networks in a predetermined way. All that is required is that the inputs to the two networks are sampled from essentially the same distribution of concepts/objects. This feature has a number of potential uses. It can be used when (1) the inputs to two networks are necessarily different because of differences in the architectures used for representing inputs, (2) the inputs to two networks are represented in different languages or representational formats and a direct translation is not available, and (3) there is a potentially infinite number of inputs, and one does not have control over the input that is presented to a network on a given trial. As an example of this third application, imagine that an infinite number of stimuli fall into 10 clusters with varying similarities to each other. Using some competitive specialization algorithm (Kohonen, 1995; Rumelhart & Zipser, 1985), a set of originally homogeneous hidden units can be trained so that each becomes specialized for one of the input clusters. By Laakso and Cottrell's approach, we can measure the similarity of two networks (that may differ in their number of hidden units) by feeding in known and matched input patterns to each network, and comparing the similarity of the networks' similarity assessments, measured by hidden unit activations, for all pairs of stimuli. Using ABSURDIST, we do not even need to feed the two networks exactly the same input patterns or match input patterns across the two networks. ABSURDIST can still measure whether the entire pattern of similarities among the hidden units is similar in the two networks. Along the way to determining this, it will align clusters of inputs across the networks according to relative, rather than absolute, input values.

Human Analogy and Comparison

As described earlier, ABSURDIST offers a complementary approach to analogical reasoning between domains. Most existing models of analogical comparison, including SME, SIAM, LISA, Drama, and ACME (Eliasmith & Thagard, 2001; Falkenhainer et al., 1989; Goldstone, 1994; Holyoak & Thagard, 1989; Hummel & Holyoak, 1997, 2003), represent the domains to be compared in terms of richly structured propositions. This is a useful strategy when the knowledge of a domain can be easily and unambiguously expressed in terms of symbolic predicates, attributes, functions, and higher-order relations. Determining the right symbolic encoding of a domain is a crucial, yet often finessed problem (Hofstadter, 1995). In many cases, such as single words or pictures, it is difficult to come up with propositional encodings that capture an item's meaning. In such cases, ABSURDIST's unstructured representation is a useful addition to existing models of analogical reasoning. From this perspective, ABSURDIST may apply when these other models cannot, in domains where explicit structural descriptions are not available, but simple similarity relations are available. For example, a German-English bilingual could probably provide subjective similarity ratings of words within the set {Cat, Dog, Lion, Shark, Tortoise} and separately consider the similarities of the words within the set {Katze, Hunde, Löwe, Hai, Schildkröte}. These similarities would provide the input needed by ABSURDIST to determine that "Cat" corresponds to "Katze." However, the same bilingual might not be able to provide the kind of analytic and structured representation of "Cat" that the other models require.

Psychologists have recently argued that similarity and difference are more similar than might appear at first (Gentner & Markman, 1994; Markman, 1996; Markman & Gentner, 2000; Medin, Goldstone, & Gentner, 1990). In particular, assessing the similarity *or* dissimilarity of two objects typically involves placing the objects into alignment as well as possible. One empirically confirmed prediction of this view is that it in many cases it is easier to find differences between similar than dissimilar objects (Gentner & Markman, 1994). Antonyms, while ostensibly opposites, are in fact quite similar in that all of their semantic attributes are perfectly aligned, with one attribute having opposite values. For example, **Black** and **White** are both terms for monochromatic, pure colors. ABSURDIST is consistent with this perspective. The alignment that it calculates would precede both similarity and dissimilarity judgments. Similarity judgments are based on the quality of alignment, and on the similarity of the corresponding elements. Dissimilarity judgments would be based on the dissimilarity of corresponding elements, and thus would also involve determining an optimal alignment between systems.

Large-Scale System Translation

The small-scale simulations conducted leave open the promise of applying ABSURDIST to much larger translation tasks. Although logistical and technical problems will certainly arise when scaling the algorithm up to large databases, the presented approach should theoretically be applicable to systems such as dictionaries, thesauri, encyclopedias, and social organizational structures. For example, ABSURDIST could provide automatic translations between dictionaries of two different languages using only co-occurrence relations between words within each dictionary. The input to the network would be the full matrix of co-occurrences between every word in English to every other word in English, and the same kind of matrix for a second language. The output would be a set of correspondences across the two language. If such a project were successful, it would provide a striking argument for the power of within-system relations. If unsuccessful, it could still be practically valuable if supplemented by a small number of external hints (e.g., that French "chat" and English "cat" might correspond to each other because of their phonological similarity).

We are not optimistic that a completely unseeded version of ABSURDIST would recover a very accurate translation between two dictionaries. We have collected a set of subjective similarities among a set of 134 animal words from two groups of subjects. The two groups were obtained by randomly assigning each of 120 Indiana University students to one of the groups. We used ABSURDIST to try correctly align the animal words across the two groups of subjects using only each groups' matrix of similarity assessments. ABSURDIST's performance rate of 34% correctly aligned animals was encouraging, but not nearly good enough to be practically useful. Furthermore, performance was higher than might generally be expected because of the high similarity between the groups, and the large number of subjects reducing extraneous noise. If we had tried to align a single pair of randomly selected subjects, ABSURDIST's performance would have been much worse. Although the unseeded ABSURDIST's performance was lackluster, we again found dramatic improvements when even a few animal terms were correctly seeded. An automatic dictionary translator could well be useful even if it needed to be seeded with 5% of the correct matches.

Translating Structured Representations

The simplicity of ABSURDIST's input representations is both a strength and a weakness. From a philosophical and rhetorical perspective, the ability of ABSURDIST to recover translations using only two similarity matrices to represent the compared systems is the strongest possible argument for the sufficiency of purely relational, within-system information for translation. However, in many cases, more structured information is readily available and could be used to improve translation performance. With this in mind, we have extended ABSURDIST so as to be able to handle generalized graph representations, including graphs with labeled relations, bidirectional or unidirectional relations, and sparse network topologies. These extensions allow ABSURDIST to be applied to many of the domains of focus for graph matching algorithms (e.g. Larkey & Love, 2003; Melnik, Garcia-Molina, & Rahm, 2002). One particularly timely application is the translation between XML documents on the web. XML is a meta-language for expressing databases using terminology created by database programmers. Unfortunately, different database programmers may use different terminology for describing essentially the same entities. For example, in Figure 12.7, two different designations, "worker" and "employee" have been used to describe essentially the same entity. The extended version of ABSURDIST, like other graph matching algorithms and analogical reasoning systems (Falkenhainer et al., 1989; Hummel & Holyoak, 2003), can be used to determine translations between the databases. Whether ABSURDIST offers tangible benefits over existing algorithms for structured graphs is not completely clear yet, but some potential advantages of ABSURDIST are: (1) it has a soft, rather than hard, 1-to-1 mapping constraint and so can find translations even when one database makes a distinction between two or more entities that are conceptually fused together in the other database; (2) it can integrate structured graph representations

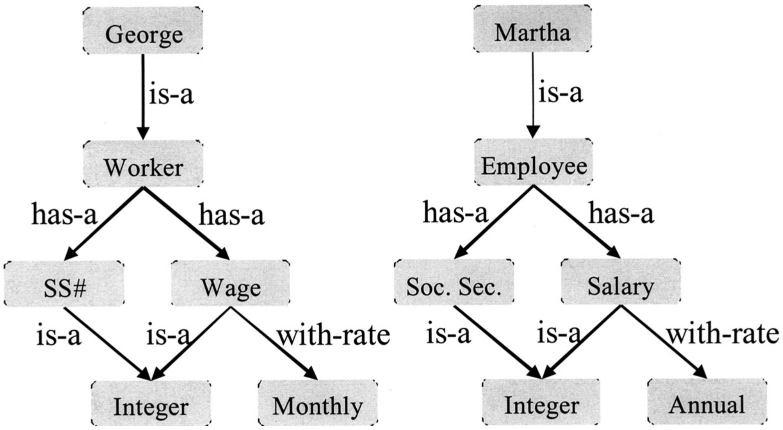

FIGURE 12.7. An example of the structured representations that the extended ABSURDIST algorithm can align.

with association-based similarity matrices; (3) it provides a natural mechanisms for incorporating external and internal determinants of meaning; and (4) the neural network architecture offers a framework for solving "correspondence problems" that has been shown to have neurophysiological plausibility in related perceptual domains (Marr & Poggio, 1979; Ullman, 1979).

Initial results with the graph-based version of ABSURDIST suggest a number of interesting trends (Feng, Goldstone, & Menkov, 2004). First, as expected, adding multiple types of relations such as **is-a** and **has-a** (see Figure 12.7) allows ABSURDIST to more accurately construct translations than it did with the single, generic similarity relation. Second, translations are more accurately found with asymmetric relations like **left-of** rather than symmetric relations like **next-to**. Third, the best translations are found when approximately half of the possible connections between concepts are present. The full connectivity that we have thus far been assuming does not give rise to the best translations. More sparsely connected systems actually have more unambiguous features that facilitate alignment. All three of these results are consistent with the idea that information that locally distinguishes concepts from each other leads to the best global translations.

CONCLUSION

Our simulations indicate that appropriate translation between conceptual systems can sometimes proceed on the basis of purely within-system conceptual relations. It is not viciously circular to claim that concepts gain their meaning by their relations to other concepts that, in turn, gain the meaning in the same fashion. Remarkably, a concept's connections to other concepts in the same system can suffice to give it a meaning that can transcend the system, at least in the sense of establishing proper connections to concepts outside of the system.

Despite the surprising power of within-system relations, our claim is not that translation typically *does* proceed with only this information. To the contrary, our simulations point to the power of combining intrinsic, within-system relations and external grounding. Conceptual-web accounts of meaning can offer an account of meaning, but they will be most effective when combined with an externally grounded component. In most real-world translation scenarios, external grounding is crucial. To return to our introductory example, translating between John and Mary's **Horse** concepts is immeasurably facilitated by virtue of the fact that they use the same words to refer to their **Horse** concepts, they originally learned their **Horse** concepts from the same kinds of inputs, and they gesture to the same things when asked to provide examples of horses. These are all external cues to translation, and play at least as large a role as within-system relations. However, our point is that we do not need to select either internal or external sources as the definitive provider of meaning. By choosing between or even separating these sources, their potential is weakened.

Like many of the other contributors to this book, we are impressed by the tight connection between our supposedly abstract, sophisticated concepts and our perceptual and motor capacities. The traditional split between research on low-level perceptual processes and high-level cognition is misleading and counterproductive. Elsewhere we have attempted to describe bridges between our perceptual and conceptual systems (Goldstone, 2003; Goldstone & Barsalou, 1998; Goldstone et al., 2000). In the present work, we have tried to strengthen this bridge by arguing that proponents of an approach to concepts grounded in perceptual-motor activity need not snub intraconceptual relations or claim that all thought reduces to sense data.

We believe that separating embodied cognition approaches from sensory reductionism makes embodied cognition substantially more palatable. One can maintain that all concepts have perceptual-motor components without claiming that concepts reduce to sensations and actions. A sensory reductionist easily falls into the trap of claiming that **Horse's** meaning is given by its parts – hoof, mane, tail, etc. These parts, in turn, are described in terms of their parts, and so on, until eventually the descriptions are grounded in elementary sensory elements. This account is dissatisfying because it overlooks much of people's rich knowledge associated with horses. Horses are associated with cavalries, Paul Revere, races, and cowboys. Several of the associated concepts that imbue **Horse** with its meaning, such as **Freedom**, **Domesticated**, and **Strength**, are less concrete than **Horse** itself. Sensory reduction can be avoided by positing interconceptual relations such as these, that coexist with perceptual groundings. We do not have to choose between meaning based on interconceptual relations or perceptual senses, and by not choosing we make it much more plausible that all of our concepts have perceptual-motor components. Moreover, it is not simply that these complementary aspects of meaning coexist. Internal and external bases of meaning are mutually reinforcing, not mutually exclusive.

ACKNOWLEDGMENTS

The authors wish to express thanks to Eric Dietrich, Shimon Edelman, Gary Cottrell, John Hummel, Michael Lynch, Arthur Markman, Robert Nosofsky, Diane Pecher, Richard Shiffrin, Mark Steyvers, and two anonymous reviewers. This research was funded by a grant from Lockheed Martin and NSF grant 0125287. Correspondence concerning this article should be addressed to rgoldsto@indiana.edu or Robert Goldstone, Psychology Department, Indiana University, Bloomington, Indiana 47405. Further information about the laboratory can be found at http://cognitrn.psych.indiana.edu/.

References

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

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences 7*, 84–91.

Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy 10*, 615–78.

Block, N. (1999). Functional role semantics. In R. A. Wilson, & F. C. Keil (Eds.), *MIT Encyclopedia of the Cognitive Sciences* (pp. 331–332). Cambridge, MA: MIT Press.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition 75*, 1–28.

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science 13*, 185–189.

Burgess, C., Livesay, K., & Lund, K. (1998). Explorations in context space: Words, sentences, and discourse. *Discourse Processes 25*, 211–257.

Burgess, C., & Lund, K. (2000). The dynamics of meaning in memory. In E. Diettrich & A. B. Markman (Eds.), *Cognitive Dynamics: Conceptual Change in Humans and Machines* (pp. 117–156). Mahwah, NJ: Lawrence Erlbaum Associates.

Carey, S. (1999). Knowledge acquisition: Enrichment or conceptual change. In E. Margolis & S. Laurence (Eds.), *Concepts: Core Readings* (pp. 459–487). Cambridge, MA: MIT Press.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an *n*-way generalization of "Eckart-Young" decomposition. *Psychometrika 35*, 283–319.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic priming. *Psychological Review 82*, 407–428.

Edelman, S. (1999). *Representation and Recognition in Vision*. Cambridge, MA: MIT Press.

Eliasmith, C., & Thagard, P. (2001). Integrating structure and meaning: A distributed model of analogical mapping. *Cognitive Science 25*, 245–286.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence 41*, 1–63.

Feng, Y., Goldstone, R. L., & Menkov, V. (2004). ABSURDIST II: A Graph Matching Algorithm and its Application to Conceptual System Translation. FLAIRS 2004.

Field, H. (1977). Logic, meaning, and conceptual role. *Journal of Philosophy 74*, 379–409.

Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Clarendon Press.

Fodor, J., & Lepore, E. (1992). *Holism*. Oxford: Blackwell.

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science 5*, 148–152.

Gentner, D., & Goldin-Meadow, S. (2003). *Language in Mind: Advances in the Study of Language and Thought*. Cambridge, MA: MIT Press.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review 9*, 558–565.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language 43*, 379–401.

Goldstone, R. L. (1994). Similarity, Interactive Activation, and Mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition 20*, 3–28.

Goldstone, R. L. (1996). Isolated and Interrelated Concepts. *Memory and Cognition 24*, 608–628.

Goldstone, R. L. (2003). Learning to perceive while perceiving to learn. In R. Kimchi, M. Behrmann & C. Olson (Eds.), *Perceptual Organization in Vision: Behavioral and Neural Perspectives* (pp. 233–278). Mahwah, NJ: Lawrence Erlbaum Associates.

Goldstone, R. L., & Barsalou, L. (1998). Reuniting perception and conception. *Cognition 65*, 231–262.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition 78*, 27–43.

Goldstone, R. L., & Rogosky, B. J. (2002). Using relations within conceptual systems to translate across conceptual systems. *Cognition 84*, 295–320.

Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual Interrelatedness and Caricatures. *Memory and Cognition 31*, 169–180.

Goldstone, R. L., Steyvers, M., Spencer-Smith, J., & Kersten, A. (2000). Interactions between perceptual and conceptual learning. In E. Diettrich & A. B. Markman (Eds.), *Cognitive Dynamics: Conceptual Change in Humans and Machines* (pp. 191–228). Mahwah, NJ: Lawrence Erlbaum Associates.

Harnad, S. (1990). The symbol grounding problem. *Physica D 42*, 335–346.

Hofstadter, D. (1995). *Fluid concepts and creative analogies*. New York: Basic Books.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science 13*, 295–355.

Hume, D. (1740/1973). *An Abstract of a Treatise on Human Nature*. Cambridge: Cambridge University Press.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review 104*, 427–466.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review 110*, 220–264.

Kohonen, T. (1995). *Self-Organizing Maps*. Berlin: Springer-Verlag.

Kuhn. T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Laakso, A., & Cottrell, G. (1998). "How can I know what you think?": Assessing representational similarity in neural systems. In *Proceedings of the 20th Annual Cognitive Science Conference*. Madison, WI: Lawrence Erlbaum (pp. 591–596).

Laakso, A., & Cottrell, G. (2000). Content and cluster analysis: Assessing representational similarity in neural systems. *Philosophical Psychology 13*, 47–76.

Lakoff, G., & Nunez, R. E. (2000). *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. New York: Basic Books.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis Theory of the acquisition, induction, and representation of knowledge. *Psychological Review 104*, 211–240.

Larkey, L. B., & Love, B. C. (2003). CAB: Connectionist analogy builder. *Cognitive Science 27*, 781–794.

Lenat, D. B., & Feigenbaum, E. A. (1991). On the thresholds of knowledge. *Artificial Intelligence 47*, 185–250.

Locke, J. (1690). *An Essay Concerning Human Understanding*. (http://www.ilt.columbia.edu/Projects/digitexts/locke/understanding/title.html).

Markman, A. B. (1996). Structural alignment in similarity and difference judgments. *Psychonomic Bulletin and Review 3*, 227–230.

Markman, A. B., Gentner, D. (2000). Structure mapping in the comparison process. *American Journal of Psychology 113*, 501–538.

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London 204*, 301–328.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1990). Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science 1*, 64–69.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review 100*, 254–278.

Melnik, S., Molina-Garcia, H., & Rahm, E. (2002). Similarity flooding: A versatile graph matching algorithm and its application to schema matching. In *Proceedings of the International Conference on Data Engineering (ICDE)* (pp. 117–128).

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

Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.

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

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science 12*, 410–430.

Quine, W. V., & Ullian, J. S. (1970). *The Web of Belief*. New York: McGraw-Hill.

Rapaport, W. J. (2002). Holism, conceptual-role semantics, and syntactic semantics. *Minds and Machines 12*, 3–59.

Regier, T. (1996). *The Human Semantic Potential: Spatial Language and Constrained Connectionism*. Cambridge, MA: MIT Press.

Regier, T., & Carlson, L. A. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General 130*, 273–298.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science 27*, 767–780.

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 (Eds.), *Parallel Distributed Processing: Volume 2* (pp. 7–57). Cambridge, MA: MIT Press.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science 9*, 75–112.

Saussure, F. (1915/1959). *Course in general linguistics*. New York: McGraw-Hill.

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

Shepard, R. N., & Cooper, L. A. (1986). *Mental images and their transformations*. Cambridge, MA: MIT Press.

Simmons, K., & Barsalou, L. W. (2003). The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology 20*, 451–486.

Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition 21*, 177–237.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science 12*, 153–156.

Stich, S. P. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, MA: MIT Press.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition 32*, 193–254.

Ullman, S. (1996). *High-Level Vision*. Cambridge, MA: MIT Press.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shape of objects. *Psychological Science 13*, 168–171.