ELSEVIER

Cognition 84 (2002) 295–320

COGNITION

www.elsevier.com/locate/cognit

# Using relations within conceptual systems to translate across conceptual systems

Robert L. Goldstone*, Brian J. Rogosky

*Psychology Department, Indiana University, Bloomington, IN 47405, USA*

Received 2 July 2001; received in revised form 11 March 2002; accepted 15 March 2002

**Abstract**

According to an "external grounding" theory of meaning, a concept's meaning depends on its connection to the external world. By a "conceptual web" account, a concept's meaning depends on its relations to other concepts within the same system. We explore one aspect of meaning, the identification of matching concepts across systems (e.g. people, theories, or cultures). We present a computational algorithm called ABSURDIST (Aligning Between Systems Using Relations Derived Inside Systems for Translation) that uses only within-system similarity relations to find between-system translations. While illustrating the sufficiency of a conceptual web account for translating between systems, simulations of ABSURDIST also indicate powerful synergistic interactions between intrinsic, within-system information and extrinsic information. 

*Keywords*: Concepts; Meaning; Symbol grounding; Conceptual-role semantics; Translation; Neural networks

## 1. Introduction

> "Mr. Martin: I have a little girl, my little daughter, she lives with me, dear lady. She is two years old, has a white eye and a red eye, she is very pretty, and her name is Alice, dear lady.
> Mrs. Martin: What a bizarre coincidence! I, too, have a little girl. She is two years old, has a white eye and a red eye, she is very pretty, and her name is Alice, too, dear sir!
> Mr. Martin: How curious it is and what a coincidence! And bizarre! Perhaps they are the same, dear lady!"
> Eugene Ionesco (1958), "The Bald Soprano"

What gives our concepts their meaning? There have been two major answers to this

* Corresponding author. Tel.: +1-812-855-4853.
*E-mail address:* rgoldsto@indiana.edu (R.L. Goldstone).

0010-0277/02/$ - see front matter 
PII: S0010-0277(02)00053-7

question. The first answer is that concepts' meanings depend on their connection to the external world. By this account, the concept **Dog** means what it does because of its causal connection to dogs in the world. Perceptual processes are critical to this account in that they mediate between the external and internal worlds. **Dog** is characterized by features that are either perceptually given, or can be reduced to features that are perceptually given. This will be called the "external grounding" account of conceptual meaning. The second answer is that concepts' meanings depend on their connections to each other. By this account, **Dog**'s meaning depends on **Cat**, **Domesticated**, and **Loyal**, and in turn, these concepts depend on other concepts, including **Dog**. The dominating metaphor here is of a conceptual web in which concepts all mutually influence each other (Quine & Ullian, 1970). A concept can mean something within a network of other concepts but not by itself, similar to how stability may be a property of a thread within a web but not any thread taken in isolation. This will be called the "conceptual web" account.

Some researchers find the notion of a network of concepts wherein each concept characterizes, and is characterized by, the others as viciously circular. Others embrace the prospect, acknowledging its circularity, but diagnosing it as benign rather than vicious. The agenda of the present paper is three-fold. First, we will provide a brief and partial survey of the motivations for these contrasting perspectives on conceptual meaning. Second, we will focus on one argument against the conceptual web account. According to this argument, the conceptual web account cannot explain how two people can be said to possess the same concept if they have even slightly different conceptual systems. Third, we will present a computational algorithm that is able to find corresponding concepts across people or systems solely on the basis of relations between concepts within a person or system. A quantitative assessment of the algorithm will evaluate the robustness of the algorithm and its sensitivity to various parameters governing a conceptual network. Although the mere presence of the algorithm might be taken as evidence in favor of the conceptual web account, a closer examination of the algorithm's performance reveals important interactions between internal and external characterizations of a concept. Given these interactions, our final conclusion will be that the conceptual web and external grounding accounts of meaning are not only compatible with each other, but that they strengthen one another.

## 2. The conceptual web

The notion that the meaning of a concept depends on the other concepts within a system has been highly influential in all of the major fields that comprise cognitive science: linguistics, computer science, psychology, and philosophy. Representative theories in each of these fields will be described, although many other candidates could have easily been chosen.

In a standard linguistic treatment of concepts, concepts are defined or characterized in terms of other concepts. 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. 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.

In psychology, researchers have argued that concepts are frequently characterized by their associative relations to other concepts. Barr and Caplan (1987) provide evidence, by having subjects list features associated with words, that many concepts are characterized by what they call "extrinsic features", features that are "represented as the relationship between two or more entities" (p. 398).[1] An extrinsic feature of **Hammer** is that it is "used to strike nails". This feature makes recourse to an object other than hammer. If this feature is part of one's natural concept of **Hammer**, then one cannot possess **Hammer** without also possessing **Nail**. Goldstone (1993, 1996) presented empirical evidence that concepts are often interrelated in the sense that each simultaneously acquired concept not only influences how often each concept is used as a label for a presented stimulus, but also influences the absolute representation of each concept. In particular, when two concepts are highly interrelated (e.g. when they are presented in close temporal proximity to each other, or when their labels are similar), there is a tendency for people to create representations of the concepts that are systematically distorted away from each other. This is accomplished by de-emphasizing features that are possessed by both concepts, and by encoding caricatured rather than veridical dimension values for the concepts (see also Goldstone, 1995).

[1] We will use the term "extrinsic" to refer to the dependency of a concept on information outside of the system of concepts, rather than Barr and Caplan's use of the term to refer to inter-conceptual dependencies.

Finally, in computer science, semantic networks have been a major approach to knowledge representation. In these 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. Lenat and Feigenbaum (1991) have argued that inter-conceptual linkages are sufficient for establishing conceptual meanings even without any external grounding of the concepts: "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). 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 the word to other words.

A claim common to these briefly considered theories in linguistics, philosophy, psychology, and computer science is that concepts can only be understood once an entire system of interrelated concepts has been acquired. The concept **Strike** from baseball depends on concepts such as **Batter**, **Ball**, **Strike Zone**, and **Swing**, and these concepts themselves depend on other baseball concepts. Understanding a psychologist's notion of a **Conditioned Response** is possible only after a theory of stimulus–response association is learned. Until then, the definition "a conditioned response is behavior that is elicited when a neutral cue is presented that has been paired with a positive or negative reinforcer" will be of little help in teaching the concept.

## 3. Externally grounded concepts

Although the notion that concepts gain their meaning by their relations to other concepts has been popular in cognitive science, it is not without its detractors. Some have argued that the role of a concept within a network is insufficient to establish its meaning. The concept 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)

This skepticism with the sufficiency of conceptual web accounts has several earlier precedents in philosophy. 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 argues 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. Viewing sensory information as the ultimate ground for our concepts and beliefs is commonplace in philosophy. For example, Quine and Ullian (1970) argue "Thus the ultimate evidence that our whole system of beliefs has to answer up to consists strictly of our own direct observations – including our observations of our notes and of other people's reports" (p. 21).

In psychology, the importance of conceptual meanings that are grounded in something other than other concepts has recently manifested itself in a call for perceptually-based concepts. In the Barsalou's (1999) theory of perceptual symbol systems, concepts are not amodal, completely abstracted symbols, but rather are intrinsically perceptually based. He finds that detailed perceptual information is represented in concepts and that this information is used when reasoning about those concepts. 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. This research, together with research on the bi-directional influences between our concepts and perceptions (Goldstone, Steyvers, Spencer-Smith, & Kersten, 2000; Schyns, Goldstone, & Thibaut, 1998), suggests that apparently high-level conceptual knowledge and low-level perception may be more closely related than traditionally thought.

In computer science, there is a growing interest in conceptual systems that are grounded in embodied systems (Brooks, 1991). Researchers in robotics and artificial life have argued that the concepts that an agent learns should be grounded in the agent's perceptual and motor systems (Almasi & Sporns, 2001). By embodying a neural architecture in a real device, the capabilities and adaptability of the system are augmented. Part of the rationale for the embodied systems research program seems to be a mistrust with purely symbolic representations that are disconnected from the physical environment and the organism's bodily affordances.

Thus far, we have used externally grounded concepts to mean those that are connected to the world via our senses. However, there is a second, philosophical use of external meaning to refer to meaning that is causally connected to the external world without necessarily being mediated through the senses. In this tradition, both the perceptual and conceptual components to meaning would be considered internal because they are centered in a single person rather than world (Block, 1986; Miller & Johnson-Laird, 1976). The famous Putnam (1973) "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 difference between these two interpretations of "external" thus hinges on whether perceptual components are considered to be within or outside the boundary of the internal system. A convincing argument has been made for a graded rather than clear-cut boundary around a system because the same device can be interpreted as being part of an organism's perceptual system or as a transducer external to the system (Clark & Chalmers, 1998). Consistent with this desire to dissolve the boundary around a system, Harman (1987) has proposed a conceptual role semantics account in which the role that a concept plays is construed widely to include its relations to external objects as well as its relations to other concepts. In the following discussion, we will interpret external grounding to refer to any component of conceptual meaning that does not depend on other concepts. This definition coincides with our modeling work in which external grounding will be instantiated as any source of information that is external to the conceptual web, and may include perceptual processes, labels, or teacher signals. In adopting this narrower construal of external grounding, we remain silent on whether or how the world could causally connect to our conceptual systems in a manner not reducible to its impact as mediated through our perceptual and motor systems.

## 4. Translation across conceptual systems

The goal of this article is to argue for the synergistic integration of conceptual web and externally grounded accounts of conceptual meaning. However, in pursuing this argument, we will first argue for the sufficiency of the conceptual web account for a particular task associated with conceptual meaning. Then, we will show how the conceptual web account can be ably 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. Consider two individuals, Joan and John, who each possesses a number of concepts. Suppose further that we would like some way to tell that Joan and John both have a concept of, say, **Mushroom**. Joan and John may not have exactly the same concept of **Mushroom**. John may believe mushrooms grow from seeds whereas Joan believes they grow from spores. More generally, Joan and John will differ in the rest of their conceptual networks because of their different experiences and levels of expertise. Still, it seems desirable to say that Joan's and John's **Mushroom** concepts correspond to one another. We will describe a network that translates between concepts in two systems, placing, for example, Joan's and John's **Mushroom** concepts in correspondence with each other.

Translation across systems is generally desirable and specifically necessary in order to say things like "John's concept of mushrooms is less informed than Joan's". The existence of this kind of translation has been taken as a challenge to conceptual web accounts of meaning. Fodor and Lepore (1992) offer an extended critical examination of "translation

holism", by which they mean the view that nothing can translate a concept from a system L unless it belongs to a system containing many concepts that are translations of many concepts of L. To take the Kuhn (1962) example, translation holism asserts the impossibility that Lavoisier's notion of oxygen can translate into a pre-Lavoisier chemistry simply by creating a corresponding term in the pre-Lavoisier chemistry. The only way for **oxygen** to have a corresponding concept would be to generate many terms in this pre-Lavoisier chemistry that correspond to concepts in Lavoisier's chemistry.

Cross-system translation's challenge to conceptual web accounts, by Fodor and Lepore's interpretation, is that if a concept's meaning depends on its role within the larger system, and if there are some differences between the systems, then the meanings of the concepts in the two systems would be different. They write:

> But now suppose that holism is true about thought content. Then, since you and I surely have widely different belief systems (think of all the things you know that I don't) and since, by definition, a property is holistic only if nothing has it unless many other things do, it may well turn out that none of your thought has the property of bearing T* to any of mine. [T* is the property which a belief has if and only if it expresses a proposition that is the content of some belief of mine]. It would follow that that not more than one of us ever has thoughts about color or thoughts about red. (p. 14)

We are left in the position of not being able to tell that Joan's and John's **Mushroom** concepts correspond to each other. This result is bad enough when considering the two systems to be different people or different scientific theories, but is devastating when one considers the two systems to be the same person at two different times.

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 (Fodor, 1998; Fodor & Lepore, 1992) insists that the notion of similarity is not adequate to establish that Joan and John both possess a **Mushroom** concept. Fodor (1998) considers a situation where you and I have some shared beliefs about GW (George Washington), but some different beliefs as well:

> The similarity of our GW concepts is thus some (presumably weighted) function of the number of propositions about him that we both believe […] But the question now arises: what about the shared beliefs themselves; are they or aren't they *literally* shared? This poses a dilemma for the similarity theorist that is, as far as I can see, unavoidable. If he says that our agreed upon beliefs about GW are literally shared, then he hasn't managed to do what he promised; viz. introduce a notion of similarity of content that dispenses with a robust notion of publicity [a notion that requires identity of beliefs]. But if he says that the agreed beliefs aren't literally shared (viz. that they are only required to be similar), then his account of content similarity begs the very question it was supposed to answer: his way of 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. (pp. 31–32)

Fodor (1998) goes on to argue that all approaches in cognitive science that attempt to determine identical concepts across individuals by measuring conceptual similarities are

misguided. For example, if concepts are assumed to consist of sets of features, then two people's concept of **bachelor** may vary in similarity depending on how many features they have in common. However, Fodor argues that identity of *features* is required of this account in order to determine how many features two person's concepts share. Again, the claim is that conceptual similarity in fact assumes a notion of identity. This identity problem also holds for multidimensional scaling notions of similarity, and for similarity in terms of strength of beliefs. In situations where conceptual similarity appears to explain how concepts are placed in correspondence across people "it really does seem to be *identity* of belief content that's needed here. If our respective beliefs […] were supposed to be merely *similar*, circularity would ensue: since content similarity is the notion we are trying to explicate, it mustn't be among the notions that the explication presupposed (I think I may have mentioned that before)" (p. 33).

We will now present a simple neural network called ABSURDIST (Aligning Between Systems Using Relations Derived Inside Systems for Translation) 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 inter-conceptual 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. ABSURDIST will take 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 **alligator** concept may be similar to Joan's **crocodile** concept, but the two do not correspond because John's **crocodile** 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.

The existence of ABSURDIST provides evidence against Fodor's argument that similarities between people's concepts are an insufficient basis for determining that two people share an equivalent concept. Moreover, the network also explores the larger issue of whether conceptual meanings can be determined solely on the basis of inter-conceptual similarity relations. To avoid potential misunderstandings, four 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 Joan's **mushroom** concept corresponds to John's **mushroom** 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.

Second, our 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, & Gentner, 1990, 1993), using many kinds of similarity (Goldstone, 1994a), 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.

The third disclaimer is that ABSURDIST is hardly a complete model of conceptual meaning. 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. However, simply determining that concepts are equivalent across systems does not tell us what the concepts *mean*, as is made abundantly clear by the intentionally impoverished conceptual representations of ABSURDIST.

Fourth, our intention is not to describe a human simulation of translation, conceptual alignment, or analogy (e.g. Falkenhainer, Forbus, & Gentner, 1989; Hofstadter, 1995; Hummel & Holyoak, 1997). ABSURDIST finds correspondences between concepts across systems, and would not typically be housed in any one of the systems. The exception to this would be if a system was interested in finding translations between entities in two distinct subsystems within it. In this case, the algorithm could be considered a model of human conceptual alignment, albeit one that uses a much simpler representation than the models cited above. In other simulations of human conceptual alignment, such as SME (Falkenhainer et al., 1989), SIAM (Goldstone, 1994b), LISA (Hummel & Holyoak, 1997), Drama (Eliasmith & Thagard, 2001), and ACME (Holyoak & Thagard, 1989), the concepts themselves are richly structured in terms of hierarchical feature sets, propositions, or attribute/value sets. 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, Hund, 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. Apart from this practical benefit, the theoretical contribution of ABSURDIST is to show that it is possible to find correspondences across systems even when the entities within a system are completely defined by their relations to other entities within the same system. With the SME, SIAM, LISA, and ACME systems, the representations of concepts include both relations to other concepts within the domain and stand-alone properties.

If the primary interpretation of ABSURDIST 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. 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.

## 5. The ABSURDIST algorithm

Elements $A_{1,\ldots m}$ belong to System $A$, while elements $B_{1,\ldots n}$ 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 $q$th element of $A$ and the $x$th element of $B$. There will be $m \times n$ 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:

$$\text{if } N(C_t(A_q,B_x)) \geq 0 \text{ then } C_{t+1}(A_q,B_x) = C_t(A_q,B_x) + N(C_t(A_q,B_x))(\max - C_t(A_q,B_x))L$$

$$\text{else } C_{t+1}(A_q,B_x) = C_t(A_q,B_x) + N(C_t(A_q,B_x))(C_t(A_q,B_x) - \min)L \quad (1)$$

If $N(C_t(A_q,B_x))$, the net input to a unit that links the $q$th element of $A$ and the $x$th element of $B$, is positive, then the unit's activation will increase as a function of the net input, a squashing function that limits activation to an upper bound of $\max = 1$, and a learning rate $L$ (set to 1). If the net input is negative, then activations are limited by a lower bound of $\min = 0$. The net input is defined as

$$N(C_t,A_q,B_x) = \alpha E(A_q,B_x) + \beta R(A_q,B_x) - \chi I(A_q,B_x) \quad (2)$$

where the $E$ term is the external similarity between $A_q$ and $B_x$, $R$ is their internal similarity, $I$ is the inhibition to placing $A_q$ and $B_x$ into correspondence that is supplied by other developing correspondence units, and $\alpha + \beta + \chi = 1$. 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

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

where $D(A_q,A_r)$ is the psychological distance between elements $q$ and $r$ in System $A$, and $S(E,F)$ is the similarity between distances $E$ and $F$, and is defined as

$$S(E, F) = e^{-|E-F|}$$

The amount of inhibition is given by

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

These equations instantiate a fairly standard constraint satisfaction network, with one twist. According to the equation for $R$, elements $q$ and $x$ will tend to be placed into correspondence to the extent that they enter into similar similarity relations with other elements. For example, in Fig. 1, $q$ has a distance of 7 to one element ($r$) and a distance of 9 to another element ($s$) within its System $A$. These are similar to the distances that $x$ has to the other elements in System $B$, and accordingly there should be a tendency to place $q$ in correspondence with $x$. 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 $q$ and $x$, 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 one-to-one correspondences exist between systems.

The inhibitory $I$ term 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 $A$.

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, Eq. (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 the fixed number of iterations have been completed. Thus, the network gives as output a complete set of proposed correspondences/translations between Systems $A$ and $B$.

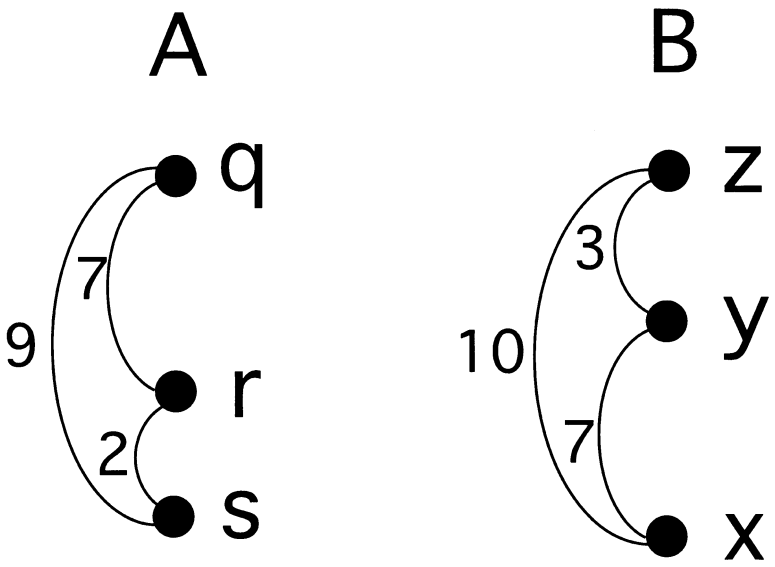


Fig. 1. An example of the input to ABSURDIST. Two systems, *A* and *B*, are each represented solely in terms of the distances/dissimilarities between elements within a system. The correct output from ABSURDIST would be a cross-system translation in which element *q* was placed in correspondence with *x*, *r* with *y*, and *s* with *z*. Arcs are labeled with the distances between the elements connected by the arcs.

## 6. An empirical assessment of ABSURDIST's performance

The general form of the ABSURDIST model presented above does not constrain relations between elements within a system, the $D(x,y)$ values. The network takes two dissimilarity matrices as input, and the elements of these matrices can assume any non-negative values. However, in assessing ABSURDIST's performance, it will be helpful to assume that the conceptual dissimilarities obey metric assumptions, and are interpretable as distances between concepts lying in a geometric space. 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 distances between every pair of elements within a system are computed by

$$D(x,y) = \left[ \sum_{n=1}^{N} \left| V_{n,x} - V_{n,y} \right|^{r} \right]^{\frac{1}{r}}$$

where $V_{n,x}$ is the value of element $x$ on dimension $n$, and $r$ is a parameter that specifies the kind of Minkowski distance metric used ($r = 1$ for City-block distance, $r = 2$ for Eucli-

dean). Then, Eq. (1) is used to update correspondences across the two systems for a fixed number of iterations. The correspondences computed by ABSURDIST 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*.

### *6.1. 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 $7 \times 6$ factorial combination of simulations, with 7 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, 0.1, 0.2, 0.3, 0.4, or 0.5% 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 4000 iterations of Eq. (1). The Minkowski *r* value was set to 2. $\alpha$ was set to 0 (no external information was used to determine correspondences), $\beta$ was set to 0.4, and $\chi$ was set to 0.6. The values for $\beta$ and $\chi$ were selected because they were the most balanced weights that produced fewer than 5% two-to-one correspondences. For each of the 30 combinations of noise and number of items, 1000 separate randomized starting configurations were tested. The results from this simulation are shown in Fig. 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 at least modest amounts of noise.

More surprisingly, Fig. 2 also shows that the algorithm's ability to recover true correspondences generally increases as a function of the number of elements in each system, at least for small levels of noise. One might have thought that as more elements were matched between systems there would be greater confusion between elements, given that the size of the bounding region remains constant. In fact, at a noise level where the probability of correctly translating all elements for three-element systems is less than 50% (noise = 0.3%), completely correct translations for five-element and 15-element systems are found 74 and 92% of the time, respectively. 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. In the same way that more reliable multidimensional scaling solutions are found as the number of related points increases, so does the ability to identify a point on the basis of its relations to other points in the same system. 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 $1/N!$. Thus, with 0.6% noise, the 23% rate of recovering all three correspondences for a three-item system is slightly above chance performance of 16.67%. However, with the same amount of noise, the 17% rate of

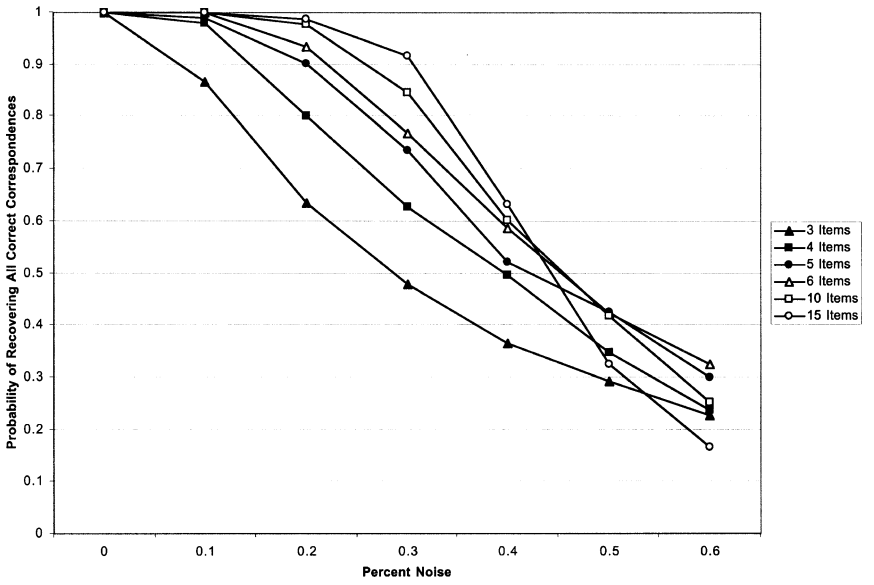


Fig. 2. Probability of correctly translating every element in one system to every element in a second system, as a function of the number of items per system, and the amount of noise with which the elements of the second system are displaced relative to their positions in the first system. For this simulation, the number of dimensions defining each element is 2, $r = 2$, number of iterations = 4000, and $\beta = 0.4$. In this graph, as well as all others, standard error bars are smaller than the height of the legend symbols.

recovering all of the correspondences for a 15-item system is remarkably higher than the chance rate of $7.6 \times 10^{-13}$. Thus, at least in our highly simplified domain, we have support for the argument of Lenat and Feigenbaum (1991) that establishing meanings on the basis of within-system relations becomes easier, not harder, as the size of the system increases.

The measure of translation accuracy shown in Fig. 2 is a conservative measure of performance because properly aligning 14 out of 15 items, for example, would be counted as a failure rather than success. Fig. 3 provides more detailed information on the distribution of partially correct and fully correct alignments. Fig. 3 graphs the frequency, over the 1000 tests, of obtaining a given percentage of correct correspondences for different items. This graph reveals that partially correct translations are rarely obtained. With relatively few exceptions, either ABSURDIST finds all of the correct correspondences, or finds none. The reason why 0% of correspondences are found more frequently than would be predicted by chance responding is that on these trials no cross-system correspondence receives activation above 0.55. Fig. 3 indicates that if some concepts are correctly translated, then all concepts are likely to be correctly translated. This is, once again, due to the cooperative, synergistic nature of the algorithm for determining correspondences.

In evaluating the efficiency of the algorithm, it is useful to know how quickly it converges to good solutions. Fig. 4 plots the probability of finding a completely correct

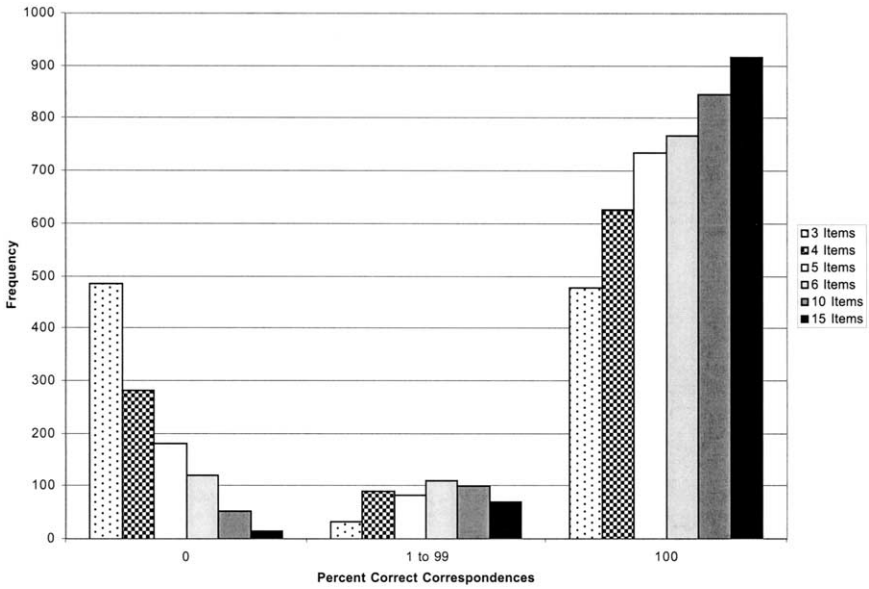


Fig. 3. Frequency distributions, out of 1000 tests, associated with different numbers of items and different percentages of correctly translated items. The number of dimensions defining each element is 2, noise = 0.3%, $r = 2$, number of iterations = 4000, and $\beta = 0.4$.

translation as a function of the number of items per system and the number of iterations of Eq. (1). By 4000 iterations, ABSURDIST's performance has attained nearly asymptotic levels, and reasonably good levels of performance are found with 1000 and 2000 iterations. One attractive feature of the algorithm shown in Fig. 4 is that the number of iterations required for good performance is not appreciably affected by the number of items per system. However, the number of network units required by the algorithm does increase as a quadratic function of the number of items per system because $N^2$ correspondence units are required for aligning two systems with $N$ items per system.

Fig. 2 shows the robustness of ABSURDIST in the face of noise resulting from displacements of elements across systems. As applied to conceptual systems, this corresponds to two people having corresponding concepts, but having somewhat different knowledge associated with the concepts. A more challenging situation arises if people do not have the same set of concepts at all. One possibility is that one system has more concepts than the other. When different-sized systems are compared, ABSURDIST's correspondences are still typically one-to-one, but not all elements of the larger system are placed in correspondence. This situation is shown in Fig. 5A, in which System *A* has three elements and System *B* has seven elements, three of which are arranged in the same configuration as those in System *A*. Given the parameters used thus far, ABSURDIST correctly places the elements from System *A* into correspondence with the three-element pattern contained within the larger seven-element pattern. In this fashion, ABSURDIST provides an algorithm for finding patterns concealed within larger contexts.

A particularly challenging situation for ABSURDIST occurs if two systems have the

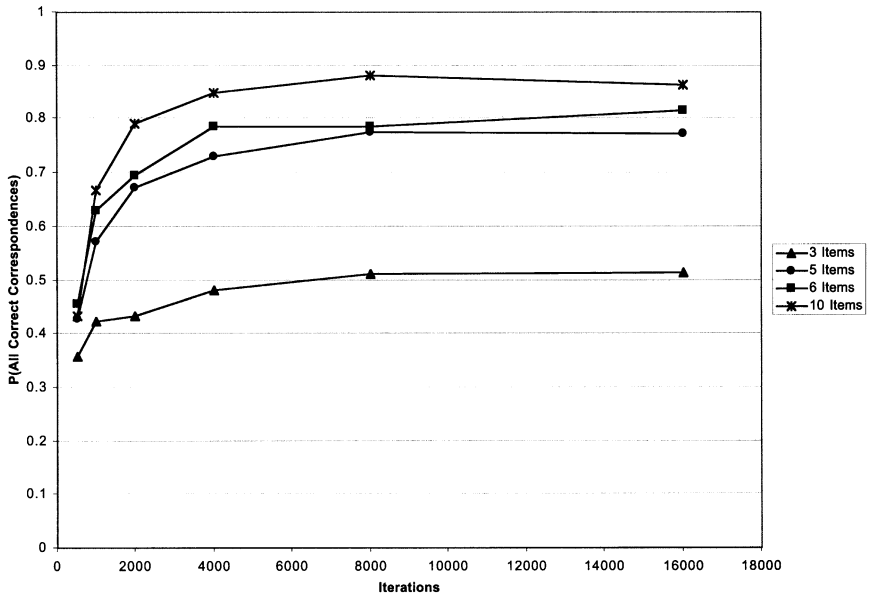


Fig. 4. Probability of ABSURDIST achieving a perfect translation between two systems, as a function of the number of iterations. Noise = 0.3%, number of dimensions defining each element = 2, number of tests per point = 1000, $r = 2$, and $\beta = 0.4$.

same number of elements, but only a subset of them properly matches. For example, Joan and John both have concepts of **Mushroom**, **Fungus**, and **Spores**, but only Joan has a concept of **Truffle** and only John has a concept of **Morel**. This situation is implemented in Fig. 5B 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 Fig. 5B and the previously used parameter values for $\alpha$, $\beta$, and $\chi$, 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.

In sum, we have considered three ways of modeling what it means for people to have different concepts. First, similarity relations among concepts may be different. Second, one person may possess more concepts than another person. Third, each person may have concepts that are unknown to the other person. In each of these cases, ABSURDIST can translate between people and determine which concepts have corresponding concepts in the other person and which concepts are untranslatable.

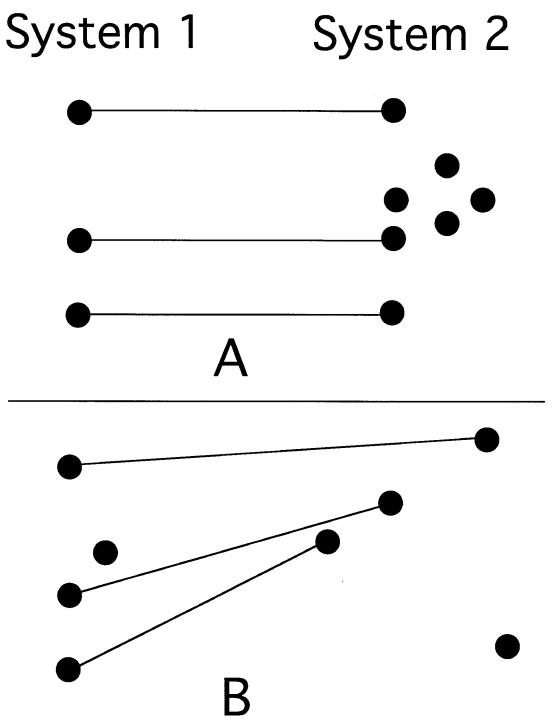


Fig. 5. Two examples of ABSURDIST translating only parts of a system, with typical alignments shown by a solid line connecting elements across the systems. (A) The pattern represented by System 1 is aligned with the subsystem of System 2 that optimally matches this pattern. (B) The three pairs of elements 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.

### 6.2. Indirect similarity relations

In ABSURDIST (when $\alpha = 0$), the cross-system correspondence between two elements is based on their within-system similarity relations. However, if two elements within a system enter into the same set of similarity relations, they still may be disambiguated. This point is clarified by the systems shown in Fig. 6. In System 1, there are two elements, *A* and *E*, that have the same set of dissimilarities, albeit reordered, to the other elements in System 1. That is, both *A* and *E* have distances of 187, 333, 278, and 400 to the other four units of System 1. System 2 is a rotation of System 1 in which *A* becomes *V*, *B* becomes *W*, and so on. Given that *A* and *E* have the same within-system distance relations, one might suspect that deciding whether *A* corresponds to *V* or *Z* of System 2 would be at chance. However, ABSURDIST is able to determine the proper correspondences, shown by the dotted lines, with perfect reliability.

The reason for this successful translation is that the all correspondences are worked out simultaneously, and completely ambiguous correspondences can be disambiguated by

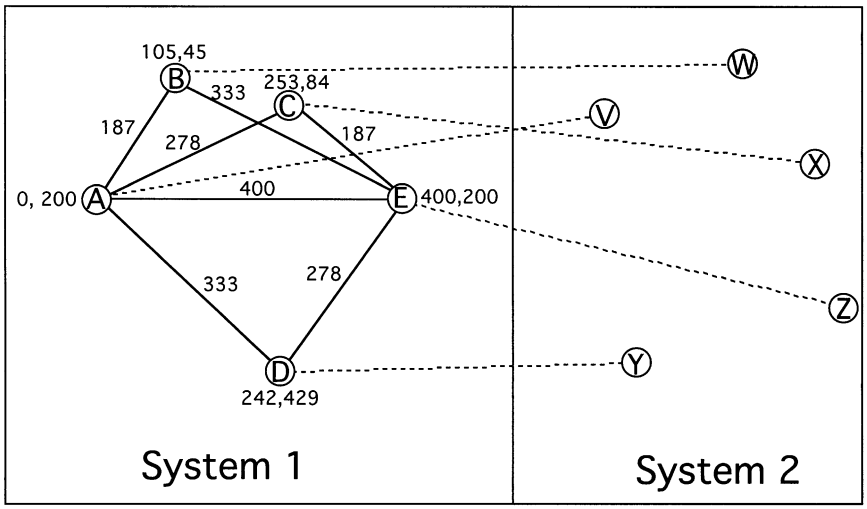


Fig. 6. The horizontal and vertical coordinates for each element of System 1 are shown next to it, and System 1 is rotated clockwise to obtain the elements of System 2. The correct translation (shown by the dashed lines) between Systems 1 and 2 is reliably found by ABSURDIST even though elements *A* and *E* enter into the same four within-system distance relationships. Elements *A* and *V*, and elements *E* and *Z*, are properly aligned because of the role of indirect, within-system similarity relations on translation.

other developing correspondences. Initially, the unit that places *A* into correspondence with *Z* will be just as activated as the unit that places *A* into correspondence with *V*. However, the identical distance between *A* and *B* from System 1 and *X* and *Z* from System 2 will not strengthen the *A*-to-*Z* correspondence much because within-system dissimilarities will indicate that *B* corresponds best to *W*, not *X*. In general, the eventual correspondence strength between two elements will be based not only on their direct similarity relations to other elements, but also on indirect relations among other elements. That is, whether *A* corresponds to *V* depends not only on how similar *A* and *V* are to other elements in their systems, but it also depends, for example, on how similar *B* is to *C*. Analogs of this effect can be found in lexical semantics (Landauer & Dumais, 1997), the interpretation of neural networks (Laakso & Cottrell, 2000), the phenomenology of color perception (Clark, 2000; Palmer, 1999), similarity judgments (Shepard, 1962), and object recognition (Edelman, 1999). In each of these domains, a multi-element, complex system provides many direct *and* indirect constraints that can determine proper translations across systems. This is part of the reason why increasing the number of items per system generally increases rather than decreases the quality of a translation.

### *6.3. Integrating internal and external determinants of conceptual correspondences*

Thus far, translations have been completely based on within-system relations. The simulations have indicated 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. ABSUR-

DIST 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 signal may have provided a hint that the two elements correspond. In these cases, the initial seed value may be significantly greater than 0.5.

Fig. 7 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 0.6% 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 is given a larger initial activation than the other correspondences. When Seed = 1.0, this single correspondence is given even a larger initial activation. Somewhat unsurprisingly, when a true correspondence is given a relatively large initial activation, then ABSURDIST

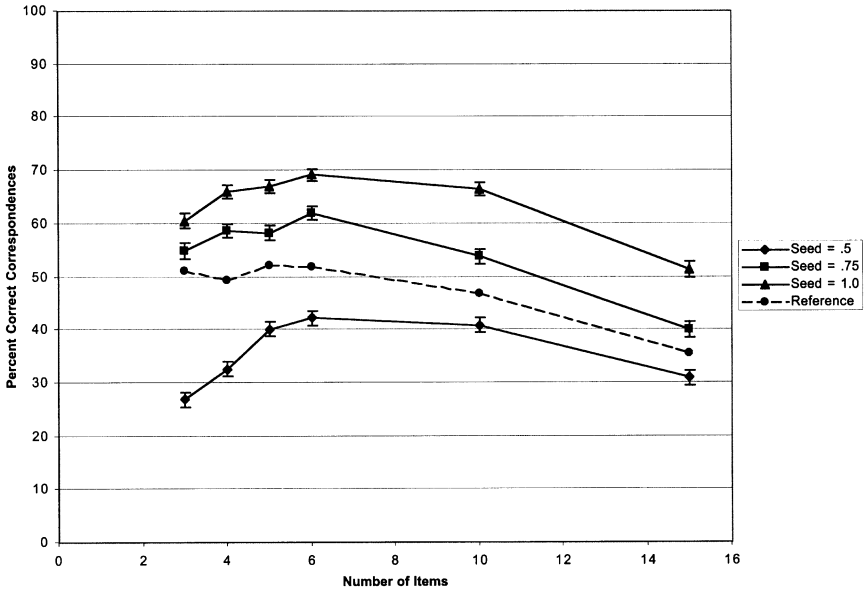


Fig. 7. Percentage of correct alignments found by ABSURDIST, as a function of the number of items per system, and the amount of external bias that seeds a single correct alignment between two elements. As the strength of external bias increases, the percentage of correct correspondences increases, and 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. Noise = 0.6%, number of dimensions defining each element = 2, number of tests per point = 1000, number of iterations = 4000, $r = 2$, and $\beta = 0.4$.

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% $(((0.31 \times 14) + 1)/15)$. The reference line in Fig. 7 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.

Eq. (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 1000 simulations that started with ten 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 Eq. (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 $q$ and $x$ will tend to be strengthened if $q$ and $x$ have similar coordinates. This is extrinsic support because the similarity of $q$'s and $x$'s coordinates can be determined without any reference to other elements. If the two dimensions reflect size and brightness, for example, then for $q$ and $x$ 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 Fig. 8, we set $\alpha = 0.4$, $\beta = 0$, and $\chi = 0.6$. For this group, correspondences are only based on the extrinsic similarity between elements. For the "Intrinsic only" results, we set $\alpha = 0$, $\beta = 0.4$, and $\chi = 0.6$. This group is comparable to the previous simulations in that it uses only a role-based measure of similarity to establish correspondences. Finally, for

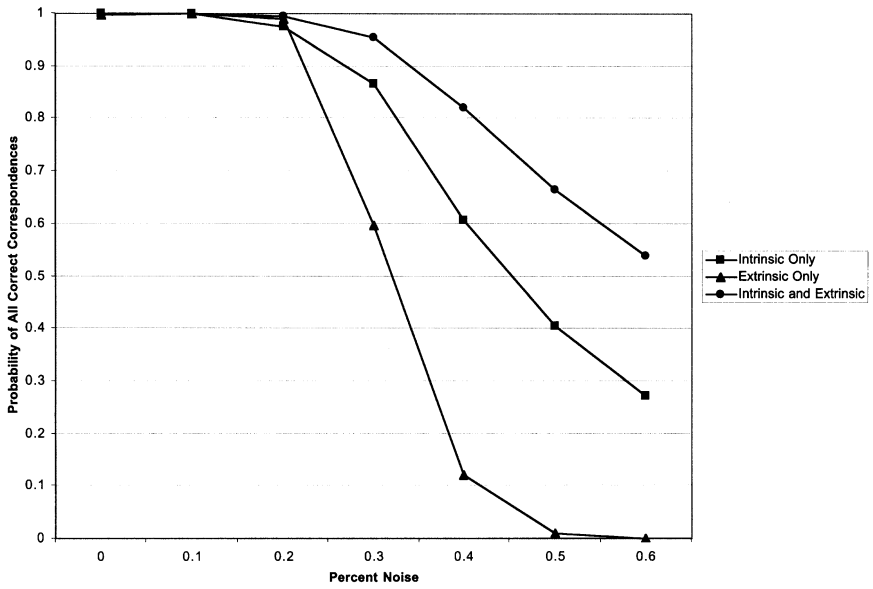


Fig. 8. Probability of ABSURDIST achieving a perfect translation between two systems, as a function of noise, and the weighting of extrinsic and intrinsic information. Better performance is achieved when all weight is given to intrinsic information than when only extrinsic information is used. However, the best performance is achieved when both sources of information are weighted equally. Number of items per system = 10, number of dimensions defining each element = 2, number of tests per point = 1000, number of iterations = 4000, $r = 2$, and $\alpha + \beta = 0.4$.

"Intrinsic and Extrinsic", we set $\alpha = 0.2$, $\beta = 0.2$, and $\chi = 0.6$. 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 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.

Fig. 8 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 Eq. (2). The superior 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. For example, a slight distortion to a point may make its pattern of distances to other points quite similar to another point. This will present difficulties to the intrinsic system, but will not necessarily affect the extrinsic system at all. 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.

## 7. Discussion of simulations

The ABSURDIST model makes two theoretically important points. First, translations *between* two systems can be found using only information about the relations between elements *within* a system. This general claim can be applied to the particular issue of identifying identical concepts across two different people. The claim is that the concept in Person *A* that matches a concept in Person *B* can be found considering only the relations between concepts in Person *A*, and the relations between concepts in Person *B*. ABSURDIST's account of meaning, impoverished though it is, is based solely on the role of a concept within its system (when $\alpha = 0$). ABSURDIST demonstrates how a holistic conception of meaning is compatible with the goal of determining correspondences between concepts across individuals. Two people need not have exactly the same systems, or even the same number of concepts, to create proper conceptual correspondences. Contra Fodor (Fodor, 1998; Fodor & Lepore, 1992), information in the form of inter-conceptual similarities suffices to find inter-system equivalences between concepts. In ABSURDIST, two concepts are treated as matching if the correspondence unit that connects them has an activation greater than a threshold value, and given the positive feedback inherent in the algorithm, correspondence units typically converge rapidly to either 0 or 1.

The simulations identify several specific characteristics of the process of conceptual web-based translation. First, in many cases it is 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 powerful, compensating advantage of many-element systems is that the roles that an object plays within a system are more elaborated and distinctive as the number of elements in the system increases. Second, as exemplified by Fig. 6, an algorithm that uses a concept's role within a system to determine its proper translation can still distinguish between concepts that have the same overall set of relations to other concepts. This is achieved by using indirect relations. The translation for concept *X* is based not only on *X*'s relations, but also on *Y*'s relation to *Z*, assuming that *X*, *Y*, and *Z* belong to the same system. Third, the particular algorithm presented converges relatively quickly on a cross-system translation, and the convergence time does not depend much on the size of systems being aligned. The number of nodes does increase quadratically with the number of elements per system, but this can be reduced by only building correspondence units for alignments that have initial support above a threshold level (Goldstone, 1998), or by using dynamic binding operations to represent correspondences (Hummel & Holyoak, 1997).

The second important 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 first evidence for this is that providing an extrinsic bias to align two concepts improves translation for more than just this pair of concepts. The beneficial ripple-effect for seeding one alignment increases with the number of elements in the system. The relations within a system amplify the effect of extrinsic information. The second evidence is that better alignments are found when both extrinsic and intrinsic information is used than when either source of information is exclusively used, even when the total amount of information is equated.

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 zero-sum. In philosophy, the debate on conceptual meaning has been framed in terms of whether concepts gain their meaning from their role in a system or their external grounding. By this framing, conceptual web accounts of meaning seem opposed to externally grounded accounts. 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 opposition to this zero-sum perspective on intrinsic and extrinsic meaning, ABSURDIST offers a framework in which a concept's meaning is both intrinsic 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. Taking the concept **Car** as an example, we need not claim that **car**'s meaning is completely exhausted by perceptually available data (e.g. a car is composed of tires, seats, and an engine, and tires are composed of wheels and hubcaps, and wheels are composed of …). A concept's meaning may also depend on concepts at the same level of abstraction (**bus** and **truck**), and higher levels of abstraction (**car** is not only characterized *by* **engine**, but it also serves *to* characterize **engine**). Yet, perceptual information, when provided, can be an integral part of the concept. To claim that all concepts in a system depend on all of the other concepts in a system is perfectly compatible with claiming that all of these concepts have a perceptual basis. These two bases of meaning are mutually reinforcing, not mutually exclusive.

## 8. Conclusions

With respect to the application of ABSURDIST to conceptual systems, we agree with Fodor (1998) that concepts should be shareable. An account of concepts needs to provide a way of saying that John's and Joan's **Mushroom** concepts correspond to one another despite their different knowledge about **Mushrooms** and **Tapioca**. Without this correspondence, John and Joan would not be able to achieve communicative contact with one another. They would no longer feel that they are thinking and talking about the same thing. Where we disagree with Fodor is on the question of whether this impression of thinking about the same thing requires literal identity between John's and Joan's concepts. According to Fodor, "to say that two people share a concept (i.e. that they have literally the same concept) is thus to say that they have tokens of literally the same concept type" (p. 28). In

contrast, we have argued that establishing correspondences between concepts is sufficient to determine matching, and hence shared, concepts across systems.

The advantage of accounting for shared concepts in terms of correspondence rather than identity is that one avoids the uncomfortable conclusion that people with demonstrably different knowledge associated with something have the identical concept of that thing. Although the notion of correspondence is less restrictive than identity, it is more constrained than similarity. Many concepts may be similar to each other, but in ABSURDIST one concept in System *A* typically corresponds to at most one concept from System *B*. Unlike the completely graded notion of similarity, correspondences in ABSURDIST become all-or-none after a modest number of iterations. This all-or-none nature of correspondences explains our inclination to say that two people have the same concept, and that slight differences in the persons' knowledge do nothing to affect this claim. Joan's and John's **Mushroom** concepts are placed into complete correspondence with one another even if only Joan knows that mushrooms come from spores. Despite this difference in gradedness between similarity and correspondence, it is nonetheless true that correspondences are determined by similarities between concepts across systems. In turn, the similarity of concepts across systems can be based solely on the concepts' similarities to other concepts within their system. Even if two systems have different relations between corresponding concepts (Figs. 1 and 2), different numbers of concepts (Fig. 5A), or a subset of concepts with no correspondences (Fig. 5B), it is often still possible to recover the correct translation between the conceptual systems using this completely within-system relational information.

Our claim is not that translation between large systems should or typically does proceed using only within-system relations. To the contrary, our simulations point out the power of combining intrinsic, within-system relations and extrinsic grounding. The simulations that did involve only within-system relations indicate that relations intrinsic to a system are an effective component for identifying and translating elements within the system, and that this efficacy does not require extrinsic grounding. Conceptual web accounts of meaning can offer an account of some aspects of meaning, even though they are most effective when combined with an externally grounded component. Thus, a system in which the meanings of its elements all depend upon each other is not viciously circular. A system's elements do not need to be grounded in something outside of the system for proper correspondences between the system's elements and elements outside of the system to be formed. The notion that the meaning of an element within a system, and a component of its meaning that transcends the system, can emerge from its relations to other elements in the system need not be an ABSURDIST fantasy.

## Acknowledgements

The authors wish to express thanks to Eric Dietrich, Shimon Edelman, Gary Cottrell, John Hummel, Michael Lynch, Arthur Markman, Robert Nosofsky, Richard Shiffrin, Mark Steyvers, and three anonymous reviewers for helpful suggestions on this work. This research was funded by NIH grant MH56871, and NSF grant 0125287. Further information about the laboratory can be found at http://cognitrn.psych.indiana.edu/.

## References


Almasi, N., & Sporns, O. (2001). Perceptual invariance and categorization in an embodied model of the visual system. In T. Consi & B. Webb, *Biorobotics*. (pp. 123–143). Menlo Park, CA: AAAI Press/MIT Press.

Barr, R. A., & Caplan, L. J. (1987). Category representations and their implications for category structure. *Memory & Cognition*, *15*, 397–418.

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

Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, *10*, 615–678.

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.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence Journal*, *47*, 139–159.

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.

Clark, A. (2000). *A theory of sentience*. Oxford: Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7–19.

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.

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.

Goldstone, R. L. (1993). Evidence for interrelated and isolated concepts from prototype and caricature classifications. *Proceedings of the fifteenth annual conference of the Cognitive Science Society* (pp. 498–503). Hillsdale, NJ: Lawrence Erlbaum Associates.

Goldstone, R. L. (1994). The role of similarity in categorization: providing a groundwork. *Cognition*, *52*, 125–157.

Goldstone, R. L. (1994). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 3–28.

Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science*, *6*, 298–304.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, *24*, 608–628.

Goldstone, R. L. (1998). Hanging together: a connectionist model of similarity. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 283–325). Mahwah, NJ: Lawrence Erlbaum Associates.

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

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.

Harman, G. (1987). (Non-solipsistic) conceptual role semantics. In E. Lepore (Ed.), *New directions in semantics* (pp. 55–81). London: Academic Press.

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, UK: 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.

Ionesco, E. (1958). *Four plays* (D. M. Allen, Trans.). New York: Grove Press.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Laakso, A., & Cottrell, G. (1998). "How can I know what you think?": assessing representational similarity in neural systems. *Proceedings of the twentieth annual cognitive science conference* (pp. 591–596). Madison, WI: Lawrence Erlbaum.

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

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.

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* [On-line]. Available: http://www.ilt.columbia.edu/Projects/digitexts/locke/understanding/title.html.

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.

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

Palmer, S. E. (1999). Color, consciousness, and the isomorphism constraint. *Behavioral and Brain Sciences*, *22*, 923–989.

Putnam, H. (1973). Meaning and reference. *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.

Saussure, F. (1959). *Course in general linguistics* (W. Baskin, Trans.). New York: McGraw-Hill. (Original work published 1915)

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. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function. Part I. *Psychometrika*, *27*, 125–140.

Stich, S. P. (1983). *From folk psychology to cognitive science: the case against belief*. Cambridge, MA: MIT Press.