COGNITIVE SCIENCE Vol 24 (2) 2000, pp. 299–349 ISSN 0364-0213

Efficient Creativity: Constraint-Guided Conceptual Combination

FINTAN J. COSTELLO
Dublin City University

MARK T. KEANE
National University of Ireland

This paper describes a theory that explains both the creativity and the efficiency of people's conceptual combination. In the constraint theory, conceptual combination is controlled by three constraints of diagnosticity, plausibility, and informativeness. The constraints derive from the pragmatics of communication as applied to compound phrases. The creativity of combination arises because the constraints can be satisfied in many different ways. The constraint theory yields an algorithmic model of the efficiency of combination. The C^3 model admits the full creativity of combination and yet efficiently settles on the best interpretation for a given phrase. The constraint theory explains many empirical regularities in conceptual combination, and makes various empirically verified predictions. In computer simulations of compound phrase interpretation, the C^3 model has produced results in general agreement with people's responses to the same phrases.

I. INTRODUCTION

A fundamental aspect of everyday language comprehension is the interpretation of novel compound phrases through conceptual combination: a cognitive process that is engaged whenever people interpret new phrases like “sand gun,” “cactus fish” or “pet shark.” As a cognitive process, conceptual combination involves an interesting mixture of efficiency and creativity. The efficiency of conceptual combination is apparent in the ease and rapidity with which people can interpret novel compounds (e.g., Gerrig & Murphy, 1992; Murphy, 1990; Potter & Falconer, 1979; Springer & Murphy, 1992). The creativity of conceptual combination is evident in the diversity of interpretations that people produce (e.g., Costello & Keane, 1997a; Hampton, 1987, 1988; Murphy, 1988; Wisniewski &

Direct all correspondence to: Fintan J. Costello, Dublin City University, School of Computer Applications, Glasnevin, Dublin 9, Ireland Fintan.Costello@compapp.dcu.ie.

Gentner, 1991); in the semantic richness of these interpretations, which often include information drawn from world knowledge (e.g., Cohen & Murphy, 1984; Murphy, 1988; Gray & Smith, 1995; Hampton, 1987); and in the polysemy of novel phrases, which can often evoke a number of alternative interpretations (Costello & Keane, 1997a). In this paper, we present a computational-level theory (the constraint theory) and an algorithmic-level model (the C^3 model) that together attempt to capture the creativity and efficiency of the conceptual combination process (for more on meta-theoretical frameworks see Marr, 1982; Palmer, 1989; Keane, Ledgeway & Duff, 1994).

Our computational-level theory describes conceptual combination as controlled by three informational constraints of diagnosticity, plausibility and informativeness. In this theory the best interpretation for a given compound phrase is that interpretation that best satisfies these three constraints. These constraints explain why conceptual combination is creative: the diversity, richness, and polysemy of compound phrase interpretations arises because the constraints can be satisfied in different ways. At the algorithmic level, we propose one possible instantiation of this theory, the *C*onstraint-guided *C*onceptual *C*ombination model (C^3 model) designed to capture the efficiency of conceptual combination. Any given novel noun-noun combination can have a very large number of possible interpretations (potentially infinite according to Kay & Zimmer, 1976) but people only produce a small subset of these interpretations. An adequate model of conceptual combination must provide a tractable procedure that can tame these exponential possibilities to produce a small subset of good interpretations. The C^3 algorithmic model implements the constraints proposed by constraint theory to efficiently construct the best interpretations for novel noun-noun compounds.

In the course of this paper, we describe the constraint theory and show how it is supported by the available psychological and computational evidence. From a psychological perspective, we show how the theory accounts for many of the empirical regularities that have been found in the literature on conceptual combination. We also describe a number of specific predictions the theory makes, and show how these have been empirically verified. From a computational perspective we use the C^3 model to show that the theory can be formulated as an effective procedure that is computationally tractable. We further show that in computer simulations of the interpretation of a set of novel compound phrases, the model produces results that generally agree with people's responses to the same phrases.

This paper has six main sections. First, the scene is set by describing the empirical regularities that have a bearing on the creativity of conceptual combination. Second, the constraint theory of combination is described. Third, the C^3 model is described, and the basis for its efficient computation of combined concepts is explained. Fourth, supporting evidence for the theory is given, in terms of its account for empirical regularities, its specific verified predictions, and its computer simulation results. Fifth, constraint theory and the C^3 model are compared with two other theories of conceptual combination. Finally, the paper concludes with a discussion of the work and its implications for the wider issue of compositionality in language and thought.

II. THE CREATIVITY OF CONCEPTUAL COMBINATION

It is important to place our proposals about the creativity of conceptual combination in a wider context. First, we should be clear that, in everyday discourse settings, people's interpretation of novel phrases by conceptual combination often involves two processes: first, the construction of a combined concept, and then the use of that constructed concept to find an existing referent in the preceding discourse context. For example, interpreting the phrase "leg screws" in the context of instructions for assembling a table would involve first constructing a combined concept such as "*leg screws* are screws used to affix table legs" and then using the constructed concept to find the particular screws referred to, which may have already been identified in some other way. Reference to discourse context clearly has an influence on compound phrase interpretation (in another context, "leg screws" might refer to screws with a leg shape, or perhaps a pivoting joint). However, the process of constructing a combined concept is equally important, both when phrases are presented alone and when they occur in discourse contexts (see Gerrig & Murphy, 1992, for evidence). Our aim in this paper is solely to give an account of the construction of combined concepts; this is a necessary prelude to a more general account of how combined concepts are constructed and used in context.

Second, we are only considering noun-noun phrases when, in a sense, all possible legal syntactic sequences could be considered to be conceptual combination (e.g., adjective-noun phrases, verb-noun phrases, plural nouns, relative clauses, and so on). We focus on noun-noun combinations because they seem to involve much more semantic change than, for example, adjective-noun phrases (see e.g., Smith, Osherson, Rips, & Keane, 1988).

Third, even within the domain of noun-noun combinations there are types of combination that our theory might find hard to explain: in particular, combinations that are based on the use of metaphor or analogy (see Keane, 1993, 1997; Veale & Keane, 1994, 1997). For example, one writer has referred to the '90s as the "discount decade" (Simmons, 1995), a phrase that makes a metaphorical mapping between the domain of discount pricing and that of dates (1992, 1995, 1999). However, these overtly metaphorical combinations seem to be relatively rare, and the remaining corpus of nonmetaphorical combinations that the theory can address is likely to be large (see the General Discussion for more on metaphor, analogy and conceptual combination).

Given these caveats on our topic of interest, we can outline some empirical regularities in the creativity of conceptual combination under four main headings: diversity of interpretation type; diversity of interpretation focus; semantic richness; and polysemy.

Diversity in Interpretation Type

Syntactically a noun-noun compound has two parts: the modifier (the first word in the phrase) and the head (the second word), and people can combine these constituents in a number of different ways to interpret the compound. Five general types of interpretation have been recognized: relational, property, hybrid, conjunctive, and known-concept interpretations. In relational interpretations a relation is asserted between the concepts being combined, as in “a *horse knife* is a knife for butchering horses” (Cohen & Murphy, 1984; Murphy, 1988). In property interpretations a property of one combining concept is asserted of the other, as in “a *cactus fish* is a prickly fish” (Wisniewski & Gentner, 1991; Wisniewski, 1996). In hybrid interpretations the combined concept is a blend of both concepts being combined, as in “a *drill screwdriver* is two-in-one tool with features of both a drill and a screwdriver” (Wisniewski, 1997a, 1997b). In conjunctive interpretations the combined concept is an instance of both concepts being combined, as in “a *pet fish* is a guppy” (Hampton, 1987, 1988). Finally, known-concept interpretations describe a particular known concept that is related to the concepts being combined, as in “a *cow house* is a byre where cows are kept” (Costello & Keane, 1997a).

Surveys of people’s interpretations of novel compounds suggest that some interpretation types are more common than others (Costello & Keane, 1997a, 1997b; Wisniewski & Gentner, 1991; Wisniewski & Love, 1998). Relational and property interpretations seem to be most frequent with hybrid, conjunctive and known-concept interpretations occurring more rarely. Further, the frequency of property interpretations reliably increases with the similarity of the concepts being combined. Thus a novel compound such as “elephant pig,” with similar constituents, is more likely to evoke a property interpretation (“an *elephant pig* is a large pig”) than a compound such as “elephant car,” with dissimilar constituents (Markman & Wisniewski, 1997; Wisniewski, 1996). The occurrence of different interpretation types is also influenced by various priming effects, with combinations using relations that occur frequently with the combining words being judged sensible faster than combinations using less frequent relations (Gagné & Shoben, 1997), and with combinations more likely to be interpreted using relations (or properties) if relational (or property) interpretations have occurred in the discourse context (Wisniewski & Love, 1998).

The five categories of relational, property, hybrid, conjunctive, and known-concept interpretation do not exhaust the range of interpretation types that people produce: there are many variations on these types, and some interpretations do not fall neatly into one or other of these general categories. At the very least, however, an adequate theory of combination should explain how these five types of interpretation are produced.

Diversity of Interpretation Focus

Often, one part of a compound phrase interpretation has privileged status as the focal concept of the interpretation: that is, the concept that the interpretation is about. For example, the focal concept for the interpretation “a *horse knife* is a knife for butchering horses” is the concept *knife*: the interpretation describes a knife for butchering horses, not a horse that is butchered by a knife. The focal concept usually contributes most semantic information to a combined concept, with the other parts of an interpretation limited to modifying one particular aspect of the focal concept. The focal concept also identifies the general category of which the combined concept is a member; and is typically mentioned first in a verbal description of the combined concept. Finally, when asked to describe a combined concept people usually list properties of the focal concept only, and do not mention properties of other parts of the interpretation. When describing their interpretation for the phrase “horse knife,” for example, people might mention attributes such as LARGE, SHARP, METALLIC, which belong to the focal concept *knife*, but would not mention attributes such as FOUR-LEGGED or ANIMATE, which belong to the *horse* part of the interpretation.

For most compound phrase interpretations the focal concept is simply the concept named by the head word of the phrase (as in “horse knife”). Some compounds, however, do not follow this pattern. Exocentric compounds (Bauer, 1983) have as their focal concept some concept other than the head, as in the familiar compound “seahorse” (“a species of *fish* shaped like a horse”) or the novel compound “jellybean shovel” (“a type of *spoon* for dispensing jellybeans”). Wisniewski (1996) has identified similar cases, which he terms construals. In these construals a compound is taken to refer to a concept related to one of the phrase’s constituents (e.g., “an *artist collector* is a collector of the works of an artist”). In other interpretations the pattern of focus is reversed, with the focal concept being the modifier of the phrase being interpreted. Although these reversals are rare they have been found both in experimental investigations of conceptual combination (Costello, 1997; Costello & Keane, 1997a; Gerrig & Murphy, 1992; Wisniewski & Gentner, 1991), and in everyday discourse. For example, in a Disney Chip-and-Dale story the phrase “slipper bed” was interpreted as “a slipper in which a chipmunk can sleep.” In this interpretation, the modifier “slipper” plays the role of focal concept, contributing most to the new combined concept, best identifying the category in which the combined concept is a member, and being mentioned first in the description. These changes of focus in reversals and in exocentric compounds represent the extension of the usual focal concept (the head) to cover the modifier or some other concept. The challenge for theories of combination is to explain how these changes occur.

Semantic Richness

People’s interpretations for novel noun-noun phrases are often semantically rich, containing detailed knowledge drawn from various, apparently semantically distant, parts of world knowledge (e.g., Gray & Smith, 1995; Kunda, Miller, & Claire, 1990; Medin & Shoben, 1988; Murphy, 1988). The knowledge used to interpret one combination is often completely different to that used to interpret other combinations containing the same concepts. For example, the three interpretations

- a *street knife* it is an easily concealed knife used by muggers and petty criminals,
- a *street flower* is a small weed that grows through cracks in the pavement,
- a *street brush* is a wide tough brush that street-sweepers use,

make use of very different knowledge about the concept *street* and the concepts with which it combines. The richness of compound phrase interpretations is perhaps to be expected, given that the role of compounds in language is to convey a relatively large amount of detailed information in a concise way.

Researchers investigating the semantic richness of conceptual combination have typically emphasized the production of emergent attributes in combination. Emergent attributes for a combined concept are attributes that people rate as typically true for the combined concept but not for its constituents. Hampton (1987), for example, found that conjunctive combinations such as "pets that are also birds" produced emergent attributes such as SMALL, KEPT-IN-CAGE and PRETTY. Others have confirmed this result in several domains (see e.g., Chater, Lyon, & Myers, 1990; Gentner & France, 1988; Kunda, Miller, & Claire, 1990; Murphy, 1988; Tversky & Kahneman, 1983). The existence of these emergent attributes in combinations is often taken as evidence of the noncompositionality of conceptual combination (see General Discussion).

Two possible sources for the emergent semantic richness in conceptual combination are generally recognized. Some emergent information may come from background knowledge in the form of abstract domain theories of the concepts being combined (Hampton, 1991; Murphy, 1988; Rips, 1995). Other information seems to come from specific known instances of the combining concepts (what Hampton, 1988, called *extensional feedback*). Current empirical evidence supports the origin of semantic richness in specific known instances. Gray & Smith (1995), for example, found a high correlation between the production of emergent attributes in combined concepts, and the occurrence of those attributes in specific instances of the combined concept (see also Medin & Shoben, 1988). This influence of specific instances is perhaps to be expected given that the linguistic function of compound phrases is often to select specific subsets of more general categories (for example, "knife" refers to a relatively general category, "steak knife" refers to a specific subset of that category). There is currently little direct evidence for the influence of abstract domain theories on combination. However, this does not mean that such abstract theories are not used; the lack of evidence may simply reflect a difficulty in empirically distinguishing between information derived from abstract theories and that derived from specific instances. Explaining the origins of emergent semantic richness in combined concepts is a major challenge for current theories of conceptual combination.

Polysemy

Given the variation of interpretation type and focus in combination and the breadth of world knowledge accessible by the combination process, the range of alternative interpretations for a given novel compound can be quite broad. This range of alternatives is made explicit in novel noun-noun compounds that have a high degree of polysemy. Polysemous compounds evoke a number of different interpretations, each combining the constituents in a different way and drawing on different sources of knowledge. For example, a *shoe knife* could be "a knife used by cobblers in repairing shoes," "a knife with a broad flat blade shaped like the sole of a shoe" or "a knife which gangsters carry concealed in their shoes" (Costello & Keane, 1997a; see also Kay & Zimmer, 1976).

Unfortunately, the polysemy of conceptual combination has received little systematic study. Murphy (1990) found that the average number of meanings produced by participants to adjective-noun compounds was reliably less than that produced for noun-noun compounds, and the average number of meanings produced for predicating-adjective compounds was reliably less than that produced for nonpredicating-adjective compounds (nonpredicating adjectives, for example "corporate" or "medical," derive their meaning from nouns and are thought to be more complex than predicating adjectives such as "red" or "heavy"). Costello and Keane (1997a; Costello, 1997) in a survey of the polysemy of a large number of novel noun-noun compounds, found that the polysemy of a novel compound was reliably influenced by the types of concept it contained. Compounds containing artifact or superordinate concepts evoked significantly more alternative interpretations than those containing natural-kind or basic-level concepts. A general theory of conceptual combination should be able to explain how these variations in polysemy occur.

III. CONSTRAINT THEORY: EXPLAINING THE CREATIVITY OF CONCEPTUAL COMBINATION

There are three main components to the Constraint theory of conceptual combination: first, some high-level assumptions about the knowledge used to construct combined concepts; second, a statement of pragmatic influences on novel compound interpretation; and third, three constraints—diagnosticity, plausibility, and informativeness—which control the production of compound phrase interpretations.[1] In meta-theoretical terms, our proposal on pragmatics constitutes a computational-level statement of the goals of conceptual combination, whereas the knowledge assumptions and constraints are computational-level descriptions of what needs to be computed in the combination process. At this stage we will not say anything about how compound interpretations are constructed in the theory: this is the preserve of our algorithmic model (although in giving examples to explain the theory, we will necessarily provide hints about how things are done).

Knowledge Available to the Conceptual Combination Process

People make use of a wide variety of knowledge when interpreting novel compound phrases. As such, we assume that the combination process can make use of prototypes of the constituent concepts, specific instances of these concepts, prototypes and instances of related concepts, general domain theories and specific event representations that involve these concepts. In short, the constraint theory proposes that the combination process has direct access to the full contents of memory.[2] This position contrasts with several other theories which propose that the combination process is limited to summary, prototype information, with other types of knowledge only being used to elaborate a combined concept initially constructed from these prototypes (see Hampton, 1991; Murphy, 1988).

To make this point more concrete, consider the graphical representation in Figure 1 of a simplified knowledge base, along with one interpretation for the compound phrase "finger cup" generated from it. This severely limited knowledge base encodes knowledge about instances of the concepts *finger*, *cup*, and *bowl*, although related concepts like *hand* and *liquid* are also mentioned. The concept representations have both properties (e.g., SMALL, TUBULAR, SOLID) and relations (CONTAINS, WASHED-IN) thus giving us all we need to

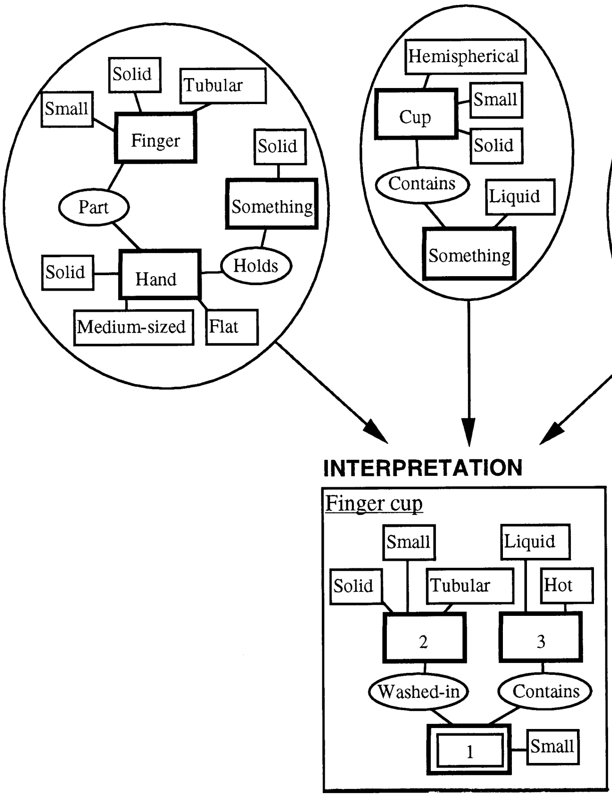

Figure 1. Graphic representation of an example knowledge-base and a possible interpretation for the phrase "finger cup".

represent prototypes, instances, domain theories and events (although the latter two are not shown in this example). The knowledge-base shows how instances of one concept can be related to instances of another via relations holding between roles: for example, the relation CONTAINS holds between the role *bowl* and the role *water*. The graphic representation finesses the detailed representation of slot names and roles simply showing the value of the slot connected to the concept name (i.e., SMALL rather than SIZE[A, SMALL]; see the model section for the actual representations used). This knowledge base has been used to construct one possible interpretation of "finger cup" which can be glossed as "a *finger cup* is something small containing hot liquids used to wash fingers." This example knowledge-base is extremely limited compared to the knowledge-base actually used in simulations, which contained multiple instances of many more concepts. In our example-knowledge base for the interpretation of "finger cup" we include only one other concept (*bowl*) to demonstrate how the theory uses knowledge from other concepts in the interpretation of a compound phrase. We use this example knowledge-base to illustrate our discussion of the theory. We will iterate through this example again at a more detailed level when describing the operation of the model.

The Goal of Conceptual Combination: Pragmatics of Novel Compound Usage

A listener hearing a novel compound phrase can make a number of pragmatic assumptions about the intentions of the speaker who produced that phrase. At the most basic level the listener can assume that the speaker, in uttering the phrase, is trying their best to indicate a particular combined concept. This basic assumption of co-operation leads to a number of more specific inferences that the listener can validly make about that combined concept. First, that the intended combined concept is one which the listener already more-or-less knows (otherwise the speaker would not have used a terse compound but would have described the intended concept in more detail). Second, that the intended combined concept is one best identified by the two words in the phrase (otherwise the speaker would have selected different words). Third, that the intended combined concept is one for which both words in the phrase are necessary and sufficient (otherwise the speaker would have used more or fewer words).

Each of the constraints in our theory instantiates one of these pragmatic assumptions. The plausibility constraint ensures the production of an interpretation describing something more-or-less known to the interpreter. Plausibility requires that acceptable interpretations contain properties that are consistent with prior experience. The diagnosticity constraint ensures the production of an interpretation best identified by the words in the phrase being interpreted. Diagnosticity requires that interpretations contain some properties diagnostic of both constituent concepts in the phrase being interpreted (the diagnostic properties of a concept are those which best identify that concept). Finally, the informativeness constraint ensures that an interpretation conveys the requisite amount of new information such that no more words are needed and no fewer words would suffice to convey that information. To interpret a novel compound correctly the listener constructs an interpretation that best satisfies these three constraints; such an interpretation will meet the pragmatic assumptions and thus be the correct combined concept as intended by the speaker.

What Needs to Be Computed: Constraints on Conceptual Combination

The pragmatics of novel compound use provides the basis for the three constraints—diagnosticity, plausibility, and informativeness—proposed to guide the process of concept combination. In this section, we describe these constraints in detail. When describing the C^3 model, we will iterate through these constraints again, and provide specific formal descriptions of how they are instantiated in the model. If the description of the constraints seems vague at this level, it is important to remember that they are instantiated later by clearly defined effective procedures.

Diagnosticity. When confronted with a novel compound phrase an interpreter can justifiably assume that the correct combined concept (as intended by the speaker) is one best identified by the two words in the phrase (otherwise the speaker would have selected different words). The diagnosticity constraint instantiates this assumption: a combined concept that satisfies the diagnosticity constraint will be one best identified by the words in the phrase being interpreted. Diagnosticity requires that an interpretation contains some predicates that are diagnostic of one of the constituent concepts in the phrase being interpreted, and some predicates that are diagnostic of the other constituent concept. The diagnostic predicates of a concept are those which best identify instances of that concept and differentiate it from other concepts: a predicate is diagnostic for a concept if it occurs frequently in instances of the concept and rarely in instances of other concepts (for related ideas see Tversky's, 1977, diagnosticity and Rosch's, 1978, cue validity[3]). Diagnosticity predicts that the first interpretation should be more acceptable than the second:

- a *cactus fish* is a prickly fish
- a *cactus fish* is a green fish

because PRICKLY is more diagnostic of cacti than GREEN, the latter being a predicate possessed by almost every plant (note that both interpretations implicitly contain diagnostic predicates for the concept *fish*, because fish are mentioned in the interpretations). This constraint does not demand that all the diagnostic predicates of each constituent concept are used in an interpretation; only enough diagnostic predicates to identify an instance of a concept well. Thus the interpretation "a *cactus fish* is a prickly fish" does not contain the diagnostic predicate GROWS-IN-DESERT; the diagnostic predicate PRICKLY is enough to identify the interpretation as describing something that could be justifiably named a "cactus fish."

The diagnosticity constraint applies equally to all interpretation types, with diagnostic predicates occurring in different places in different types of interpretations. In property interpretations diagnostic properties of one concept are asserted of the other, as in the "cactus fish" example. In conjunctive and hybrid interpretations a concept is constructed containing diagnostic predicates of both concepts, as in the interpretation "an *apple pear* is a fruit with the color of an apple and the shape of a pear." Finally, in relational interpretations the diagnostic predicates of the constituent concepts occur implicitly in different parts of the interpretation, joined by a relation, as in the interpretation "a *horse knife* is a knife for butchering horses," where the diagnostic predicates of both *knife* and *horse* are implicitly present.

As we saw earlier, one part of a compound phrase interpretation has the special status of being that interpretation's focal concept. In the constraint theory the focal concept of an interpretation is identified by diagnosticity: the focal concept is that part of an interpretation that possesses diagnostic properties of the head concept of the phrase being interpreted. Note that this does not mean that the focal concept of an interpretation need *be* the head concept of the phrase being interpreted; it need only possess that concept's diagnostic predicates. Thus in head-focus interpretations such as "a *horse knife* is a knife for butchering horses" the focal concept is the head concept *knife* (possessing its usual diagnostic predicates), whereas in exocentric interpretations such as "a *seahorse* is a fish whose head is the shape of a horse's head," the focal concept is some other concept that possesses diagnostic predicates of the head concept (in the "seahorse" case the diagnostic predicate being shape). Similarly, in focus reversals such as "a *slipper bed* is a slipper which a pet chipmunk can sleep in" the focal concept is the modifier, which has been given predicates diagnostic of the head ("used for sleeping in" being diagnostic of the concept *bed*).

In Figure 1 some of the diagnostic predicates of *finger* (i.e., TUBULAR and SMALL) and *cup* (i.e., SIZE and CONTAINS) occur in the "finger cup" interpretation. Because the diagnostic predicates of the two concepts occur in two different roles linked by a relation, the interpretation is a relational one. The focal concept of this interpretation is role 1 because it has the diagnostic predicates of the head word "cup." As we shall see, the other predicates that occur in this interpretation largely derive from the action of the plausibility constraint.

Plausibility. When confronted with a novel compound phrase an interpreter can justifiably assume that the correct combined concept (as intended by the speaker) is one describing something they already more-or-less know (otherwise the speaker would not have used a terse compound but would have described the intended concept in more detail). The plausibility constraint instantiates this assumption by requiring interpretations to contain predicates that are consistent with prior experience. Interpretations that satisfy the plausibility constraint will describe something already more-or-less known to the interpreter. Clearly, interpretations describing instances that are already known to exist are fully plausible; all predicates in them are already known to be consistent. Interpreting "stilt bird" as referring to birds with long legs such as flamingos is highly acceptable because flamingos are known to exist (and possess diagnostic properties of both stilt and bird). In many cases, however, compounds must be interpreted by constructing some novel concept. The acceptability of such interpretations will vary according to the degree to which the properties they contain are consistent with previous knowledge. The plausibility constraint predicts that the first interpretation will be more acceptable than the second:

- a *shovel bird* is a bird with a flat beak it uses to dig for food
- a *shovel bird* is a bird that uses a shovel to dig for food

because the first is much more consistent with what actually occurs in the world, whereas the second needs the support of a special context in which the bird is anthropomorphised (e.g., in a cartoon context).

In the "finger cup" example, the interpretation produced is plausible because it is consistent with knowledge available in the example knowledge-base (see Figure 1). So, the washing interpretation is consistent with known aspects of cups, but more precisely with aspects of bowls. Note that plausibility is determined relative to all the knowledge in the knowledge base, so it can be influenced by concepts that were not explicitly mentioned in original combination (like *bowl*). What is hidden in this account, but which we will expand on more fully in the model section, is how predicates from other concepts can be imported into the interpretation during this stage. The plausibility constraint has an important role in fleshing out interpretations as well as in simply evaluating their acceptability. Others have described related mechanisms for fleshing-out mental models in deductive reasoning (e.g., Byrne & Handley, 1992; Johnson–Laird, Schaeken, & Byrne, 1992; Schaeken, Byrne, & Johnson–Laird, 1995).

Informativeness. When confronted with a novel compound phrase an interpreter can justifiably assume that the correct combined concept (as intended by the speaker) is one for which both words in the compound are necessary and sufficient (otherwise the speaker would have used more or fewer words). The informativeness constraint instantiates this assumption: an interpretation that satisfies informativeness conveys new information such that both words in the phrase being interpreted are necessary and sufficient for that information. The informativeness constraint requires that an interpretation convey new information in comparison to both the modifier concept and the head concept of the phrase being interpreted. The informativeness constraint predicts that both of the following interpretations will be unacceptable

- a *head hat* is a hat worn on the head
- a *car vehicle* is a car

because the first provides no new information relative to the head concept *hat*, whereas the second provides no new information relative to the modifier concept *car*. This account fits with Downing's (1977) finding that people find it difficult to interpret novel compounds (such as "head hat") which have redundant modifiers; that is, where the modifier provides no new information beyond that conveyed by the head concept alone. Indeed, our informativeness constraint can be seen as a more general case of Downing's proposals about the influence of informativeness in compound interpretation; the primary difference being that Downing gives informativeness relative to the head of a phrase an important role in compound interpretation, whereas we address informativeness relative to both the head and the modifier.

In the "finger cup" example in Figure 1, the *washing* interpretation is informative because it conveys new information relative to both the head (*cup*) and modifier (*finger*); relative to *cup* the interpretation contains extra information about containing hot liquid and being used for washing; relative to *finger* the extra information is about being washed. As we shall see in the Model section, informativeness also determines which interpretations are more informative than other interpretations.

Interacting Constraints and Acceptability. The three constraints of diagnosticity, plausibility and informativeness act together to control the relative acceptability of compound phrase interpretations: interpretations that satisfy all constraints well will be good interpretations for the phrase in question, interpretations that satisfy the constraints less well will be less acceptable. The constraints do not contribute equally, however; the constraints of diagnosticity and plausibility together determine the primary acceptability of an interpretation, with informativeness only entering in a logical sense to determine if an interpretation is or is not informative. We describe the constraints and the interactions between them in more detail in the next section, which describes the C^3 model of the efficiency of combination.

IV. THE C^3 MODEL: SIMULATING THE EFFICIENCY OF CONCEPTUAL COMBINATION

When confronted with a novel compound phrase, people can usually produce the correct interpretation rapidly and almost effortlessly by conceptual combination; selecting the correct interpretation type from the range of possible types available, accessing and integrating knowledge from various distant sources, and rejecting possible alternative interpretations (e.g., Potter & Falconer, 1979; Murphy, 1990; Gerrig & Murphy, 1992). In this mental act the conceptual combination mechanism has solved a serious computational problem. Given the creativity and diversity of compound phrase interpretations, the number of potential interpretations for a given phrase may be very large. The efficiency of people's conceptual combination means that the combination mechanism can rapidly extract from this set of potential interpretations the best interpretation for the phrase in question. This efficiency places an important requirement on theories of conceptual combination: they must be computationally tractable, able to produce the best interpretation for a given novel phrase in a reasonable time. At the same time, theories of conceptual combination must admit the full creativity of combination, and should not a priori exclude any potential interpretations from consideration in the search for the best interpretation for the phrase. The C^3 model provides one way in which this requirement of efficiency can be met: through the use of a constrained search of the space of possible interpretations.

Constrained search in the C^3 model

The problem for an algorithmic model of conceptual combination is to efficiently extract the best interpretation for a given phrase from the set of all potential interpretations for that phrase. We represent the situation graphically in Figure 2, which shows the potential interpretations for a given phrase as points in interpretation-space, where each interpretation's height corresponds to its acceptability for the phrase in question. An efficient algorithmic model should find the most acceptable interpretation in the space (the highest point in the graph), without having to construct every possible interpretation in the space. An efficient algorithm should find the best interpretation whereas constructing as few potential interpretations as possible.

As Figure 2a shows, the overall acceptability of an interpretation has two components: its diagnosticity and its plausibility (assuming informativeness). Setting aside the details of how diagnosticity and plausibility are computed for the moment, the action of C^3 model can be seen as a constrained search through this space of potential interpretations. The model's search for the best interpretation in interpretation-space begins by constructing those interpretations with the highest degree of diagnosticity and moves through the space in steps of decreasing diagnosticity (see Figure 2b). The plausibility and informativeness constraints control the construction of the most plausible interpretations at each step of diagnosticity. The resulting interpretations will have a high degree of diagnosticity and plausibility, and hence high overall acceptability (with any uninformative interpretations being rejected).

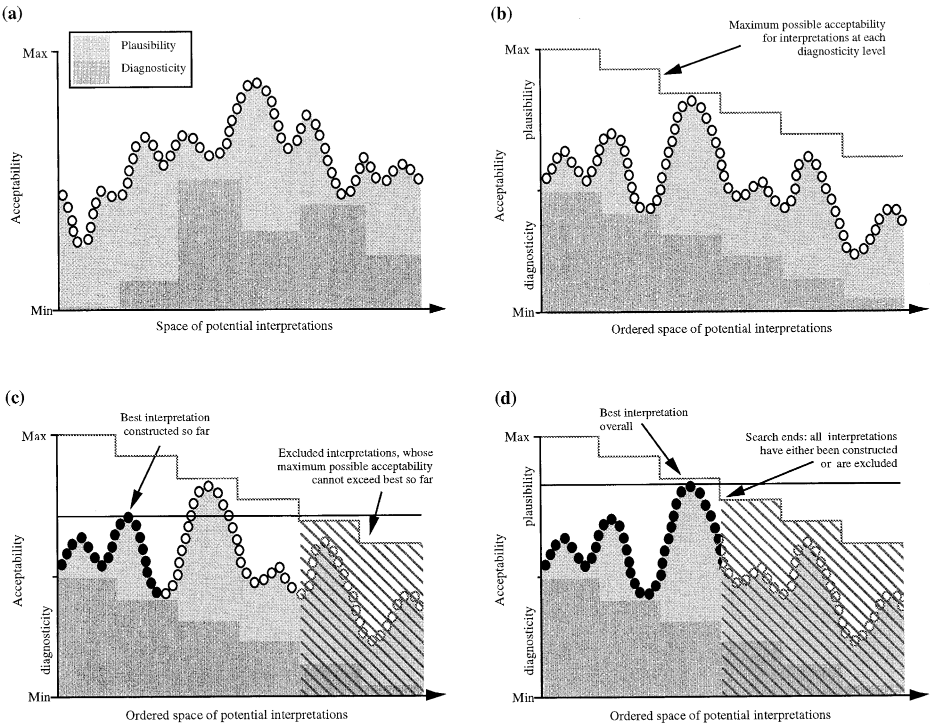

Figure 2. (a) A representation of the set of potential interpretations for a given compound phrase as points in interpretation-space, where each interpretation's height corresponds to its acceptability for the phrase in question. (b) An interpretation-space ordered by the underlying stratum of diagnosticity. As diagnosticity decreases the maximum achievable acceptability decreases also (dotted line). (c) Initial progression of the C^3 model. Some interpretations have been constructed (black circles) whereas others are excluded from consideration. (d) Final state of the C^3 model. A small subset of potential interpretations have been considered (black circles) and the guaranteed best interpretation has been found.

The model constructs the most diagnostic interpretations first for reasons of efficiency. In constraint theory, the most diagnostic interpretation will be the most acceptable, all else being equal. By constructing the most diagnostic interpretations first, the model has a better chance of quickly finding the most acceptable interpretation. This approach is consistent with psychological accounts in which the most diagnostic properties of concepts are the most easily accessible (e.g., Barsalou, 1982).

A problem for constrained-search algorithms such as the C^3 model is to ensure that they do not become trapped at "local maxima," returning only the locally most acceptable interpretation rather than the best overall interpretation. For example, in Figure 2b, the local maximum is the highest interpretation in the first step of diagnosticity. As the model proceeds by steps of decreasing diagnosticity, it will find that interpretation first; if it erroneously returned that local maximum as the best interpretation overall, it would miss the best interpretation, which occurs at a lower level of diagnosticity. To ensure that it returns the best interpretation overall, rather than the local maximum, the C^3 model always records the best interpretation it has found so far. The acceptability level of the best interpretation so far is used to exclude some potential interpretations from the search. The interpretations excluded are those whose diagnosticity level is so low that even with the highest possible level of plausibility they could never have a higher acceptability score than the best interpretation found so far. For example, in Figure 2c the black circles represent the set of interpretations constructed at an intermediate stage in processing. The best of those interpretations sets the value for the highest score so far (horizontal line in Figure 2c). Interpretations with particularly low levels of diagnosticity could not exceed this score even if they had the maximum acceptability possible for their diagnosticity level. These interpretations (the circles in the crosshatched area in Figure 2c) need never be constructed, and are excluded from consideration.

The C^3 model searches through all nonexcluded interpretations by steps of lower and lower diagnosticity, until it arrives at a point at which all interpretations have either been considered or excluded from the search (Figure 2d). At this point the model returns the best interpretation it has found: that interpretation is guaranteed to be the best possible interpretation for the phrase in question. Because the C^3 model's search progressively excludes from consideration potential interpretations that have low diagnosticity, it need only consider a subset of the interpretation-space, and hence can find the best interpretation in a reasonable amount of time.

Components of the C^3 model

In its solution to the computational problem of conceptual combination, the C^3 model instantiates the computational-level constraints and knowledge assumptions of the constraint theory. The model takes as input a knowledge base of predicate calculus instance representations and a phrase to be interpreted, and outputs the interpretation that best meets the constraints of diagnosticity, plausibility and informativeness. The model goes through three stages to construct interpretations (see Figure 3). First, the diagnosticity component produces a set of partial interpretations based on subsets of the diagnostic predicates of both constituent concepts, computing diagnosticity scores for each. Second, the plausibility component takes these partial interpretations and generates full interpretations by adding consistent predicates from its knowledge-base, assessing the diagnosticity and plausibility of these full interpretations. Finally, the informativeness and acceptability component computes the informativeness and overall acceptability of each full interpretation. The model goes through these three stages at each level of diagnosticity, constructing the most acceptable interpretation at that level. The process iterates until the best interpretation has been found and no other interpretations need be considered.

In the remainder of this section, we describe these three components using a more elaborated version of the “finger cup” example we used earlier. We consider each of the components in turn and show how they produce a variety of interpretations for this example case.

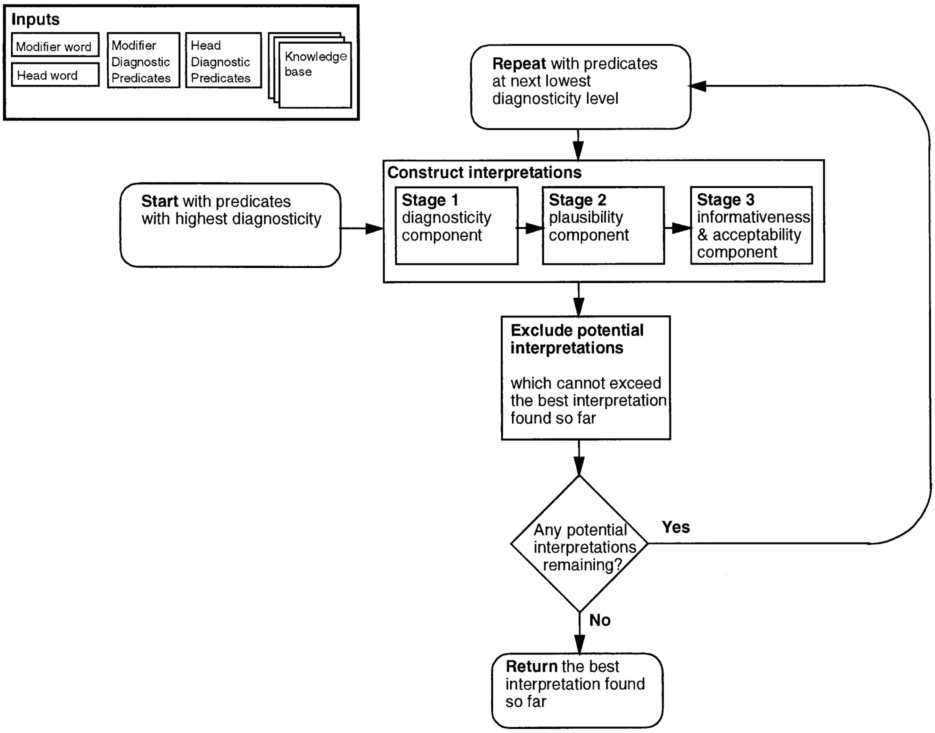
Inputs
Modifier word
Head word
Modifier Diagnostic Predicates
Head Diagnostic Predicates
Knowledge base
Repeat with predicates at next lowest diagnosticity level
Construct interpretations
Start with predicates with highest diagnosticity
Stage 1 diagnosticity component
Stage 2 plausibility component
Stage 3 informativeness & acceptability component
Exclude potential interpretations which cannot exceed the best interpretation found so far
Any potential interpretations remaining?
Yes
No
Return the best interpretation found so far

Figure 3. Flow-chart representation of constrained search algorithm used in the C^3 model.

Representations Used in the C^3 model

Table 1 shows handcrafted examples of the predicate representations used by the model, which are essentially more detailed versions of those graphically depicted in Figure 1. As we said earlier, the graphical depiction finessed three details shown here: the predicates governing the attribute values (e.g., TUBULAR was simply shown instead of SHAPE[F, TUBULAR]); the explicit NAME predicates (e.g., NAME[F, “FINGER”]); and the details of the various roles used [for example, the role labeled W in the *finger* instance has CONSISTENCY [W, SOLID]). In these representations, attributes are notated as SIZE[C, SMALL], which can be read as role C having the value small on the dimension size; relations are notated as connecting two roles, so CONTAINS[C, L] connects role C and role L. The representations in Table 1 also go beyond those in Figure 1 by representing two different instances of a particular category (instance 2 and instance 3). The NAME predicate identifies them as instances of *cup* (i.e., NAME[C, “CUP”] and name[D, “CUP”]). In a larger knowledge-base such as that used in the simulations described below, there would be many different instances of *cup* (and other concepts) along with domain knowledge about the sorts of situations in which those instances occur. As we shall see, this predicate calculus representation formalism can also be used to represent compound phrase interpretations.

TABLE 1
Example Predicate Representations of Instances of Concepts Finger, Cup, and Bowl, as Used by the C^3 Model

Instance			Roles and predicates	
1	''Finger''	NAME[F, ''Finger'']	NAME[H, ''Hand'']	NAME[W, _]
		SHAPE[F, TUBULAR]	CONSISTENCY[H, SOLID]	CONSISTENCY[W, SOLID]
		SIZE[F, SMALL],	SIZE[H, MEDIUM]	
		CONSISTENCY[F, SOLID]	SHAPE[H, FLAT]	
		PART-OF[F, H]	HOLDS[H, W]	
2	''Cup''	NAME[C, ''Cup'']	NAME[L, _]	
		SHAPE[C, HEMISPHERICAL]	CONSISTENCY[L, LIQUID]	
		SIZE[C, SMALL],		
		CONSISTENCY[C, SOLID]		
		CONTAINS[C, L]		
3	''Cup''	NAME[D, ''Cup'']	NAME[E, ''Expresso'']	
		SHAPE[D, HEMISPHERICAL]	CONSISTENCY[E, LIQUID]	
		SIZE[D, SMALL]	COLOUR[E, BROWN]	
		CONSISTENCY[D, SOLID]	TASTE[E, BITTER]	
		CONTAINS[D, E]	TEMPERATURE[E, HOT]	
4	''Bowl''	NAME[B, ''BOWL'']	NAMES[G, _]	NAME[K, _]
		SHAPE[B, HEMISPHERICAL]	CONSISTENCY[G, LIQUID]	CONSISTENCY[K, SOLID]
		SIZE[B, MEDIUM]	TEMPERATURE[G, HOT]	WASH-IN[K, B]
		CONSISTENCY[B, SOLID]		
		CONTAINS[B, G]		

In the following sections we will show how the C^3 model uses the knowledge in these example representations to construct interpretations for our illustrative combination "finger cup."

Constructing Interpretations in the C^3 Model

The C^3 model moves through the space of potential interpretations by steps of decreasing diagnosticity, at each step going through three stages to construct interpretations at that diagnosticity level (see Figures 2 and 3). Below we describe these three stages in detail.

Stage 1: Constructing Partial Interpretations Using Diagnosticity. The diagnosticity constraint requires that a compound phrase interpretation contains diagnostic predicates of both constituent concepts of the phrase being interpreted. An interpretation containing diagnostic predicates is pragmatically acceptable because it describes something best identified by both the modifier and head of the phrase being interpreted. Further, the focal concept of an interpretation is that part of the interpretation containing diagnostic predicates of the head concept of the phrase being interpreted. Diagnosticity has been characterized in a number of different ways (see e.g., Tversky, 1977; Rosch's, 1978, cue validity). In the C^3 model, a predicate is diagnostic if it occurs frequently in instances of the concept and rarely in instances of other concepts (see section A1 of Appendix A for a formal definition). The C^3 model uses the diagnosticity of single predicates or sets of

TABLE 2
Some Diagnostic Predicates for the Concept Cup and the Concept Finger, With Their Diagnosticity Scores

Diagnostic predicates	Score	Occurrence
Diagnostic for concept *cup*		
SIZE[C, SMALL] & CONTAINS[C, L]	1	Only in both cup instances
CONTAINS[C, L]	2/3	Bowl and both cups
CONSISTENCY[C, solid]	1/2	Both cups, finger & bowl
Diagnostic for concept *finger*		
SHAPE[F, TUBULAR]	1	Only in finger instance
PART-OF[F, H]	1	Only in finger instance
SIZE[F, SMALL]	1/3	In finger and both cups

predicates for the constituent concepts of phrases being interpreted. These diagnosticities are computed for the whole knowledge-base before the interpretation process takes place.

In the C^3 model, the process of constructing interpretations begins with the diagnosticity component, which produces partial interpretations based on the diagnostic predicates of the constituent concepts of the phrase being interpreted. For the "finger cup" example, some of the computed diagnosticities of the different predicate sets are shown in Table 2 (these are computed from just the concepts in the knowledge-base shown in Table 1). For *cup*, the predicate set (SIZE[C, small] & CONTAINS[C, L]) gets the highest diagnosticity score: these predicates appear in both instances of *cup* and do not appear in any other concept in the knowledge-base. For *finger*, SHAPE[F, TUBULAR] receives the highest diagnosticity score.

The diagnosticity component constructs partial interpretations by combining these sets of diagnostic predicates by unifying their roles. As there are a number of different possible role unifications, for any given pair of predicate-sets the component may produce a number of alternative unifications. For example, Table 3 shows the three possible role-unifications produced by the diagnosticity component from the predicate sets (SHAPE[F, TUBULAR]) from *finger* and (CONTAINS[C, L] & SIZE[C, SMALL]) from *cup*. These

TABLE 3
Three Alternative Unifications of the Predicates ''SHAPE[F, TUBULAR]'' and the Predicates ''CONTAINS[C, L], SIZE[C, SMALL]''

	Role unifications	Reinstantiated predicates
1	<>	SHAPE[X, TUBULAR] SIZE[Y, SMALL] CONTAINS[Y, Z]
2	F <-> C	SHAPE[X, TUBULAR] SIZE[X, SMALL] CONTAINS[X, Y]
3	F <-> L	SHAPE[X, TUBULAR] SIZE[Y, SMALL] CONTAINS[Y, X]

different unifications are important because they represent very different partial interpretations for the combined concept, since they essentially determine the way multiple concepts are related together in the interpretation. These partial interpretations only become full interpretations after the next stage is completed.

Stage 2: Plausibility and the Generation of Full Interpretations. The plausibility constraint requires that interpretations should be consistent with prior experience, following the pragmatic assumption that an acceptable interpretation is one that is more-or-less already known to the interpreter. Plausibility can be instantiated in a number of different ways (e.g., it could be based on similarity to known concepts; Osherson, Smith, Wilkie, López, & Shafir, 1990). In the C^3 model, an interpretation's plausibility is based the degree to which its predicates co-occur in stored instances. Co-occurrence is measured as the amount of overlap or number of shared predicates between an interpretation and stored instances in the knowledge-base. A completely plausible interpretation is one which overlaps completely with a stored instance (i.e., the interpretation is something that is already known). When an interpretation has a number of partial overlaps with different stored instances, its degree of plausibility is proportional to the average size of those overlaps (see section A2 in Appendix A for a formal definition).

In the second stage of the C^3 model's interpretation-construction process, the plausibility component constructs full interpretations around the partial interpretations constructed by the diagnosticity component in stage 1. To construct a full interpretation with the highest possible plausibility, the overlap between a partial interpretation and every instance in the knowledge-base is determined. The most plausible full interpretation is produced by collapsing across all these overlaps, producing an interpretation whose predicates occur together in as many instances as possible. That full interpretation will have the highest degree of overlap with stored instances, and hence highest possible degree of plausibility. Often the plausibility component will construct several different full interpretations around each partial interpretation, by collapsing across the overlaps in different ways. Plausibility guides the construction of the full interpretations in two ways. First, it helps identify the predicates that can be used to construct the full interpretations. Second, in contributing to overall acceptability, it aids the evaluation of full interpretations by supplying their plausibility score. Of course, because diagnosticity also contributes to overall acceptability, it also plays a role in this stage in evaluating full interpretations as they are constructed.

For the "finger cup" example, Table 4 shows the full interpretation produced around partial interpretation 1 from Table 3 (SHAPE[X, TUBULAR], SIZE[Y, SMALL], CONTAINS[Y, Z]). As Table 4 shows, this full interpretation has a high degree of overlap with the various stored instances, and hence has a high degree of plausibility. This full interpretation could perhaps be glossed as "a f*inger cup* is a cup containing hot liquid in which fingers are washed." The average overlap between this full interpretation and each instance in the example knowledge base is equal to the proportion of predicates in the interpretation that also occur in the instance, for each role the instance and interpretation share. Computing this average overlap gives the interpretation a plausibility score of 0.82. Note that this full

TABLE 4
The Full Interpretation Produced Around Partial Interpretation 1 (SHAPE[X, TUBULAR], SIZE[Y, SMALL] CONTAINS[Y, Z]) by the Plausibility Component, and the Overlap Between That Interpretation and the Stored Instances

Instance		Roles and predicates		
''Finger cup'' interpretation		NAME[X, ''Finger''] SHAPE[X, TUBULAR] SIZE[X, SMALL] CONSISTENCY[X, SOLID] WASH-IN[X, Y]	NAME[X, ''FINGER CUP''] SIZE[Y, SMALL] SHAPE[Y, HEMISPHERICAL] CONSISTENCY[Y, SOLID] CONTAINS[Y, Z]	CONSISTENCY[Z, LIQUID] TEMPERATURE[Z, HOT]
Overlap of roles in stored instances with roles in ''Finger cup'' interpretation				
1	''Finger''	SHAPE[F, TUBULAR] SIZE[F, SMALL] CONSISTENCY[F, SOLID]	————	————
2	''Cup''	————	SHAPE[C, HEMISPHERICAL] SIZE[C, SMALL] CONSISTENCY[C, SOLID] CONTAINS[C, L]	CONSISTENCY[L, LIQUID]
3	''Cup''	————	SHAPE[D, HEMISPHERICAL] SIZE[D, SMALL] CONSISTENCY[D, SOLID] CONTAINS[D, E]	CONSISTENCY[E, LIQUID] TEMPERATURE[E, HOT]
4	''Bowl''	CONSISTENCY[K, SOLID] WASH-IN[K, B]	SHAPE[B, HEMISPHERICAL] CONSISTENCY[B, SOLID] CONTAINS[B, G]	CONSISTENCY[G, LIQUID] TEMPERATURE[G, HOT]

Note. The degree of plausibility of the ''finger cup'' interpretation is equal to the average overlap between roles of the interpretation and roles of the stored instances. The overlap between a role of the interpretation and a stored instance's role is equal to the number of predicates which occur in both roles, divided by the total number of predicates in the interpretation role. The interpretation shown has an overlap of 3/4 with the first stored instance (3 predicates occur in both overlapping roles, 4 predicates in total in interpretation role). The overlaps between the interpretation and instances 2, 3, and 4 are 5/6, 6/6 and 7/10, respectively. The interpretation thus has a plausibility value of 0.82, equal to the average of these four overlaps.

interpretation contains predicates like SIZE[B, SMALL] from *cup* and WASH-IN[A, B] from *bowl*. Although it is perhaps not so surprising that an interpretation acquires predicates from the constituent concepts and related instances, it is important to note that any concept that has some overlap with the partial interpretation can contribute predicates (e.g., the WASH-IN and HOT predicates from *bowl*). This mechanism plays a crucial role in generating semantically-rich interpretations, as knowledge can be drawn from diverse sources in memory.

Full interpretations can similarly be produced for the other partial interpretations shown in Table 3. Table 5, for example, shows a full interpretation produced around partial interpretation 2 (SHAPE[X, TUBULAR], SIZE[X, SMALL], CONTAINS[X, Y]) with that interpretation's overlap with the instances in the example knowledge-base. This interpretation could perhaps be glossed as "a *finger cup* is a cup that is tubular, and shaped like a finger." Notice that this interpretation is a property interpretation (the TUBULAR property of *finger* is asserted of *cup*), whereas the previous interpretation was a relational interpretation (asserting a WASHED-IN relation between *finger* and *cup*). The full interpretation produced

TABLE 5
The Full Interpretation Produced Around Partial Interpretation 2 (SHAPE[X, TUBULAR], SIZE[X, SMALL], CONTAINS[X, Y]) by the Plausibility Component, and the Overlap Between That Interpretation and the Stored Instances

Instance		Roles and predicates	
''Finger cup'' interpretation		NAME[X, ''Finger cup''] SHAPE[X, TUBULAR] SIZE[X, SMALL] CONSISTENCY[X, SOLID] CONTAINS[X, Y]	CONSISTENCY[Y, LIQUID]
Overlap of roles in stored instances with roles in ''Finger cup'' interpretation			
1	''Finger''	SHAPE[F, TUBULAR] SIZE[F, SMALL] CONSISTENCY[F, SOLID]	————
2	''Cup''	SIZE[C, SMALL] CONSISTENCY[C, SOLID] CONTAINS[C, L]	CONSISTENCY[LIQUID]
3	''Cup''	SIZE[D, SMALL] CONSISTENCY[D, SOLID] CONTAINS[D, E]	CONSISTENCY[E, LIQUID]
4	''Bowl''	CONSISTENCY[B, SOLID] CONTAINS[B, G]	CONSISTENCY[G, LIQUID]

Note. The interpretation shown has an overlap of 3/4 with the first stored instance (3 predicates occur in both overlapping roles, 4 predicates in total in interpretation role). The overlaps between the interpretation and instances 2, 3, and 4 are 4/5, and 3/5, respectively. The interpretation thus has a plausibility value of 0.74, equal to the average of these four overlaps.

around partial interpretation 2 has a lower degree of plausibility (0.74) than the previous full interpretation, and would thus have a lower degree of overall acceptability.

Stage 3: Informativeness and Overall Acceptability. The informativeness constraint requires that interpretations should convey more information than could be obtained from the constituents of the compound phrase in question on their own. In the C^3 model, an interpretation is informative if it contains new predicates relative to the prototype representations of constituent concepts of the combination.[4] Prototypes are used in this stage because good interpretations should convey something more that is usually conveyed by the word under most circumstances. To satisfy the informativeness constraint, an interpretation must not be a subset of either the modifier's or the head's prototype representation (see section A3 of Appendix A for a formal definition). Informativeness is not a matter of degree but rather is treated logically as being either present or absent. Technically, the overall acceptability of a full interpretation is based on its combined diagnosticity and plausibility scores (so-called primary acceptability), and its passing the informativeness test (see section A4 of Appendix A).

In the C^3 model, the informativeness component acts to reject full interpretations that are under- or overinformative relative to one another. An underinformative interpretation is one that conveys less information than some other interpretation of which it is a subset (see section A3 of Appendix A). For example, "a *drill screwdriver* is a screwdriver with a side-handle like a drill" is underinformative relative to "a *drill screwdriver* is a screwdriver with a side-handle like a drill, and which is used to bore holes." An overinformative interpretation is one that conveys extra information that does not improve its primary acceptability (see section A3 of Appendix A). For example, the interpretation "An *elephant pig* is an very large pig with two eyes" conveys more information than "an *elephant pig* is a very large pig" (specifically, "has two eyes"), but this extra information does not increase the interpretation's primary acceptability (i.e., the property "has two eyes" is not very diagnostic of elephants). The interpretation "An *elephant pig* is an very large pig with two eyes" would be rejected as overinformative.

To continue with the "finger cup" examples, we can see that both full interpretations produced in Table 4 and 5 would satisfy the informativeness constraint: each contains some predicates not contained in instances of *finger*, and others not contained in instances of *cup*.

Summary of the C^3 Model

The C^3 model's constrained search algorithm and its three-stage interpretation construction process allow the model to avoid the intractability of considering all possible interpretations for a given phrase on the way to finding the best. This model has simulated the interpretation of a large number of novel compound phrases, producing results that generally agreed with people's responses to the same phrases. In the next section we describe these simulations in detail, along with other empirical evidence for constraint theory and the C^3 model. Although the C^3 model is computationally tractable, it still generates a large number of possibilities: in the simulations, on average about 4000 to each compound phrase. Clearly, one future issue for the model is the extent to which this number could be reduced. It could be the case that a parallel constraint satisfaction model might do a better job at only generating a few possible interpretations (as has been shown in analogy; see Holyoak & Thagard, 1989) but the details of such a model are not obvious at this point. Further, we have not considered how context and the differential accessibility of different knowledge might simplify the computational task by limiting the knowledge entering the combination process, or changing the diagnosticities of predicates. Both of these directions are natural and fruitful extensions of the model.

V. EVIDENCE FOR THE CONSTRAINT THEORY OF COMBINATION

In the previous two sections we have described the constraint theory and its implementation in the C^3 model. Here, we describe evidence supporting the constraint theory as an account of conceptual combination. First we outline the empirical support for the constraint theory in terms of its general account for the empirical regularities of conceptual combination. We then describe a number of specific verified predictions that the theory makes. Finally, we describe the simulated interpretation of a large number of compound phrases, showing how these simulations produced results generally agreeing with people's responses to the same phrases.

General Empirical Regularities

The constraint theory explains the wide range of observed regularities in people's conceptual combination in terms of the three constraints of diagnosticity, plausibility and informativeness. In this subsection, we return to the empirical regularities outlined in the first part of the paper to explain how the theory accounts for each one.

Diversity of Interpretation Type. As we saw earlier, five general types of compound phrase interpretation have been recognized: conjunctive, property, hybrid, relational, and known-concept interpretations. Each interpretation type represents a different way of satisfying the three constraints proposed by constraint theory (known-concept interpretations are dealt with later).

Conjunctive interpretations, which describe concepts that are equally an instance of the modifier and the head of the phrase being interpreted, represent cases in which the diagnostic properties of the modifier and the diagnostic properties of the head co-occur plausibly in an instance of both those concepts. For instance, the interpretation "a *pet fish* is a guppy" is acceptable because it describes an already-known instance (satisfying plausibility) which contains diagnostic properties of *pet* (guppies are kept by people) and of *fish* (guppies have fins, gills, and so on).

Property interpretations represent cases where a diagnostic property of one constituent concept is asserted of an instance that already contains diagnostic properties of the other constituent. For example, the property interpretation "a *cactus fish* is a prickly fish" is acceptable because it contains diagnostic properties of cactus (prickly) and of fish (because "fish" is mentioned in the interpretation its diagnostic predicates are implicitly present). Property interpretations in which a diagnostic property of one concept is "overwritten" by a diagnostic property of the other, as in "a *zebra dalmatian* is a Dalmatian with stripes, not spots," are also acceptable because, even though the "overwritten" diagnostic property of *dalmatian* (i.e., SPOTS) is no longer available, other diagnostic properties are still present: a "zebra dalmatian" still possesses the diagnostic shape of a Dalmatian, for example, and the diagnostic contrasting colors. (Note that an interpretation such as "a *zebra* dalmatian is a dalmation with stripes as well as spots" would not be acceptable to the plausibility constraint, because the properties SPOTS and STRIPES do not plausibly occur together). Hybrid interpretations represent cases where many diagnostic properties of one constituent are asserted of the other. For example, the hybrid interpretation "a *drill screwdriver* is a screwdriver with a side-handle like a drill, an electric motor, and which is used to bore holes as well as place screws" has many diagnostic predicates of *drill* asserted of *screwdriver*.

Property or hybrid interpretations based on the assertion of diagnostic properties will only be acceptable if they also satisfy the plausibility constraint; that is, if they describe something consistent with background knowledge. The interpretation "a *cactus fish* is a prickly fish" is consistent with background knowledge because some fish are known to have spines; the interpretation "a *drill screwdriver* is a screwdriver with a side-handle like a drill, an electric motor, and which is used to bore holes as well as place screws" is consistent with background knowledge because similar tools are known to exist: drills with interchangeable heads for boring holes, placing screws, and removing bolts, for example. Both interpretations are thus highly acceptable because they satisfy the diagnosticity and the plausibility constraints. In general, if two similar concepts are combined by asserting a diagnostic property of one concept as true of the other, the resulting property interpretation will have a high degree of consistency with background knowledge, and hence acceptability to the plausibility constraint. Specifically the interpretation will have a high degree of similarity to, and therefore overlap with, stored instances of both concepts; and the more an interpretation overlaps with stored instances, the higher its plausibility score. The plausibility constraint can thus provide an account for the increased occurrence of property and hybrid interpretations when combining concepts are similar (Markman & Wisniewski, 1997; Wisniewski, 1996).

Finally, relational interpretations represent cases where the two concepts being combined (and hence their diagnostic properties) occur in two different parts of an interpretation joined by a relation. For example, an interpretation like “a *horse knife* is a knife for butchering horses” is acceptable to the diagnosticity constraint because the diagnostic properties of both *horse* and *knife* are present implicitly in different parts of the interpretation; that is, the diagnostic properties of knives are present in the concept *knife*, which is part of the interpretation, and the diagnostic properties of horses are present in the concept *horse*, which is also part of the interpretation. Relational interpretations will be acceptable to the plausibility constraint only if the interpretation contains a relation in which both concepts plausibly fit. A relational interpretation will satisfy the plausibility constraint if the new concept taking part in the relation is consistent with background knowledge about other concepts that previously took part in that relation. The interpretation “a *horse knife* is a knife for butchering horses” is acceptable because the concept *horse* possesses many properties consistent with other concepts known to take part in the relation is-butchered, such as cows, pigs and sheep. These properties, such as four-legged, animal, bred by humans, and found-on-farms, mean that the interpretation satisfies the plausibility constraint well. (It should be noted that this differs from other approaches in which concepts can take part in relations only if they possess specific properties necessary for those relations; see e.g., Finin, 1980.)

Diversity of Focal Concept. As we saw earlier, interpretations can have three different types of focal concept. Some interpretations have as their focus the head concept of the phrase being interpreted; other interpretations have as their focus the modifier concept (focus reversals); others have as their focus some concept other than the modifier or head (exocentric compounds). The three constraints account for this observed variety of focal concepts, and explain why that variation occurs. Again each type of focal concept corresponds to a different way of satisfying the constraints. In the constraint theory the focal concept of an interpretation is that part of the interpretation that possesses diagnostic properties of the head concept of the phrase being interpreted. This does not mean that the focal concept of an interpretation need be the head concept of the phrase being interpreted; it need only possess that concept’s diagnostic properties. Phrases such as “horse knife,” whose focal concept is the head concept of the phrase, represent cases where the diagnostic properties of the head occur in conjunction with other properties of that concept. Exocentric phrases such as “seahorse” represent cases where diagnostic properties of the head occur in some other concept distinct from the head (i.e., a fish that has a horse-shaped head). Finally, in focus reversals such as “slipper bed” (a slipper in which a chipmunk sleeps) the focal concept is the modifier, which has been given properties diagnostic of the head.

Semantic Richness and Polysemy. The three constraints also account for the semantic richness of conceptual combinations. Compound phrase interpretations containing knowledge from many different sources are acceptable as long as that knowledge is consistent with background knowledge and with diagnostic properties of the concepts being combined. A number of studies have shown that extra properties arise in conceptual combinations when they are consistent with background knowledge (see Medin & Shoben, 1988; Hampton, 1987, 1988); people interpret the combination “wooden spoon” as describing something longer than the prototypical spoon because they know particular instances of the concept *spoon* which are wooden and long. Thus, these results fit well with the account given by the plausibility constraint.

Furthermore, the informativeness constraint gives an account of the differing levels of detail in different compound phrase interpretations. An interpretation is too detailed if the extra information in it does not increase the interpretation’s acceptability; such interpretations are rejected in favor of less detailed interpretations. Conversely, an interpretation is not detailed enough if extra information could be added to it to increase its acceptability; such interpretations are rejected in favor of more detailed interpretations with higher acceptability.

Finally, the proposed constraints give a natural account of the polysemy of novel compound phrases. Alternative meanings for polysemous compounds correspond to different ways in which interpretations can be constructed so that they satisfy the constraints well.

Specific Empirical Findings

As well as accounting for the general empirical regularities in the creativity of conceptual combination, the constraint theory makes a number of novel predictions about the combination process that have been confirmed in recent empirical studies. We briefly describe these results under four headings: the diagnosticity of property interpretations; property interpretations in focus reversals; known-concept interpretations; and the influence of concept-type on polysemy.

The Diagnosticity of Property Interpretations. A fundamental prediction of the constraint theory is that interpretations will include diagnostic properties of the combining concepts. More specifically, in the case of property interpretations (in which a property of the modifier concept is asserted of the head concept), the asserted property should be diagnostic. Other accounts have suggested that structural alignment is the crucial determinant of the property used (see Markman & Wisniewski, 1997, Wisniewski & Gentner, 1991; Wisniewski & Markman, 1993; Wisniewski, 1996; and later review section). According to this view, property interpretations are produced by a comparison process in which the combining concepts are compared and their shared conceptual structure is placed into alignment. By aligning this shared structure people find differences between the concepts; differences which are interconnected or related to the shared structure (differences which are related to a shared structure are called alignable differences; see Markman & Gentner, 1993; Markman & Wisniewski, 1997). One of these differences then forms the basis for a property interpretation. Stated simply, this view predicts that acceptable property interpretations will be based on alignable differences between the two combining concepts.

Costello & Keane (in press) compared the influence of these two factors—diagnosticity and alignability—in conceptual combination using comprehension and production tasks. In the comprehension task people were shown four different property interpretations for novel noun-noun phrases, and were asked to rate those interpretations in terms of "how good or bad they were as guesses at the meaning of that phrase." The interpretations varied on the dimensions of alignability and diagnosticity (as determined in various pretests). For example, the four interpretations for the novel compound "bumblebee moth" were:

1. moths that are black and yellow (aligned diagnostic)
2. moths that are the size of a bumblebee (aligned nondiagnostic)
3. moths that sting (nonaligned diagnostic)
4. moths that fertilize plants (nonaligned nondiagnostic)

The constraint theory would predict that only those interpretations based on diagnostic properties (i.e., interpretations 1 and 3) would be deemed acceptable, irrespective of whether they were alignable or nonalignable. This prediction was confirmed by the results that showed that subjects reliably rated property interpretations based on diagnostic properties as good interpretations, and reliably rated interpretations based on nondiagnostic interpretations as bad interpretations. Significantly, participants rated nonaligned interpretations as good interpretations as long as they were diagnostic, and rated aligned interpretations as bad interpretations when they were not diagnostic.

In the production task, participants were given the same novel noun-noun phrases and asked to produce what they judged was the best interpretation for each phrase. The interpretations produced were then matched with those used in the comprehension task. The results confirmed the pattern found in the first task, with participants producing many interpretations based on diagnostic properties, and few based on nondiagnostic properties. Significantly, participants often produced nonaligned interpretations, as long as they were diagnostic, and rarely produced aligned interpretations if they were not diagnostic.

Predicted Property Interpretations in Focus Reversals. Constraint theory also makes a prediction about the interpretation type of focus reversals; that is, interpretations where the focal concept is based on the modifier word rather than the head word. The theory's diagnosticity constraint requires that the focal concept of an interpretation must possess diagnostic properties of the head concept of the phrase. In focus reversals, the focal concept is the modifier. For focal reversals to satisfy the diagnosticity constraint, diagnostic properties of the head must be transferred to the modifier. In other words, focal reversals must be property interpretations in which a diagnostic property of the head is transferred to the modifier, as in Wisniewski & Gentner's example "a *chair ladder* is a chair used for climbing on" (Wisniewski & Gentner, 1991).

The status of focal reversals is controversial, with some researchers arguing that they are mistakes or artifactual (see Gleitman & Gleitman, 1970; Gerrig & Murphy, 1992). Nevertheless, people do produce them as interpretations to novel combinations. Notwithstanding their status, constraint theory predicts that focus reversals should be property interpretations rather than relational interpretations. The diagnosticity constraint proposes that an interpretation has to have some of the diagnostic properties of the head word, so if the focal concept of the interpretation is the modifier then it has to have asserted properties from the head (e.g., in the "chair ladder" interpretation *chair* has the diagnostic, functional properties of *ladder*).[5]

In a series of experiments, Costello & Keane (1997a, 1997b) presented participants with novel noun-noun phrases asking them to "say what the phrase could plausibly mean and if you can think of more than one possible meaning for a phrase then report them in the order in which they occur to you." In these polysemy experiments, most of the interpretations people produced used head words as the focal concept (70%), but a significant minority were focal reversals (10%). Typically, in this head-focal category, relational interpretations were more common than property interpretations (roughly 50% versus 35%). However, in the focal reversal category, this trend was reliably reversed with relational interpretations being much less common than property interpretations (roughly 25% versus 55%).

Predicted Occurrence of Known-Concept Interpretations. Constraint theory predicts the occurrence of known-concept interpretations. The plausibility constraint proposes that, all else being equal, people will sometimes produce interpretations with the highest possible degree of plausibility (i.e., an interpretation that is a known concept from prior experience). So, sometimes people will say that "a *clothes tool* is a washing-machine" or "a *stilt bird* is a flamingo." Because the acceptability of an interpretation is an interaction between diagnosticity, plausibility and informativeness, however, these known-concept interpretations will not occur very often: only in cases where there is an already-known concept that satisfies the three constraints. In their polysemy experiments, Costello & Keane (1997a, 1997b) found that known-concept interpretations occurred approximately ten percentage of the time.

Predicted Influence of Concept-Type on Polysemy. A basic proposal of constraint theory is that there many possible interpretations for a novel compound. Hence, the theory naturally expects polysemy to occur. Indeed, when the theory is combined with assumptions about the representation of different conceptual classes—like artifacts and natural kinds—it makes predictions about relative polysemy of different compounds. First, the theory predicts that compounds with artifact heads should be more polysemous that those with natural-kind heads, because the functional models associated with artifacts admit more possible interpretations. For example, “wasp gun” would be more polysemous than “wasp cow,” because a *wasp gun* could be “a gun for shooting at wasps,” “a gun that uses wasps as bullets,” “a gun used by wasps to shoot things” and so on, but a *wasp cow* limited to interpretations such as “a striped cow” or “a cow that stings.” Second, the theory predicts that superordinate-head compounds should be more polysemous than those with basic-level heads; because superordinates admit more different interpretations than basic-level concepts (e.g., a *pavement vehicle* suggests many interpretations based on the subordinates of *vehicle* like “a bicycle for city use,” “a skateboard,” or “rollerblades,” but a “pavement bicycle” is much more restricted by the diversity of subordinates of *bicycle* to being some type of bicycle).

Again these predictions have been confirmed in Costello & Keane’s experiments (Costello & Keane, 1997a, 1997b). In their first experiment, people produced interpretations for novel phrases made up of artifacts and natural-kinds. The results showed that artifact-head phrases evoked significantly more interpretations than natural-kind head phrases. In the second experiment, people produced interpretations for novel phrases made up of artifacts or natural-kinds, and superordinate or basic-level concepts. To control for specific effects due to materials, the same items occurred in modifier position in some phrases and head position in others. Again, artifact-head phrases evoked significantly more interpretations than natural-kind head phrases. Also, superordinate-head phrases produced significantly more interpretations than basic-level head phrases. It has to be granted that these findings rely more on the representational assumptions about conceptual classes than on the constraint theory. However, the theory still plays a significant role in making the predictions. For instance, it is the theory’s proposals on diagnosticity that identify the polysemy as arising from the concept in the head-position rather than the concept in the modifier-position. Furthermore, the predicted superordinate/basic-level distinction hinges on the theory’s proposal that interpretations have to be informative relative to one another to be acceptable. Finally, the C^3 model has simulated the interpretation of the specific phrases used in these experiments, and has been quite successful in reproducing the observed pattern of results, and the other effects outlined in this section.

Computer Simulation of the Conceptual Combination Process

We tested the computational efficiency of the C^3 model by implementing it in a computer program and using that program to simulate the interpretation of a large number of novel compound phrases. The compound phrases used were those given to people in the two experiments described above, allowing us to compare the model’s pattern of responses with people’s responses to the same phrases (see Costello & Keane, 1997a, 1997b, for the specific phrases used). It is clear that there are several levels of detail at which computational models can parallel people’s behavior. At the most general level, a model can simply solve the computational problem posed by a phenomenon in the form of an implemented, effective procedure. This is a very abstract correspondence, but can nevertheless be significant if no previous model has done this. At a more specific level, a model can capture the broad shape of diverse psychological phenomena; for example, as we show below, the C^3 model can produce the different interpretation types observed and simulate different rates of polysemy for different compounds. More specifically, one might expect the model to parallel the exact responses given by people on the task: the interpretations produced by the model should correspond to those produced by people. Finally, at the most specific level, a model can parallel people's actual response time and error performance on the task (see Keane, 1997, for a model of analogy that makes predictions at this task-performance level).

Few models manage to capture all of these levels of correspondence and the C^3 model is no exception to this rule. Specifically, it makes no attempt to parallel people's actual response-time performance, and the closeness of its outputs to people's exact interpretations is questionable. Indeed, we will argue that there are principled difficulties for any knowledge-intensive model in achieving a good correspondence at this level. However, at the higher levels of correspondence the model does a reasonable job. In particular, the model produces results mirroring the observed empirical regularities in the production of different interpretation types and polysemy effects. Before reviewing these parallels we first consider the set-up for the model in these simulations.

Set-Up For C^3 Model Simulations. To conduct the simulations, the C^3 model was implemented in Procyon Common Lisp running on a high-end PowerPC workstation. The model was programmed to produce the 10 best interpretations for the compound phrases presented to it. On average, it took up to two hours to produce the 10 best interpretations for a single phrase, after considering on average 4000 alternative interpretations in the process (earlier runs on a Macintosh LC took 2 days per combination). This set of 10 best interpretations was further divided into good and bad interpretations using a threshold acceptability score. The program was run on a knowledge-base of 76 instance descriptions represented as frames, each of which on average consisted of 22 predicates. These descriptions were produced by the authors using a "combination blind" methodology (see below).

The compound phrases used were drawn from Costello & Keane's experiments investigating the polysemy of novel noun-noun compounds. The materials in Experiment 1 consisted of 24 compound phrases (varying on the dimensions artifact/natural-kind) which were presented to each participant (Costello & Keane, 1997a). To simulate that experiment the simulation program generated interpretations for each of those 24 compound phrases. The materials in Experiment 2 consisted of 32 randomly-generated sets of 16 phrases (varying on the dimensions artifact/natural-kind and superordinate/basic-level), each of which was presented to one participant (Costello & Keane, 1997, b). To simulate that experiment the program generated interpretations for each of the 16 phrases in each of those sets. The total number of phrases interpreted by the program, across both simulations, was 536. The interpretations produced by the program were automatically classified as being relational, property, conjunctive, or hybrid based on the operational definitions used in the psychological experiments (known-concept interpretations could not be classified reliably because of the model's limited knowledge-base). Two important aspects of the model's set-up require further discussion: the knowledge-base and the cutoff threshold for determining the set of best interpretations.

C^3's Knowledge Base. A major worry in any computational model is that the knowledge base used may have been specifically tailored to produce the outputs sought. To avoid this tailored construction, the descriptions used in the simulation's knowledge base were constructed independently by each of the authors in a blind fashion (from a pool of agreed slot and value descriptors). That is to say, the authors did not consult with one another in constructing the representations and they did not know which conceptual combinations would be finally computed by the program. Each concept instance could be an instance of a number of different concepts (e.g., *espresso cup* could be an instance of *cup* and *utensil*). For the purposes of simulating effects such as the effect of head type (artifact/natural-kind) on polysemy, representational distinctions between artifacts/natural-kinds and superordinate/basic-level classes were also agreed. First, artifacts were represented as having more relations than natural kinds (e.g., *knife* had relations indicating that it was "used to cut things," "used to spread things" and "used to stab things," whereas *tulip* only had a single relational entry indicating that it was "used for decoration"). Second, superordinates were represented as referring to multiple instances that were very *different* to one another (in their predicate descriptions) whereas basic-level concepts referred to a smaller number of instances that were *similar* to one another (by virtue of their predicates). All the concepts represented by both authors were used in the knowledge-base without any modification, so that differences in representations of even the same concept reflected some of the variability that exists in different people's knowledge of the world.

Threshold Acceptability Score. The results of the simulations consisted of 536 sets of 10 interpretations, one set for each phrase in the materials. For each interpretation the program also returned the interpretation's acceptability score on a scale from 0 to 1. Each set of 10 interpretations was divided into "good" and "bad" interpretations using a threshold acceptability score. Again, there is always a worry that such a threshold might be used in an unprincipled fashion to fit the outputs of the model to the known results. To avoid such a bias we chose a threshold value on principled grounds and used the same threshold value in all the interpretation sets computed. To allow the broadest possible comparison between the simulation results and the psychological experiments we chose the threshold value that discriminated between good and bad interpretations for as many compound phrases as possible. On the basis of the full set of interpretations produced by the simulation a threshold value of 0.875 was found to be most discriminatory as it divided 420 (out of 536) interpretation sets into good and bad interpretations (a threshold of 0.870 divided 418 interpretation sets into good and bad interpretations, and a threshold of 0.88 divided only 311 interpretation sets). In the following sections we consider how these simulations captured some of the more specific aspects of conceptual combination.

How "Good" Are The Model's Good Interpretations? How well do the model's interpretations correspond to those produced by people? The short answer to this question is that the model's outputs are as good as one could expect given its knowledge-base. So, when it has the relevant knowledge, many of the simulation's good interpretations correspond to ones that people produce. For example, one of the most acceptable interpretations generated by the model was produced for the compound phrase "train hat" (Acceptability = 0.903). This interpretation could be paraphrased as "a *train hat* is a hat which worn on a train."[6] People interpreting the same phrase often produced similar "worn on" interpretations. Similarly, many of the interpretations that the model rated as bad are, indeed, ones that people tend not to produce. For example, one of the least acceptable interpretations generated by the model was produced for the compound phrase "potato ball" (Acceptability = 0.76). This interpretation could be paraphrased as "a *potato ball* is a plastic ball that grows on a potato"; an interpretation that people presented with the phrase never produced. However, the model often does not produce the typical interpretations produced by people, because it often lacks the specific knowledge used in a given interpretation. For example, people often interpreted the phrase "gun horse" as something like "a *gun horse* is a horse used by hunters which is trained to be unafraid of gunshot." The simulation, because it did not have knowledge saying that hunters used horses, or that gunshot provoked fear, could not produce this interpretation. As we saw earlier in our discussion of the semantic richness of interpretations, there is evidence that people have a preference for using specific detailed information when interpreting compound phrases (see e.g., Gray & Smith, 1995). This preference fits with the role of compounds in language, which is to convey detailed information easily, and select specific instances from more general categories. This preference could be one reason why the fit between the model's specific outputs and what people produce is less than perfect, the model's knowledge-base being severely limited relative to the range of specific knowledge available to people.

These difficulties reveal a deeper issue for knowledge-intensive models, like the C^3 model. When a model has processes that depend crucially on the structure/content of the knowledge used, there are in-principle problems in achieving close correspondences to people's behavior. At present, the sheer quantity and uniqueness of people's knowledge cannot be adequately captured by computational models. Better tests of this type of model may become feasible if large-scale knowledge bases become available. However, even these knowledge-bases will not guarantee the production of better simulations because, on the whole, they tend to involve normative knowledge rather than the idiosyncratic facts that may prove to be the bedrock of the creativity in people's combination behavior. Until we can plot, in detail, the diverse contents of an individual's long-term memory the outputs of models like the C^3 model will always be approximate. Given this state of affairs it makes more sense to concentrate on the simulation of broad empirical regularities, and accept that the specific outputs of such simulations will always be just indicative rather than conclusive.

Simulation of the Different Interpretation Types. As we said above, the model produced interpretations of different types when processing the compound phrases used by Costello

& Keane (1997a, 1997b). Here, we report the details of the model's results to show their correspondence of people's behavior on this dimension.

The simulations generated a diverse range of interpretation types, reflecting the diversity of interpretation types found in the psychological experiments. For example, the three most acceptable interpretations generated for the phrase "pencil bed" could be paraphrased as "a bed colored orange like a pencil" (a property interpretation; acceptability = 0.901), "a bed with pencils as legs" (a relational interpretation; acceptability = 0.899) and "a bed with pencil-writing on the sheets" (another relational interpretation; acceptability = 0.88). Furthermore, the distribution of different interpretation types produced by the program was a relatively close reflection of the distribution found in the psychological experiments. In Experiment 1, relational interpretations dominated (46%) followed by property interpretations (33%), with conjunctive/hybrid interpretations being quite rare (0.3%; Costello & Keane, 1997a). In the simulated interpretation of the phrases from Experiment 1, relational interpretations similarly predominated (41%) with property interpretations also frequent (25%) and conjunctive interpretations (3%) occurring rarely. In Experiment 2 the percentage of property interpretations rose (39%) whereas the number of relational interpretations fell (40%) relative to Experiment 1. In the simulated interpretation of the phrases from Experiment 2, the number of property interpretations also rose (33%) whereas the percentage of relational interpretations fell (22%) relative to the first simulation. As such, the simulations captured the relative ordering of the different interpretation types even though the frequencies of production are not identical (exact correspondences are again constrained by the program's limited knowledge-base).

The simulation's outputs also generally paralleled the variation in focal concepts found in people's compound phrase interpretations. Aggregating across all phrases, the program was much more likely to produce interpretations with the head as focal concept (37%) than with the modifier as focal concept (17%). In the experiments, people similarly produced more head-focus interpretations (70%) than modifier-focus interpretations (10%). However, the simulation produced many more exocentric interpretations (interpretations whose focus was some concept other than the modifier or the head; 46%) than were produced in the psychological experiments (20%).

Considering the variation of interpretation types produced within the head-focal and focal reversal categories, the simulations showed a preference for property-interpretations over relational interpretations in focal reversals. In the psychological experiments, head focal interpretations tended to be relational (47%) rather than property interpretations (39%), whereas focal reversals tend to be mainly property interpretations (54%) rather than relational interpretations (27%). Similarly in the simulations, head focal interpretations tended to be relational (27%) rather than property ones (19%), whereas focal reversals tended to be mainly property (93%) rather than relational interpretations (7%). Again, although the exact percentages do not correspond the direction of the differences are captured by the model.

The Model's Simulation of Polysemy Effects. In Experiment 1, people produced significantly more interpretations for head artifact phrases (M = 2.36, SD = 1.06) than for phrases with natural-kind heads (*M* = 2.12, *SD* = 1.11). In the simulation of Experiment 1, the program also produced more interpretations for head artifact phrases (*M* = 3.4, *SD* = 2.3) than for head natural-kind phrases (*M* = 2.3, *SD* = 2.4). In Experiment 2 people also produced significantly more interpretations for head artifact phrases (*M* = 1.96, *SD* = 1.07) than for phrases with natural-kind heads (*M* = 1.77, *SD* = 1.0). People also produced significantly more interpretations for phrases with superordinate head words (*M* = 1.94, *SD* = 1.07) than for phrases with basic-level heads (*M* = 1.79, *SD* = 1.0). In the simulation of Experiment 2, the program again produced more interpretations for head artifact phrases (*M* = 3.25, *SD* = 2.55) than for head natural-kind phrases (*M* = 3.19, *SD* = 2.36), and also produced more interpretations for superordinate head phrases (*M* = 3.39, *SD* = 2.52) than basic-level head phrases (*M* = 3.04, *SD* = 2.37). Its clear that the simulations captured the various polysemy effects found in Experiments 1 and 2. The main difference between the simulation and experimental results are that the model tends to generate slightly more acceptable interpretations than people do; about 3 on average versus a mean of 2 in people.

VI. COMPARISON WITH OTHER THEORIES OF CONCEPTUAL COMBINATION

The Constraint theory is one in a series of theories that have tried to explain conceptual combination. Many of these theories and the empirical work they suggested have played a formative role in shaping the proposals made here. In this section, we review the relevant theories of conceptual combination and assess the advance that constraint theory makes on this previous work. But, even before this review, we can say that a major advance made by the constraint theory is that, unlike previous theories, it has been fully implemented in a computer program that has been tested on a large number of novel phrases. Obviously, we would argue that the proper appreciation of the computational problem underlying conceptual combination and the provision of a tractable solution to this problem is a considerable contribution in itself.

Many different theories of conceptual combination have been proposed in the cognitive science literature: the concept-specialization theory (Cohen & Murphy, 1984; Murphy, 1988), selective-modification theory (Smith, Osherson, Rips, & Keane, 1988), composite-prototype theory (Hampton, 1988, 1991), dual-process theory (Wisniewski & Gentner, 1991, Wisniewski, 1996, 1997a, 1997b), and the CARIN model (Shoben, 1993; Shoben & Gagné, 1997; Gagné & Shoben, 1997). We will not review all of these theories because some have limited applicability and are subsumed by others. Selective-modification theory has only been applied to adjective-noun combinations and does not extend to noun-noun combinations. Composite-prototype theory only accounts for conjunctive combinations such as “a *pet fish* is a guppy.” Finally, concept-specialization theory was mainly proposed to account for relational interpretations and is subsumed into dual-process theory to perform this function. This leaves us with dual-process theory and the CARIN model to review as the main competitors to the constraint theory.

Dual-Process Theory: Scenario Creation and Comparison and Alignment.

Dual-process theory (Wisniewski, 1996, 1997a, 1997b) proposes two main mechanisms for conceptual combination: scenario creation and comparison and alignment. Scenario creation produces relational interpretations, and comparison and alignment produces property and hybrid interpretations. The two mechanisms act independently, though in some cases they may operate in parallel and compete to produce the best interpretation for a given compound (Wisniewski, 1997b). Below we outline each mechanism in turn and the empirical support for them.

Dual-process theory's comparison & alignment account for property and hybrid interpretations makes use of a structural alignment mechanism originally proposed in accounts of analogy (Gentner, 1983; see Keane, 1993, for a review). In structural alignment, the two constituent concepts of a combination are compared by aligning the relational structure that is common to both. The output of this comparison consists of commonalities between the two concepts (properties and relations both share), and two kinds of differences: those linked to those commonalities and interconnected with that shared relational structure (called alignable differences), and those not linked to the commonalities, and not part of the shared structure (called nonalignable differences; Markman & Gentner, 1993, Markman & Wisniewski, 1997).

This structural alignment process makes an important prediction about the properties used in property interpretations. To quote Wisniewski (1996, pp. 449), the process ". . . helps to constrain which properties of the modifier are mapped to the head concept. On this account, properties linked to commonalities between the head and modifier concept would be mapped." In other words, dual-process theory predicts that the properties used in property interpretations will be alignable differences of the concepts being combined (see also Markman & Wisniewski, 1997; Wisniewski & Gentner, 1991; Wisniewski & Markman, 1993, Wisniewski, 1996). For example, consider the phrase "zebra horse." Alignment of the concepts *zebra* and *horse* would yield extensive common relational structure: both animals have similarly-shaped heads, related in similar ways to similarly-shaped torsos, in turn related in similar ways to similar legs and tails. This alignment would also yield an alignable difference linked to those commonalities: a zebra's torso is striped, whereas a horse's torso is typically brown. This alignable difference could then be transferred from *zebra* to *horse*, to produce the property interpretation "a *zebra horse* is a horse with stripes." Further, if the comparison process yields many commonalities and many alignable differences, then people may combine the representations of the two combining concepts, to produce a hybrid interpretation. A number of studies give evidence supporting the role of alignment in property and hybrid interpretations (Markman & Wisniewski, 1997; Wisniewski & Markman, 1993; Wisniewski, 1996).

As well as alignment, dual-process theory gives other factors such as diagnosticity and systematicity a role in property interpretations. Presumably these other factors choose between competing alignable differences if more than one is available (Wisniewski, 1996). It should be noted that, in some alternative versions of the theory, alignment does not play such a central part. Alignment was initially described as selecting or constraining the properties to be used in property interpretations. In alternative versions of the theory, the properties used can be selected by some factor other than alignment. After a property has been selected, alignment plays a role in integrating that property with the concept to which it is transferred.

The scenario creation mechanism in dual-process theory is similar to the concept-specialization mechanism suggested by Cohen & Murphy (1984; see also Murphy, 1988). In concept-specialization, relational interpretations are generated by specializing a slot of the head concept using the modifier concept; so, in the interpretation "an *apartment dog* is a dog that lives in an apartment" the LIVES-IN slot of *dog* is specialized by *apartment*. In a similar fashion, the scenario creation mechanism produces relational interpretations by placing one combining concept into a role in a scenario associated with the other constituent concept. For example, the concept *knife* would be associated with a CUTTING scenario, with roles for agent, object and instrument, corresponding to who did the cutting, what was cut, and what tool was used. The relational interpretation "a *horse knife* is a knife for butchering horses" would be produced by creating a CUTTING scenario in which the concept *horse* filled the object role and *knife* filled the instrument role. In scenario creation, a concept can fill a particular role in a scenario if it possesses the preconditions for that role: those properties which fillers of that role must have.

Dual-process theory uses two subsidiary mechanisms to characterize some other effects. First, the theory has a mechanism of construal, by which a concept used in a combination may be construed as referring to some other, associated concept (Wisniewski, 1996), as in the interpretation "an *artist collector* is a person who collects *the works of* an artist." Second, the content of an interpretation produced by the two processes of alignment and scenario creation may be further elaborated using background knowledge (including domain theories and specific instances). In particular, property interpretations produced by structural alignment may be altered by a construction process that produces a new version of the transferred property more appropriate to the head concept (Wisniewski, 1997a, 1997b). Murphy (1988) proposed a similar elaboration stage, explaining how the phrase "apartment dog," for example, might be understood to describe a dog that is smaller than normal, based on either the knowledge that large dogs would not fare well in confined spaces or known instances of lapdogs who live in apartments. These secondary mechanisms support the dual-process theory's explanation of creativity in conceptual combination. Dual-process theory explains the occurrence of diverse interpretation types (relational, property and hybrid interpretations). The construal mechanism may allow the theory to explain how the focus of some interpretations can be something other than the constituent concepts and elaboration could account for the semantic richness of interpretations.

Dual-process theory is a well-developed account that makes several novel and interesting predictions about conceptual combinations, many of which have been confirmed empirically. However, we would argue that the dual process theory is less parsimonious than the constraint theory because, though both theories assume complex processing mechanisms, the constraint theory does not assume that the different interpretation types are "special cases" requiring specific independent explanations. We would also argue that, unlike the constraint theory, the dual process theory lacks a rationale for some of its proposals. In particular the theory does not tell us why two separate processes should be used to understand novel compounds rather than just one. Finally, from a computational perspective, even though the structural alignment process is well understood from models of analogy (see Falkenhainer, Forbus, & Gentner, 1986; Holyoak & Thagard, 1989; Keane, 1988, 1993, 1996, 1997; Keane et al., 1994; Veale & Keane, 1994, 1997) many other aspects of the model have not been implemented and tested. The elaboration or construction process is to be singled out in this respect, as it is clearly a very complex process that is, as yet, unspecified (c.f. Murphy, 1988).

We know of only one piece of evidence that seems to present difficulties for the dual-process theory; namely, Costello & Keane's (in press) study of people's comprehension and production of property interpretations with properties that were systematically varied in their alignability and diagnosticity (e.g., the "bumblebee moth" examples described earlier). These experiments found that people prefer property interpretations using nonalignable properties (if they are diagnostic) to alignable properties (if they are not diagnostic). This finding runs contrary to dual-process theory's structural alignment account which predicts that alignable properties will be preferred in property interpretations (Wisniewski, 1996). We believe that studies of this type are a fruitful direction for future research, as part of a research program designed to assess which aspects of the constraint and dual-process theories are correct.

The CARIN Model: Relational Templates for Compound Interpretation

The second general theory of conceptual combination comes from a linguistic tradition, which essentially maintains that conceptual combination is not as creative as it first seems. Based on the observation that many compound interpretations conform to a limited number of standard relations, several theorists have proposed a template-filling account (e.g., Levi, 1978; Lees, 1970). Basically, this approach proposes that the concepts in a combination are fitted to existing relational templates, encoding standard relationships like MADE-OF, FOUND-IN, PART-OF, COLOR, SHAPE (see Downing, 1977, for reviews and critiques). This approach has recently been instantiated by the Shoben & Gagné's CARIN model (Competition Among Relations In Nominals), which provides a set of 16 standard relational templates (Shoben, 1993; Gagné & Shoben, 1997; Shoben & Gagné, 1997).

In the CARIN model relational interpretations are formed by placing the combining concepts into one of the 16 relational templates that are commonly used in interpreting compound phrases. The correct relational template for a given pair of combining concepts is the one most often used to interpret other combinations containing those concepts. For example, most people would interpret phrases of the form "chocolate X" as referring to an X made of chocolate. In the CARIN model, this is because people know that MADE-OF is the relation which most commonly occurs in phrases containing the modifier "chocolate" (e.g., "chocolate cake," "chocolate bar," "chocolate egg," and so on). Although there are other possible relations (e.g., in "chocolate allergy") they are less common; the MADE-OF relation is preferred for phrases of the form "chocolate X" because people know many previous phrases of that form in which the relation occurs. Although relational templates may be thought of as general structures abstracted over a range of phrases, they may also be derived from specific cases by analogy; for example, the interpretation of *Iran-gate* or *contra-gate* by analogy to the meaning of *Watergate*.

The CARIN model gives an account for some of the observed diversity of interpretation-types in combination; the relational interpretation-type is clearly handled and the conjunctive interpretation-type is accounted for by the use of an is-a relational template. However, the model does not deal with property interpretations.[7] The model does not explicitly address the issue of semantic richness, though it might arise from analogical instances. Finally, Gagné and Shoben (1997) have also found empirical support for the model's predictions using a sensibility-judgment task; novel compounds containing relations that occurred frequently with the modifier were judged to be sensible faster than compounds using less frequent relations (e.g., "mountain stream," using the frequent LOCATION relation, was judged sensible faster than "mountain magazine," using the less frequent ABOUT relation). They also found that compounds with a small number of frequent relations were judged sensible faster than those with many frequent relations (supporting the idea of competition among relational templates).

The template approach takes up a very different philosophical position to the constraint theory on the issue of creativity in conceptual combination. The problem with this sort of account is that it does not handle the creativity of combination well; there are many interpretations that fall outside the coverage of any proposed set of templates (see e.g., Downing, 1977 for evidence). Models such as the CARIN model lack the fundamental generativity needed to produce the diversity of interpretations found. By analogy to AI planning research, we would argue that the constraint and dual-process theories provide the "first-principles" generativity needed for conceptual combination, whereas the carin model reflects "speed-up learner" aspects of combination (see Smyth & Keane,1996, 1998).

VII. GENERAL DISCUSSION

In this paper we have advanced a theory that attempts to capture both the efficiency and the creativity of conceptual combination. Our main aims have been theoretical: to lay out the computational-level account of the theory and to describe one algorithmic instantiation of this account (the C^3 model). From our perspective, conceptual combination operates within the pragmatics of communication and is a cognitive process guided by the high-level constraints of diagnosticity, plausibility and informativeness. We have tried to show that these constraints can be concretely specified and implemented in an effective procedure that solves the computational problem implicit in the task. The C^3 model admits the full creativity of the combination process and yet efficiently finds the best interpretations for a given novel compound phrase. We would argue that one of the important contributions of this work has been to address the computational problem involved in combination: many previous approaches have tended to ignore this issue, simply because they were not modeled computationally. Finally, we have shown that the theory as a whole can give an account for the broad empirical regularities found in the literature; that it can generate novel predictions; and that many of these predictions have been confirmed.

The constraint theory was developed with a relatively specific focus: it attempts to explain how people construct meanings for noun-noun compound phrases. Having said this, we see it as having implications for general theories of language and meaning. In particular, the theory takes up a stance on the compositionality of language that we will expand on briefly, as a close to the paper.

Constraint Theory and Compositionality

A fundamental principle underlying many general theories of language is the principle of compositionality. This states that the meanings of complex linguistic expressions (such as phrases or sentences) are determined solely and completely by the meanings of their component words and the structural relations between those words. According to the principle of compositionality, anyone who knows the meaning of each word in a complex expression should need no further information to grasp the meaning of the expression. If language is noncompositional, even someone who knows the meaning of each word in a complex expression may not be able to understand the expression, because they lack some further specific information. Conceptual combination has been something of a battleground for the issue of compositionality (see e.g., Butler, 1995; Fodor & Lepore, 1996; Kamp & Partee, 1995). Some have argued that conceptual combinations are in principle noncompositional, drawing on knowledge from abstract theories beyond the two concepts being combined (e.g., Murphy, 1988; Rips, 1995). Others have shown examples of this noncompositionality in practice (e.g., Murphy, 1988; Medin & Shoben, 1988; Springer & Murphy, 1992; Gray & Smith 1995). Typically, these show that in interpreting compound phrases, people sometimes go beyond the meaning of the constituent words of the phrase and make use of other emergent information. For example, people naturally interpret the compound "pet fish" as referring to a small, brightly colored fish kept in a glass bowl, such as a guppy or a goldfish. However, when asked to interpret the words "pet" or "fish" on their own, people never mention the emergent properties SMALL, BRIGHTLY-COLORED or IN-GLASS-BOWL (see Hampton, 1987). As well as these "emergent property" examples, other examples of noncompositionality include exocentric compounds (which refer to concepts other than the two being combined) and compounds based on metaphor or analogy.

In this section we address the issue of compositionality from the standpoint of constraint theory. We have two aims: to show that constraint theory is a compositional theory of combination, and to show that constraint theory can account for many proposed examples of noncompositionality. We begin by describing a number of different properties that provide a functional definition for the idea of compositionality. We then argue that strict compositionality can only occur with strictly defined or "classical" categories, and that, given such categories, constraint theory is strictly compositional. Next, we consider less strictly defined "family resemblance" categories, for which strict compositionality is not possible; we identify some weaker grades of compositionality that can arise for categories of this kind. Finally, we reassess the proposed examples of noncompositionality and show that some in fact represent these weaker grades of compositionality, and can be explained within constraint theory's compositional account.

Unpacking Compositionality

The principle of compositionality plays a central role in theoretical accounts of language and meaning. Different properties of compositionality can help to explain many cognitive functions, such as the ability to communicate, the ability to learn the meanings of words, and the ability to access information about word meanings. Together these explanatory properties provide a functional definition of the idea of compositionality: cases in which all these properties are satisfied will exhibit complete or strict compositionality, whereas cases that only possess some of these properties will show weaker grades of compositionality.

First, compositionality is important in theories of language because it allows communication between people who have different knowledge. Under compositionality two language users will be able to understand each other as long as they both know the meaning of words in their language. Any differences in any other knowledge they have is irrelevant. If language is noncompositional, however, people will only understand each other if they share all other specific knowledge necessary to understand the complex expressions they produce.

Second, compositionality is important because it provides for the generative nature of language. An almost infinite number of new expressions can be produced by combining the words in a language in novel ways; under compositionality, all new expressions can be understood by anybody who knows the meaning of words in the language. If language is noncompositional, even someone who knows the words in the language could nevertheless be unable to understand some new expressions, if they lack the further specific information necessary for those expressions.

Third, compositionality is important for accounts of language learning (Butler, 1995). Under compositionality, once a learner has grasped the meaning of the words in a language they will be able to understand any complex expression they come across, without needing to learn any further information. If language is noncompositional, a learner's task may never be complete: before understanding any complex expression they would have learn not only its constituent words, but also any further specific information necessary for understanding that expression.

Finally, compositionality is important for accounts of access to information in comprehension. Under compositionality, the information accessed in understanding a complex expression is exactly that information accessed in understanding the constituent words of that expression. The same information is accessed in comprehending a word no matter what complex expression it occurs in. If language is noncompositional, different information will be accessed in comprehending a word when it occurs in different complex expressions.

These four properties are at least part of a functional definition of compositionality in terms of the role it plays in accounts of language and meaning. This definition is graded:

different cases may possess these properties to a greater or lesser degree, and hence have a greater or lesser degree of compositionality. In the next two sections we describe how different degrees of compositionality are possible for different types of category.

Strict Compositionality in Classical Categories

In the classical view of categories, category membership is specified by a set of individually necessary and jointly sufficient properties, which are the defining properties for the category (Smith & Medin, 1981). This set of defining properties provide a strict rule for category membership: all instances possessing these defining properties are equally good members of the category; all instances that do not possess those properties are nonmembers. Further, these defining properties are the only properties relevant for making inferences about and classifying members of the category (Komatsu, 1992). In a category that has a classical structure, all instances that are category members possess the defining properties of that category, and instances that are nonmembers do not. All other properties of these instances are completely uncorrelated with category membership.

Combinations of classical categories can be produced by uniting the defining properties of the two categories being combined. These combinations are strictly compositional: they completely satisfy the four properties that give the graded definition of compositionality described above. First, all language users, even if they have very different stored instances of a classical category (different knowledge), will have the same set of defining properties for that category (those properties being the only ones which occur in all members of the category and in no nonmembers). All members of a language community will thus be equally able to understand an expression involving that category. Second, any new combination involving that category will be understood in the same way by all language users who know the defining properties for that category. Third, once a learner has grasped the defining properties for a category from whichever specific set of instances they have seen, they will be able to understand any complex expression containing that category. Finally, exactly the same defining properties are accessed for a category no matter what combination it occurs in.

Constraint theory, when applied to categories with a classical structure, is strictly compositional. In constraint theory, a combination of two classical categories would be interpreted just as all other combinations would be interpreted: by constructing the interpretation which best satisfies the constraints of diagnosticity, plausibility, and informativeness. The defining properties of classical categories are fully diagnostic for those categories: they occur in all members of those categories, and no nonmembers. All other properties are not diagnostic (not correlated with category membership). Thus the interpretation produced for a combination of two classical categories would contain only the defining properties of both categories, and would be strictly compositional.

Other Grades of Compositionality

Most natural-language categories are not classical in nature; rather they have a "family resemblance" or probabilistic structure (Rosch, 1978). In classical categories, the defining properties can be seen as strict rules dictating category membership. Because these rules are strict, combinations of those categories are strictly compositional. In categories with a family resemblance structure, there are no defining properties: there are no properties or sets of properties that occur in all category members and no nonmembers. However, there are some properties that occur in most members of a category, and few nonmembers. These properties have a certain degree of diagnosticity for category membership: an instance that has one of these properties is likely to be a member of the category, but is not definitely a member. The more diagnostic properties an instance has for a given category, the higher that instance's family resemblance to members of the category (the more properties it shares with category members and the fewer properties it shares with nonmembers; Rosch, 1978). The higher an instance's family resemblance to a category, the more typical that instance is as a member of the category.

Combinations of family resemblance categories can be produced by selecting some set of properties from each category being combined. However, no matter what properties are selected, strict or complete compositionality cannot be achieved. Because there are no strictly defining properties for family resemblance categories, two language users who have different stored instances of a family resemblance category could have different sets of properties for that category. Language users who share the same set of category instances would be better able to understand a complex expression involving that category than those who have very different category instances. Further, any new combination involving that category will be understood in the same way by language users who have the same category instances, and in different ways by those who have different category instances. Even when a learner has learnt some specific set of instances for a category, they may need to see some other instances to understand some particular complex expression containing that category. Finally, different properties can be accessed for a category when it occurs in different combinations, depending on the instances available.

We can identify two types of compositionality that are appropriate for family resemblance categories. The first is "subset" compositionality. In subset compositionality a combination contains properties that occurs in some subset of instances of the categories being combined (rather in all instances, as in the case of strict compositionality with defining properties). Different subsets could be used in different combinations. Subset compositionality meets some, but not all, of the properties in our graded definition of compositionality. It provides for a degree of communication between people who have different knowledge (as long as they share some subsets of instances). It provides for the generativity of language (any novel combination can be understood by anyone who has the same or a similar subset of instances), and for language learning (as long as a learner has the right set of instances). However, it does not provide for the equal access to information in all combinations, because different sets of instances could be accessed for different combinations, producing different properties.

Subset compositionality presumes that categories have strict boundaries; it presumes that instances are either category members or not. A second, much weaker, grade of compositionality is "minimal" compositionality. This arises when categories do not have clear membership boundaries but are graded in membership. In minimal compositionality, a combination contains properties that occur in some subset of instances that have a certain (possibly low) degree of membership in the categories being combined (rather than in instances that are definitely members, as in the case of subset compositionality). Minimal compositionality provides for a limited degree of communication between people who have different knowledge (as long as they share both the same subsets of instances, and the same membership gradients for those instances). It provides for the generativity of language (any novel combination can be understood by anyone who has the same subset of instances and membership gradient), and for language learning (as long as a learner has the right set of instances). Again, it does not provide for the equal access to information in all combinations, because different instances could be accessed for different combinations, producing different properties.

In this section we have described two grades of compositionality which, unlike strict compositionality, are appropriate for categories with a family resemblance structure (strict compositionality cannot occur in family resemblance categories). Our reformulation of compositionality as graded in nature is not intended to weaken the idea of compositionality unnecessarily, or turn it into a meaningless, all-inclusive construct. Even though our weaker grades of compositionality include many examples thought to be noncompositional, various kinds of noncompositionality can still occur.

Reassessing Noncompositional Compounds

Possible examples of noncompositional compounds can be classified into three broad categories according to how their noncompositionality arises. First, some compounds are deemed noncompositional by virtue of their use of "emergent" properties; properties not typically true of the combining concepts (as in the "pet fish" example). Second, others are deemed noncompositional because the senses of the combining words are extended, to refer to instances outside the categories usually named by those words. Third, some compounds are classified as noncompositional because they make use of cognitive processes such as metaphor, analogy or metonymy in their interpretation. We argue that the first two categories of noncompositionality in fact represent subset and minimal compositionality respectively; only combinations in the third category are fully noncompositional. Below we describe constraint theory's account of these three types of combination, using as illustration a number of alternative interpretations for the phrase "shovel bird":

1. A "shovel bird" could be a bird with a flat beak for digging up food
2. A "shovel bird" could be a bird that comes to eat worms when you dig the garden
3. A "shovel bird" could be a plane that scoops up water from lakes to dump on fires
4. A "shovel bird" could be a company logo stamped on the handle of a shovel
5. A "shovel bird" could be someone allowed out of jail (free as a bird) as long as he works on a road crew[8]

Noncompositionality via Atypical Instances. The first category of noncompositional compounds are those which make use of properties of atypical instances of the combining concepts (e.g., the "pet fish" example). Clearly, these compounds are not strictly compositional. In constraint theory's account of meaning, compounds of this sort are examples of subset compositionality because properties of a subset of atypical instances contribute to the meaning of the term. In our "shovel bird" examples, interpretation 1, "a bird with a flat beak for digging up food," and interpretation 2, "a bird that comes to eat worms when you dig the garden," would fall into this category. Both these interpretations are formed using knowledge of a subset of instances of the category *bird* (instances of birds with flat beaks and instances of birds that eat worms from broken ground). According to the constraint theory, both interpretations would meet the constraints of diagnosticity, plausibility and informativeness, because they contain diagnostic properties of the concepts *bird* and *shovel* co-occurring plausibly and informatively. Constraint theory could thus account for the production of interpretations of this type.

Some types of privative compound can also be dealt with in this way (e.g., "toy gun"). In privative compounds those instances which are members of the compound ("toy gun"), are necessarily not members of the head concept (*gun*) by itself: anything that is a member of the category "toy gun" is by definition not a real gun. Privative combinations such as these show noncompositionality because their interpretations often contain properties not typically true of the combining concepts: a toy gun might typically shoot plastic arrows with suckers on them, something that is not typically true of either toys or guns by themselves. But in the graded view of compositionality these compounds exhibit subset compositionality: assuming that the meaning of the term "toy" includes knowledge of a subset of instances of toys that shoot plastic arrows, the properties of the combination "toy gun" derive from those instances. Interpretations of privatives such as "toy gun" type are acceptable to the constraints of diagnosticity, plausibility and informativeness, in that they contain diagnostic properties of combining concepts (GUN-SHAPED being diagnostic of *gun*; PLASTIC and USED-BY-CHILDREN being diagnostic of *toy*) co-occurring plausibly and informatively.

Noncompositionality via Sense Extension. The second category of noncompositional compounds are those in which the senses of the combining words are extended, to refer to instances outside the categories usually named by those words. Examples of this category are exocentric compounds such as "seahorse" ("a species of *fish* shaped like a horse") or "jellybean shovel" ("a type of *spoon* for dispensing jellybeans"). Clearly, these compounds are neither strictly compositional, nor examples of subset compositionality. In the graded view of compositionality these compounds are examples of minimal compositionality, because the instances referred to in these combinations have a certain weak degree of membership in a combining concept. Typically, these instances possess some diagnostic properties of a combining concept that identify the instances as possible members (the "seahorse" interpretation having the diagnostic property HORSE-SHAPED; the "jellybean shovel" interpretation having diagnostic properties such as HAS-HANDLE, and USED-FOR-CARRYING). In our "shovel bird" examples, interpretation 3, "a plane that scoops up water from lakes to dump on fires," would fall into this category. This exocentric interpretation would be acceptable to the constraints of diagnosticity, plausibility and informativeness, because it contains properties diagnostic of the concept *bird* (FLIES-IN-AIR) and of the concept *shovel* (USED-FOR-CARRYING) co-occurring plausibly and informatively. However the interpretation would be less acceptable than interpretations 1 and 2, because those interpretations contain properties with a higher degree of diagnosticity.

Noncompositionality via Other Cognitive Mechanisms. Our final category of noncompositional compounds involves examples whose interpretation makes use of cognitive mechanisms apart from conceptual combination, such as metonymy, analogy or metaphor. In our "shovel bird" examples, interpretation 4, "a company logo stamped on the handle of a shovel," and interpretation 5, "someone allowed out of jail (free as a bird) as long as he works on a road crew," represent this type of interpretation. These examples do not show even minimal compositionality, and under our graded definition they would be classified as noncompositional. The question of how theoretical accounts of conceptual combination such as constraint theory could be applied to these examples depends on the relationship between conceptual combination and those other cognitive processes. If conceptual combination and metonymy, analogy, and metaphor are related, perhaps sharing underlying semantic mechanisms, then metaphorical and metonymic meanings could part of the inputs to the combination process, and these cases could perhaps be explained by models of combination such as constraint theory. For example, if the concept *bird* contained the meaning "free as a bird," and the concept *shovel* contained the metonymy "people working with shovels," then perhaps constraint theory could account for examples 4 and 5. If conceptual combination and other processes are not related, however, then theories of conceptual combination will be unable to account for such examples, and they will only be explained by calling on distinct analogical and metaphorical processes.

Precisely how the processes underlying conceptual combination, metonymy, analogy and metaphor are related is currently an open question. Some suggest that these separate parts of language may be closely linked: Wisniewksi (1997b), for example, proposes a link between nominal metaphors such as "my job is a jail" and property interpretations for compounds such as "jail job." There are clear similarities between nominal metaphors and property interpretations: both involve two concepts with a property of one (*jail*) being asserted of the other (*job*). There are equally clear differences, however. Compound phrases and nominal metaphors serve different functions in language: compounds are names that serve to identify a category, whereas metaphors are descriptions rather than names. Compound phrases can have alternate meanings that metaphors may not have (a jail job could be a job working in a jail, or a job which is like a jail; the nominal metaphor "my job is a jail" could not reasonably mean "I work in a jail"). Finally, metaphors seem more productive than many conceptual combinations: the "my job is a jail" metaphor can license a new metaphor "my computer is a pair of handcuffs," but the "jail job" combination does not seem to license a new compound "handcuff computer." Further work should reveal whether the similarities between nominal metaphor and combination are such that the two processes can be seen as sharing a single underlying mechanism, or whether the differences between the two processes outweigh the similarities, requiring a separate account for each. Indeed an important aim for future research on creative language understanding in general is to outline the relationships between the areas of metaphor, analogy, sense extension and conceptual combination. One goal will be to unite; to connect these domains and show the underlying similarities they share. Another, however, must be to divide; to draw a line in the sand to separate analogy from sense extension, metaphor from combination. Only by appreciating the differences, as well as the similarities, between these domains will we come to a full understanding of their separate roles in the process of language comprehension.

Acknowledgments: This research was supported by a Forbairt Presidential Post-Doctoral Award to the first author. Some of the research reported in this paper was completed as part of a PhD dissertation submitted to the University of Dublin, Trinity College (Costello, 1997). The authors would like to thank Art Markman (University of Texas, Austin), Nick Chater (University of Warwick), Ronan Reilly (N.U.I. Dublin), Angela Cotter, and all the members of the Cognition group at Trinity College, especially Ruth Byrne and Mike Mulligan.

NOTES

1. Following Sperber & Wilson (1986) we focus on pragmatic *ex*plicature, in which the primary meaning of communication is derived pragmatically, as opposed to pragmatic *im*plicature, in which a secondary meaning is derived (as in Grice, 1975). Others have suggested a pragmatic influence on both compound creation (e.g., Downing, 1977; Bauer, 1983) and on compound interpretation (Wisniewski, 1997b). Constraint theory is, as far as we know, the first account that explicitly explains how pragmatic principles guide the construction of compound phrase interpretations.
2. At present in our model, the combination process has equal access to all knowledge in memory. However, it is clear that various factors may intervene to increase the accessibility of some knowledge (e.g., priming by context, frequency of memory trace). This is an obvious direction for development of the model. At present, it is sufficient to note that inclusion of such factors would generally have the effect of simplifying the combinatorics of the combination process.
3. As Rosch describes it, the cue validity of a cue x for a category y “. . . increases as the frequency with which cue x is associated with category y increases and decreases as the frequency with which cue x is associated with categories other than y increases.” (Rosch, 1978). This description is very close to our account of the diagnosticity of a property for a category, and differs from most other definitions of cue validity in which the validity of a cue x for a category y is independent of the frequency with which the cue is associated with the category (e.g., Murphy & Medin, 1985). In these other definitions, the validity of a cue for a category depends only on the frequency with which the cue is associated with other categories: a feature could be a perfect cue for a category if it only occurred in one instance of that category, as long as it did not occur in any instances of other categories.
4. The C^3 model does not actually compute the prototypes for the categories in its knowledge base. Prototypes for some categories were generated automatically (by simple inductive generalization) because there were sufficient instances to work from, whereas others were hand-coded (in a fashion that was blind to how they would be used to evaluate interpretations).
5. Wisniewski and Gentner (1991) report focus reversals for compounds such as *stone lion* in which the modifier concept is a substance and the interpretation uses a made-of relation (a *stone lion* is a lion made of stone). Raters judged such interpretations as describing instances of the modifier concept (a *stone lion* is an instance of the concept *stone*), but not of the head concept (a *stone lion* is not a lion). Constraint theory predicts that a more extreme version of this focus reversal will occur for object-object combinations (as opposed to substance-object combinations).

6. The interpretations produced by C^3 model are in the predicate calculus formalism described above. The interpretations we report here are our translations from that formalism.
7. Shoben and Gagné (1997) argue that relational interpretations are the preferred interpretation type, and that property interpretations are rare and only produced as a last resort. However, this claim is unsupported by the results of our polysemy studies (Costello & Keane, 1997a, 1997b) and by other evidence on the order of production of property interpretations (see Wisniewski & Love, 1998).
8. We thank an anonymous reviewer for these examples.

REFERENCES

Barsalou, L. W. (1982). Context—dependent and context—independent information in concepts. *Memory & Cognition,, 10,* 82–93.

Bauer, L. (1983). *English word formation.* Cambridge, England: Cambridge University Press.

Butler, K. (1995). Content, context and compositionality. *Mind & Language, 10*(1/2), 3–24.

Byrne, R. M. J., & Handley, S. J. (1992). Reasoning strategies. *Irish Journal of Psychology, 13*(2), 111–124.

Chater, N., Lyon, K., & Myers, T. (1990). Why are conjunctive categories overextended? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 497–508.

Cohen, B. & Murphy, G. L. (1984). Models of concepts. *Cognitive Science, 8*(1), 27–58.

Costello, F. J. (1997). *Noun-noun conceptual combination: The polysemy of compound phrases.* Unpublished doctoral dissertation, University of Dublin, Trinity College, Ireland.

Costello, F. J, & Keane, M. T. (1997a). Polysemy in conceptual combination: Testing the constraint theory of combination. In *Nineteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Costello, F. J., & Keane, M. T. (1997b). Constraints on conceptual combination: A theory of polysemy in noun-noun combinations. Unpublished manuscript.

Costello, F. J., & Keane, M. T. (1998).Testing two theories of conceptual combination: Alignment versus diagnosticity in the comprehension and production of combined concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Downing, P. (1977). On the creation and use of English compound nouns. *Language, 53*(4), 810–842.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1986). Structure-mapping engine. In *Proceedings of the Annual Conference of the American Association for Artificial Intelligence.* Washington, DC: AAAI.

Finin, T. (1980). *The semantic interpretation of nominal compounds.* In Proceedings of the First Annual Conference on Artificial Intelligence, University of Illinois: USA.

Fodor, J., & Lepore, E. (1996). The red herring and the pet fish: Why concepts still can't be prototypes. *Cognition, 58,* 253–270.

Gagné, C. L., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*(1), 71–87.

Gentner, D. & France, M. (1988). The verb mutability effect: studies of the combinatorial semantics of nouns and verbs. In S. L. Small, G. W. Cottrell, & M. K. Tanenhaus (Eds.), *Lexical ambiguity resolution.* Los Altos, CA: Morgan Kaufman.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170.

Gerrig, R. J., & Murphy, G. L. (1992). Contextual influences on the comprehension of complex concepts. *Language and Cognitive Processes, 7*(3–4), 205–230.

Gleitman, L. R., & Gleitman, H. (1970). *Phrase & paraphrase.* New York: Academic Press.

Gray, K. C., & Smith, E. E. (1995). The role of instance retrieval in understanding complex concepts. *Memory & Cognition, 23,* 665–674.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3). New York: Academic Press.

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory and Cognition, 15*(1), 55–71.

Hampton, J. A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 12–32.

Hampton, J. A. (1991). The combination of prototype concepts. In P. J. Schwanenflugel (Ed.), *The psychology of word meanings.* Hillsdale, NJ: Erlbaum.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*(3), 295–355.

Johnson–Laird, P. N., Schaeken, W., & Byrne, R. M. J. (1992). Propositional reasoning by model. *Psychological Review, 99*(3), 418–439.

Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition, 57,* 129–191.

Kay, P., & Zimmer, K. (1976). On the semantics of compounds and genitives in English. *Sixth California Linguistics Association Proceedings.* San Diego, CA: Campile Press.

Keane, M. (1988). Analogical mechanisms. *Artificial Intelligence Review, 2,* 229–251.

Keane, M. T. (1993). The cognitive processes underlying complex analogies: Theoretical and empirical advances. *Ricerche di Psicologia, 17,* 9–36.

Keane, M. T. (1996). On adaptation in analogy: Tests of pragmatic-importance and adaptability in analogical problem solving. *Quarterly Journal of Experimental Psychology, 49A*, 1062–1085.

Keane, M. T. (1997). What makes an analogy difficult?: The effects of order and causal structure on analogical mapping. *Journal of Experimental Psychology: Learning, Memory & Cognition, 23,* 946–967.

Keane, M. T., Ledgeway, T, & Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science, 18,* 387–438.

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin, 112*(3), 500–526.

Kunda, Z., Miller, D. T., & Claire, T. (1990). Combining social concepts: The role of causal reasoning. *Cognitive Science, 14,* 551–577.

Lees, R. B. (1970). Problems in the grammatical analysis of English nominal compounds. In Manfred Bierwisch & Karl E. Heidolph (Eds.), *Progress in linguistics.* The Hauge: Mouton.

Levi, J. N. (1978). *The syntax and semantics of complex nominals.* New York: Academic Press.

Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language, 32*(4), 517–535.

Markman, A. B., & Wisniewski, E. J. (1997). Same and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Language, Memory & Cognition, 23,* 54–70.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W.H. Freeman.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology, 20*(2), 158–190.

Murphy, G. L. (1988). Comprehending complex concepts. *Cognitive Science, 12*(4), 529–562.

Murphy, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language, 29*(3), 259–288.

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

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*(2), 185–200.

Palmer, S. E. (1989). Levels of description in information processing theories of analogy. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. Cambridge, England: Cambridge University Press.

Potter, M. C., & Falconer, A. (1979). Understanding noun phrases. *Journal of Verbal Learning and Verbal Behaviour, 18,* 509–521.

Rips, L. J. (1995). The current status of research on concept combination. *Mind & Language, 10*(1/2), 72–104.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.) *Cognition and categorization.* Hillsdale, NJ: Erlbaum.

Schaeken, W., Byrne, R. M. J., & Johnson–Laird, P.N. (1995). A comparison of conditional and disjunctive inferences: a case-study of the mental model theory of reasoning. *Psychologica Belgica, 35*(1), 57–70.

Shoben, E. J. (1993). Non-predicating conceptual combinations. *The Psychology of Learning and Motivation, 29,* 391–409.

Shoben, E. J., & Gagné, C. L. (1997). Thematic relations and the creation of combined concepts. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes.* Washington DC: American Psychological Association.

Simmons, D. (1995). A prolegomena to any future metaphysics of Poppy. In P. Z. Brite (Ed.), *Swamp foetus,* (pp. 9–20). Middlesex, England: Penguin.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science, 12,* 485–527.

Smyth, B., & Keane, M.T. (1996). Retrieving reusable designs. *Knowledge Based Systems, 9,* 127–135.

Smyth, B., & Keane, M.T. (1998). Adaptation guided-retrieval: Questioning the similarity assumption in reasoning. *Artificial Intelligence,104,* 1–45.

Sperber, D. & Wilson D. (1986). *Relevance: Communication and cognition.* Oxford: Blackwell.

Springer, K., & Murphy, G. L. (1992). Feature availability in conceptual combination. *Psychological Science, 3*(2), 111–117.

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

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review, 90,* 293–315.

Veale, T., & Keane, M. T. (1994). Belief modeling, intentionality, and perlocution in metaphor comprehension. In *Proceedings of the Sixteenth Annual Meeting of the Cognitive Science Society,* Atlanta, GA. Hillsdale, NJ: Erlbaum.

Veale, T., & Keane, M. T. (1997). The competence of sub-optimal structure mapping on hard analogies. Paper presented at the fifteenth International Joint Conference on Artificial Intelligence, Nagoya, Aichi, Japan.

Wisniewski, E. J. (1996). Construal and similarity in conceptual combination. *Journal of Memory and Language, 35*(3), 434–453.

Wisniewski, E. J. (1997a). Conceptual combination: Possibilities and esthetics. In T. B. Ward, S. M. Smith & J. Vaid (Eds.), *Creative thought.: An investigation of conceptual structures and processes.* Washington DC: American Psychological Association.

Wisniewski, E. J. (1997b). When concepts combine. *Psychonomic Bulletin & Review, 4*(2), 167–183.

Wisniewski, E. J. & Gentner, D. (1991). On the combinatorial semantics of noun pairs: Minor and major adjustments to meaning. In G. B. Simpson (Ed.), *Understanding word and sentence.* Amsterdam: North Holland.

Wisniewski, E. J., & Love, B. C. (1998). Relations versus properties in conceptual combination. *Journal of Memory and Language, 38,* 177–202.

Wisniewski, E. J. & Markman, A. B. (1993). The role of structural alignment in conceptual combination. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society.* Boulder, CO.

APPENDIX A

Formal Descriptions of the C^3 Model's Constraints

In our description of the C^3 model we have, to ease presentation, hidden some of the formal detail of the constraint definitions used by the model. In this appendix, we provide formal definitions of each of the constraints as used by the model, stating precisely what each one computes.

A1. The Diagnosticity Constraint. A predicate or a set of predicates is diagnostic of a concept if every instance in which those predicates occur is also an instance of the concept in question, and those predicates never occur in instances of other concepts. If a new instance also possesses those diagnostic predicates it can be validly classified as an instance of the concept in question. The classification of a new instance possessing less diagnostic predicates would be less valid. We can define D(P, C), the diagnosticity of a set of predicates P for a concept C, in terms of set membership. Let A be the set of instances in a knowledge base which are members of concept C, and let B be the set of instances which possess predicates P. Then the diagnosticity of predicates P for concept C is

$$D(P, C) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

If some part of an interpretation possesses a predicate or set of predicates that are highly diagnostic of a particular concept, those predicates mark that part of the interpretation as being identified by the concept in question. An interpretation I contains many different predicate subsets $P_j \subset I$. Some of those predicate subsets (those that occur mainly in instances of concept C) will identify the interpretation I as containing concept C better than others (those which do not occur in instances of C). The diagnosticity of interpretation I for concept C is equal to the degree of identification of the predicate subset that identifies concept C the best. Formally

$$D_{max}(I, C) = MAX\ [D\ (P_j, C)] \text{ where } P_j \subset I \qquad (2)$$

The underlying pragmatic assumption for the diagnosticity constraint is that an acceptable interpretation is one best identified by both the modifier concept and head concept of the phrase being interpreted. The diagnosticity of a particular interpretation for both modifier concept M and head concept H of the phrase being interpreted is thus

$$\frac{D_{max}\ (I, M) + D_{max}\ (I, H)}{2} \qquad (3)$$

This equation represents an interpretation's score on the diagnosticity constraint. An interpretation that possesses highly diagnostic predicates for both modifier and head concept will satisfy the diagnosticity constraint well; its diagnosticity score will be 1 or close to 1.

A2. The Plausibility Constraint. The plausibility of an interpretation corresponds to the degree to which its properties are consistent with previous knowledge. In the C^3 model plausibility is computed in terms of predicate co-occurrence. The degree of co-occurrence of an interpretation is measured in terms of the amount of overlap that interpretation has with stored instances in the knowledge base of instances. The overlap between an interpretation and a stored instance is the set of predicates that the two share. A completely plausible interpretation is one which overlaps completely with a stored instance (i.e., every predicate in the interpretation also occurs in the stored instance); such an interpretation describes something which is already known. An interpretation that does not overlap completely with any one instance but has a number of partial overlaps with different stored instances has a degree of plausibility proportional to the average size of those overlaps. If those partial overlaps are large, the interpretation contains large numbers of predicates that are known to occur together, and hence is highly plausible. If the overlaps are small, the interpretation is less plausible.

A given interpretation may have a large set of different overlaps with different stored instances. Some will contain a large number of predicates; others will be smaller. If a given small overlap O_a is a subset of a different, larger overlap O_b, then the smaller overlap O_a is redundant, because the set of predicates O_a describes as co-occurring is contained within the set of predicates O_b describes as co-occurring. The plausibility of an interpretation I given a set O containing j nonredundant overlaps is then

$$P\ (I) = \frac{\sum\limits_{1..j} \frac{\text{size}\ (O_j)}{\text{size}\ (I)}}{j} \tag{4}$$

Combining Eqs. (3) and (4) the diagnosticity and plausibility of an interpretation I, given modifier concept M, head concept H, is then

$$\frac{\frac{D_{max}\ (I, M) + D_{max}\ (I, H)}{2} + P\ (I)}{2} \tag{5}$$

Eq. (5) gives the primary acceptability of an interpretation: the degree to which it satisfies the constraints of diagnosticity and plausibility. Diagnosticity and plausibility are both continuous in nature: an interpretation can have any value between 0 and 1. Primary acceptability is thus similarly continuous: an interpretation's primary acceptability can have any value between 0 and 1. The higher an interpretation's primary acceptability, the better it satisfies the constraints of diagnosticity and plausibility. The final constraint of informativeness is logical in nature: an interpretation is either informative or it is not.

A3. The Informativeness Constraint. To decide whether an interpretation is informative relative to a concept it is necessary to compare the focal concept of the interpretation with the concept. If the interpretation's predicates are a subset of the concept's predicates under that comparison, the interpretation is not informative relative to the concept, because it does not contain any predicates not contained in the concept. To satisfy the informativeness constraint, an interpretation must not be a subset of either the modifier concept or the head concept of the phrase being interpreted. Given an interpretation I, modifier concept M, and head concept H

$$\text{Informative(I)} = \begin{cases} \text{IF I} \subset \text{M OR I} \subset \text{H} & \text{FALSE} \\ \text{ELSE} & \text{TRUE} \end{cases} \tag{6}$$

As well as determining whether an interpretation is informative relative to the modifier or head of the compound being interpreted, the informativeness constraint decides whether interpretations are overinformative or underinformative relative to other interpretations. Overinformativeness and underinformativeness are defined formally in Eqs. (7) and (8). Given two interpretations I_1 and I_2,

$$\text{overinformative } (I_2) \text{ IF } I_1 \subset I_2 \text{ AND acceptability } (I_1) < \text{acceptability } (I_2) \quad (7)$$

$$\text{underinformative } (I_2) \text{ IF } I_2 \subset I_1 \text{ AND acceptability } (I_1) > \text{acceptability } (I_2) \quad (8)$$

Interpretation I_2 is overinformative relative to I_1 if I_1 contains all predicates in I_2 and I_1 has a higher acceptability score. Interpretation I_2 is underinformative relative to I_1 if I_2 contains all predicates in I_1 and I_1 has a higher acceptability score.

A4. Overall Acceptability. The full equation for the acceptability of a given interpretation I is a combination of Eq. (5) and Eq. (6):

$$\text{Acceptability (I)} = \begin{cases} \text{Informative (I)} & \dfrac{\dfrac{D_{MAX}(I, M) + D_{MAX}(I, H)}{2} + P(I)}{2} \\ \text{NOT Informative (I)} & 0 \end{cases} \quad (9)$$

Thus if an interpretation is informative, its acceptability is the average of its scores on the diagnosticity and plausibility constraints. If an interpretation is not informative, its acceptability is zero.