ACADEMIC PRESS

Journal of Memory and Language 47 (2002) 548–570

Journal of Memory and Language

www.academicpress.com

Paper has been my ruin: conceptual relations of polysemous senses☆

Devorah E. Klein[a] and Gregory L. Murphy[b,*]

[a] *University of Illinois, 6 Washington Place, NY 10003, USA*
[b] *New York University, 6 Washington Place, NY 10003, USA*

Received 26 June 2001; revision received 28 January 2002

Abstract

Polysemous words have different but related meanings (*senses*), such as *paper* meaning a newspaper or writing material. Six experiments examined the similarity of word senses using categorization and inference tasks. The experiments found that subjects did not categorize together phrases that used a polysemous word in different senses, though they did when the word was used in the same sense. Different senses of a word were categorized together no more than 20% of the time, only slightly more often than different meanings of homonyms. Pre-exposing subjects to a polysemous relation did not increase categorization of word senses that had that relation. Finally, induction from one sense of a word to a different sense was also weak. The results are consistent with the view that polysemous senses are represented separately, often with little semantic overlap, helping to explain previous results that using a word in one sense interferes with using it in another sense, even if the senses are related. Implications for lexical representations are discussed.

Keywords: Polysemy; Ambiguity; Word meaning; Lexical representation

> "Ah," said Mr. Smangle, "paper has been my ruin."
> "A stationer, I presume, sir?" said Mr. Pickwick, innocently.
> "Stationer! No, no; confound and curse me! Not so low as that. No trade. When I say paper, I mean bills."
> "Oh, you use the word in that sense. I see," said Mr. Pickwick.
>
> (Dickens, 1983, p. 549)

Many people think that most words have only one meaning, with the rare exception of homonyms. Homonym meanings are generally unrelated (e.g., *bank* refers to the side of a river or a financial institution), usually arising through historical accident. Less familiar but far more prevalent is the phenomenon of polysemous words, those with multiple related senses.[1] Most content words appear to be polysemous to at least some

☆ This research was supported by Grant NIMH41704. These experiments were reported in Devorah Klein's PhD thesis submitted to the University of Illinois, and she would like to thank her committee members, Gary Dell, Adele Goldberg, Edward Shoben, and Christopher Wickens. Richard Gerrig and Kristine Onishi provided very helpful comments on the manuscript.

* Corresponding author.
E-mail address: gregory.murphy@nyu.edu (G.L. Murphy).

[1] By linguistic convention, the different interpretations of a homonym are referred to as *meanings*, while those of a polysemous word are referred to as *senses*.

0749-596X/02/$ - see front matter
PII: S0749-596X(02)00020-7

degree. As an example, *paper* originally referred to a writing material, but it has evolved to mean the substance normally used to make that material, the content of some writing, and even an oral presentation of that content—so that we can now deliver a paper at a conference without using any paper to do so. Other directions in which *paper* has been extended include the news source (newspaper), which has been stretched to refer to the company that publishes a paper, a representative of the company, and even the editorial policies of the company (see Table 1 for examples). The material sense has also expanded to encompass financial notes, wall coverings, and gift wrap. These different extensions are not accidental, as the different meanings of a homonym are, and the historical progression is often clear (Clark & Clark, 1979; Heine, 1992; Sweetser, 1990). There is a continuum of polysemy, in which closely related senses can be repeatedly extended so that "adjacent" senses are closely related, but the more distant ones have little apparent connection (see Cruse, 1986).

These different senses are so well established in memory that people rarely question the diversity of the senses and how we are able to understand a polysemous word when it is used in many different ways. Linguists, however, have long debated the problem of how polysemous words are represented in the lexicon and have set forth competing explanations. One debate revolves around the question of whether the different senses of a polysemous word are represented distinctly (Langacker, 1987; Rice, 1992; Tuggy, 1993), or whether they can be subsumed under a single, core representation, with the particular sense determined as needed according to context (Nunberg, 1979; for an extreme view see Ruhl, 1989).

Determining which model best describes the representation of polysemous words has ramifications for theories of language processing, as the two main views represent different tradeoffs of storage and processing. Representing only a core is clearly more efficient in terms of storage, but it requires the processor to greatly elaborate the core meaning to derive the particular sense intended in any given utterance. In contrast, representing the different senses separately makes processing a matter of simply selecting the intended sense, but it complicates the lexical representation and raises the problem of determining how senses are distinguished.

The experiments in this article aim to further our understanding of the mental representations of polysemous words. They build on prior experimental research suggesting that at least some frequently occurring senses are represented distinctly. Klein and Murphy (2001) found, quite surprisingly, that using a word in one sense provided no processing advantage for using it in a different sense. In fact, using a word in two different senses tended to slow down the second use, relative to a control condition. This result appears inconsistent with the notion that polysemous senses are related. Most adult speakers of English can see the relationship between *paper* used to refer to a newspaper and *paper* used to refer to writing material more generally. Furthermore, the former developed from the latter historically. If the senses are semantically related, why does using one interfere with using the other? The present research explores the relationship between the senses, using techniques from the psychology of concepts in order to better understand the semantic structure of polysemous words.

Table 1
Sentences illustrating some of the multiple senses of *Paper*

Sense	Sentence using that sense
Substance	That statue is made out of paper
Sheets of material	He needs some paper to draw on
Material with writing	Hand her that paper to read
Meaning of the writing	Did you understand that paper?
Oral presentation	I want to go hear his paper
News source	I read the paper every morning
Newspaper company	The paper might go out of business
Company representative	The paper called about doing an interview with you
Editorial policies	The paper is very pro-Illinois
Class report	I have to go turn in my paper
Wall covering	She got the most beautiful paper for her bedroom walls
Gift wrap	He tore open the paper to get at the present
Commercial paper	The paper on that silver mine is worth 10¢ on the dollar

Background on polysemy

As mentioned earlier, linguists and psychologists have considered whether the different senses of a word have different lexical entries, the *separate sense* view, or if there is just one core representation, with the different extensions being constructed pragmatically in accordance with the context, the *single sense* view. The claim for separate senses is parallel to the usual assumption about how homonyms are represented. Because these two words are pronounced and spelled the same (by definition), they share a *lexeme*, or phonological word form (Cutting & Ferreira, 1999; Levelt, 1989). Since the meanings are so clearly different, homonymous words would have two distinct *lemma* level representations, one for each meaning—that is, they would be represented as two different words, much as they are listed separately in the dictionary.

In an influential article on polysemy, Nunberg (1979) argued that there was no need to have separate lexical entries for the multiple senses of a given word; all that must be stored is a core sense (though Nunberg also questioned whether one could determine which sense was the core). His argument is based on the fact that there are common ways to extend words polysemously. Some of the more frequent relations between senses include *object/substance*, as in (1), and *informational content/object containing that information*, as in (2).

(1) a. The cotton was dying from weevils. [plant]
b. The cotton of his sweater was warm against his skin. [substance]

(2) a. The book was unbelievably tedious. [content]
b. The book was bright yellow. [object]

In (1), the same word is used to refer to a plant and to the material made from that plant. Note that these are different senses, as only the plant can grow or die. Similar patterns can be found in other plant and animal names (*oak*, *pine*, *wheat*, *chicken*, *fish*). In (2), the same word is used to refer to a physical object and to the content carried by that object. These are also different meanings, as the content was not yellow, and the object itself was probably not tedious.

If polysemous senses are based on known patterns of meaning extension, it may not be necessary to store specific senses, since they could be easily derived. For example, one might not need to represent that *elm* can refer to a plant and to the material made from that plant, because this pattern is known more generally. If extended to a psycholinguistic theory, this proposal would suggest that only the core sense of a word needs to be represented in the mental lexicon, with the specific sense of a particular use derived by context and these familiar patterns of extension. Indeed, people do use these patterns of extension to create novel senses that have never been encountered before (Murphy, 1997), and this can be seen when familiar forms of polysemy apply to new words as they enter the lexicon. For example, as new terms like *movie*, *videotape*, *CD*, and *DVD* entered the language, they followed the same polysemic pattern as *book* does—compare (3) with (2).

(3) a. The DVD was boring. [content]
b. The DVD was scratched. [object]

Thus, there is reason to believe that people represent such patterns of extension, and some have assumed that these patterns entail that the senses generated by them are not represented in the lexicon: "Dictionaries that aim to be parsimonious, *or mirror the mental lexicon*, would include only non-predictable readings of polysemous words, along with the rules underlying regular polysemy" (Fellbaum, 2000, p. 53, emphasis added).

In an important early psychological study of polysemy, Caramazza and Grober (1976) also argued for a core representation. They listed 26 senses of the word *line*. Since they believed this to be an excessive number of senses to be represented, they concluded that there must just be a core representation, with specific senses determined online as needed (see also Miller & Johnson-Laird, 1976; Schreuder & Flores d'Arcais, 1989). Lehrer (1990) generally accepted Nunberg's (1979) view, although she also identified some limitations. She showed that there were some cases that pragmatic extension could not explain, and such exceptions would have to be listed separately. For example, the *material/product made of that material* pattern is reasonably productive, but there are instantiations that involve very specific world knowledge rather than being derived solely from this pattern. When we go from the material *iron* to the product *iron*, we mean the specific appliance used to press clothes, not anything made out of iron. In fact, many clothes irons are no longer made of iron yet retain the name. Thus, this sense of *iron* would have to be stored separately rather than derived from the core meaning.

Others have argued that the different senses of a polysemous word are generally represented

separately (Langacker, 1987; Rice, 1992; Tuggy, 1993), unless they are very similar (Cruse, 2000). Frequent extensions shouldn't have to be derived anew on every use, and these common uses may have established representations, just as the different meanings of a homonym do. For example if the news source sense of *paper* is used often, it might have its own representation, even though it is not the core meaning. One of the main arguments for separate sense representation is the claim that different senses are too diverse to derive from a common meaning. In Rice's (1992) analysis of English prepositions, she argued that the different uses clustered around several very different senses. For example, *on* means something quite different in the sentence "The book is on the table" than it does in the sentence "The police officer is on duty." These senses do not have much semantic overlap, and so it is unclear how a core meaning could represent them both. In some cases, the similarity of different senses may require sophisticated analysis to discover and so may not be known by ordinary speakers (Kay, 1992, p. 326 ff).

A hybrid approach to polysemy is also possible. Tuggy (1993) (see also Deane, 1988; Langacker, 1987) proposed a model in which information is stored about both the core (*schema* in his terminology) and the different senses.

Relevant psychological results

Experimental support for the view that different senses of a polysemous word are represented distinctly was found by Klein and Murphy (2001). Their subjects had to judge as quickly as possible whether phrases made sense. The phrases were constructed in pairs in which both phrases used the same sense or different senses of a polysemous word. For example, the phrase *wrapping paper* (which uses the material sense of the word *paper*) might be followed by the same-sense phrase *shredded paper* or a phrase using a different sense, *daily paper* (which uses *paper* in the news source sense). There was also a baseline condition, which primed *paper* in general, without specifically priming either sense. The results showed facilitation when a phrase using one sense of the word was followed by a phrase using the word in the same sense, and interference if the second phrase used the word in a different sense. This result contradicts the predictions of a core representation. If there were only a core sense, every time a phrase using the polysemous word was read, the core would be accessed and the particular extension created. This would predict no priming for consistent sense phrases compared to the baseline. Although a core model could be modified to allow this positive priming (see Klein & Murphy for discussion), the finding of interference for inconsistent sense phrases is particularly problematic for the single sense view, because there is no way to prime one sense and inhibit another if they are not functionally separate. The results are potentially compatible with a core-plus-senses model. However, if there is interference in using two different senses of a word, the core meaning does not seem to have much of an effect: To the degree that different senses of a word share semantic content, one would expect facilitation rather than inhibition. So, if there is a core meaning, it does not appear to be very useful.

The Klein and Murphy (2001) results create something of a puzzle for explaining polysemous sense representation. On the one hand, there is wide agreement that polysemous senses are highly related, that they are created partly by semantic extension along predictable patterns (Nunberg, 1979; Sweetser, 1990). On the other hand, Klein and Murphy found no cross-sense priming in either the semantic judgment task or a memory task, thereby suggesting minimal semantic overlap.

One way to resolve this apparent contradiction is to question whether the different senses are related in the right way. Klein and Murphy (2001) suggested that the different senses of polysemous words are *related* but not *similar*. For example, the substance and news source senses of *paper* are clearly related, because the newspaper is printed on the substance. However, a news source is not in fact very similar to bleached, flattened wood pulp. An even farther stretch is uses such as "The paper fired half its reporters," in which *paper* means a company that publishes a newspaper. A corporation has no semantic overlap with sheets of wood pulp, even if there is a thematic relation between the two senses (the corporation prints its product on that wood pulp). As a result, Klein and Murphy argued, the senses of polysemous words may indeed be quite related, as linguists have claimed, but they may simultaneously be quite dissimilar. When they are, it is not surprising that speakers represent those senses separately. If someone describes a publishing company firing its workers, the semantic representation of writing material would not help you understand this utterance—quite the contrary.

Categories of polysemous senses

The present experiments use techniques from the psychology of concepts in order to explore the strength and type of relations between polysemous senses. If the argument made by Klein and Murphy is correct, then polysemous senses may not have semantic overlap but may be related in other ways. There is a close relationship between the psychological representation of words and concepts (Clark, 1983; Murphy, 1991). In general, one can think of words as picking out categories of objects, events, or properties (Malt, Sloman, Gennari, Shi, & Wang, 1999). One way to phrase the question about the structure of polysemous words, then, is to ask what kind of category or categories is picked out when the word is used.

Categorization research has revealed that there are multiple ways to classify things. In *taxonomic* categories, the form of categorization studied in most categorization experiments, items are grouped according to class membership. Items in taxonomic categories such as *dog* or *furniture* are related by being similar to one another. If we are right in saying that different senses of a word are often not similar, then other types of categories may have to be considered as well.

In *thematic* categories, items are grouped by co-occurrence and functional relations rather than by similarity. For example, dog and bone or umbrella and rain might go together in thematic categories, even though they are not similar. Although thematic categories were thought to be only the province of children (see Markman, 1989), research has found that they are used more broadly, by the elderly (Annett, 1959; Smiley & Brown, 1979), in nonliterate populations (Luria, 1976; Sharp, Cole, & Lave, 1979), and recently even in college students (Lin & Murphy, 2001; Murphy, 2001; Ross & Murphy, 1999).

Another way of grouping things is by *goal-derived*, or *ad hoc* categories (Barsalou, 1983, 1991). These categories link together items that fulfill a goal, for instance, children and jewelry are in the category *things to remove from the house in case of a fire*. These are often nonce categories, created and used on the fly. Despite this, there is general agreement about their contents. A final way of linking things is seen in *radial* or *chained* categories (Lakoff, 1987; Malt et al., 1999). Items in a radial category acquire a common name through a chain of similar items. So, a laminated cardboard box containing juice can be called a *juice box* by similarity to larger cardboard boxes, but the next generation, made of plastic, retains the name, as does the following generation, which is made of colored plastic and is shaped like an animal. This instantiation may share few features with what was initially called *a box*, yet by the process of extension has acquired that name. Malt et al. (1999) argue that this process occurs because new entities need to be named and are given names of the entities that are most similar to them. As a result of multiple chaining events, an original name can get stretched to include objects that are not similar to one another. This phenomenon resembles the polysemy extension process, except that it is phrased by Malt et al. in terms of reference (names for objects, in particular), whereas polysemy refers to extension to novel meanings. For example, different shaped containers that are called *boxes* derive from the same general sense of the word, containers, even if they have somewhat different forms.

In summary, there are many different ways to form a category. In the present experiments, we leave it up to our subjects as to what kinds of categories they wish to use in grouping senses together. We mention the different sorts to point out that similarity-based categories, which are the most well-studied, are not the only form of categorization. Thus, in asking whether senses form coherent categories, we are not requiring that the senses be similar or have any particular relation that we expected. As past research has shown, even when subjects are explicitly told to respond based on taxonomic similarity, they may use other bases to form categories (Lin & Murphy, 2001). We discuss which form of categorization is the most appropriate to describe the relations among senses in the General Discussion.

Experiments 1–5 asked subjects to form categories based on polysemic senses or other kinds of category relations. Experiment 6 explored whether polysemous categories provide a basis for inference. One of the main functions of categories is to allow inference to novel members. If polysemic categories have any coherence, then they should allow induction from one member to another, as taxonomic categories do (Osherson, Smith, Wilkie, López, & Shafir, 1990).

Experiment 1

The first three experiments used a forced-choice sorting task. This type of task was chosen because it provides a good measure of how people

choose to group items with different types of connections between them and is also a very common technique in categorization studies (Gelman & Markman, 1986; Lin & Murphy, 2001; Markman, 1989; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Smiley & Brown, 1979). In our task, subjects were given a target phrase, which used a polysemous word in one sense, and two potential choice phrases. One alternative phrase used the polysemous word in a different sense, and the other used a word that was linked to the target word taxonomically or thematically. Both polysemous words were capitalized, as was the word in the corresponding position in the alternative choice. Consider the following example:

wrapping PAPER

(1) liberal PAPER (2) smooth CLOTH

The task was to read the target phrase on top and think about the capitalized word in the context of the phrase. Subjects then decided which of the two other capitalized words in the context of their phrases went best with the target to form a category, for example, whether the paper of *liberal PAPER* or the cloth of *smooth CLOTH* was more like the target use of *paper*. Selecting the different sense of a polysemous word as fitting with the target is an indication that the two are semantically similar, perhaps sharing a core representation.

Taxonomic or thematic categories were used as the alternative category, because they have been well documented as categorical relations that people use. The taxonomic choices were members of the same superordinate category as the target. For example, as shown above, *cloth* was chosen as the alternative for the material sense of *paper*, because both are kinds of materials. As is well known, superordinate categories are fairly weak and provide few processing benefits, as they do not have many features in common (Rosch et al., 1976). For example, superordinate names are not an effective prime in a perceptual task compared to basic-level names, are considerably slower in classification tasks, and are seldom used to name objects (see Murphy & Lassaline, 1997, for a review). Thus, comparing polysemous senses to superordinate category relations is not a very stringent test of the similarity of the senses.

In the thematic condition, a phrase such as *sharp SCISSORS* would be the alternative choice for *paper*. If, as we have suggested, polysemous senses are related without being very similar, then in the taxonomic category condition subjects are being asked to compare apples and oranges, because items like cloth and paper are somewhat similar without having a strong relation. The thematic alternative provides the opposite situation. Here, items such as paper and scissors or tin and biscuit are not generally similar but have strong spatial or functional relations. If subjects felt pressure to choose the taxonomic category alternative because of a preference for similarity relations, then this would not be operative in the thematic alternative condition.

Clearly, there is *some* perceived relation between different senses of a word, however remote. The question asked in Experiment 1 is whether that relation is as strong and coherent as relations between superordinate category members or thematically related items.

Method

Pretest. Several potential taxonomic and thematic choices for two senses of each polysemous target word were created and tested. Taxonomic choices were created by determining the category to which the target phrase item belonged and selecting another category member that did not have an apparent thematic relation to the target. For example, for the material sense of *paper*, the taxonomic choice selected was *cloth*, another material. In creating the thematic choices, our main concern was that the thematic choice might be less related to the target than the polysemous choice, which could yield more choosing of the polysemous choice than the thematic choice in the actual experiment. For example, the thematic choice for the material sense of *paper* was *scissors*. If this thematic choice was not strongly linked to the target, subjects might rely on the repeated word present in the polysemous choice. To eliminate this possibility, a pretest was used to select thematic choice items that were at least as related to the target as the polysemous phrase was. The taxonomic items were not tested for relatedness, since they are similar to one another rather than being related.

A total of 204 phrases was tested, with the goal of constructing 180 choice phrases for 30 polysemous words. Different lists were used to present all the words in all conditions. Fifteen subjects were presented with a target phrase and a potential choice phrase, with the polysemous word and equivalent choice word in the same position capitalized. They rated how related the two capitalized words were in the context of their phrases on a 7-point scale, with 1 being not related at all and 7 being highly related. Instructions explained that

just because the two words might be the same (as would be true for cases when the polysemous word was a choice), this did not mean they were necessarily related, and the modifying word must be taken into account.

From the resulting ratings, items were selected to best equate relatedness between thematic and polysemous choices. The resulting average relatedness rating for the polysemous choices was 2.64 and for the thematic choices, 3.77. This difference was statistically significant by both subjects and items, $t(14) = 4.17,\ p < .001$; $t(29) = 4.72,\ p < .001$. While the main goal of having the polysemous choice not be more related than the thematic choice was met, the greater relatedness for the thematic choice may be of concern. Particularly, this difference could be a problem if, in the actual experiment, the thematic choice was selected more than the polysemous choice, but the taxonomic choice was not.

Subjects. Thirty-two native English speakers attending the University of Illinois at Urbana-Champaign participated in the main experiment for the partial fulfillment of class requirements. None had participated in the pretest.

Materials. The basic polysemous words and the phrases that distinguished their senses were derived from our earlier work (Klein & Murphy, 2001), which verified that the phrases did indeed pick out different senses of each word. For each of the 30 polysemous words, a total of eight phrases were used, two target phrases, one for each of two senses tested, and three possible choice phrases for each sense: different polysemous sense, taxonomic choice, and thematic choice. Thus, there were four phrases that included *paper*, to make up the same- and different-sense pairs of phrases, as shown in Table 2. Using both senses as possible targets ensured that any results were not due to one sense being more familiar or preferred. Subjects saw each polysemous word in only one trial, either with a thematic or taxonomic alternative. Half of each subject's trials had each type of choice (taxonomic or thematic).

Procedure. The task was forced-choice categorization. Subjects were presented with the target phrase that contained a polysemous word used in one of its senses in the center of the screen; 750 ms later, two choices appeared on the screen below the target. As shown above, the critical word in each phrase was capitalized. Subjects were instructed to decide which of the two capitalized words in the choices, as used in the phrase, went best with the first item to form a category. Position of the polysemous and alternate choice phrases was randomized. Subjects responded by typing the number that corresponded to their choice and pressing the space bar to begin the next trial. They could take as much time as needed. There were 30 trials.

The instructions gave the following definition of a category (adapted from Lin & Murphy, 2001): "A category is a set of things that shares some commonalities—be it genetic makeup, functions, purposes, physical and perceptual characteristics, or behaviors." Subjects were also reminded that "Some words that seem related in isolation might not be all that related in context. You should judge the word as it is used in the particular way suggested by the context" to emphasize that they should not automatically group the polysemous words together. The experiment was run on Macintosh computers using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993).

Results and discussion

Given each target, subjects could choose either a polysemous choice or a regular category choice (either taxonomic or thematic). Results showed that people chose the polysemous choice only about 20% of the time, regardless of whether the alternate choice was taxonomic ($M = 20.2\%$) or thematic ($M = 19.6\%$). This is significantly less than chance by subjects and items, for both the taxonomic alternative choices ($t(31) = 9.96,\ p < .001$; $t(29) = 12.36,\ p < .001$) and thematic alternative choices ($t(31) = 16.45,\ p < .001$; $t(29) = 8.88$,

Table 2
Examples of all phrases used with the word *paper*

Target phrase	Choice alternative		
	Polysemous	Taxonomic	Thematic
Wrapping paper	Liberal paper	Smooth cloth	Sharp scissors
Daily paper	Shredded paper	Evening news	Smart editor

Note. Each trial included a target phrase and a polysemous choice, with either a taxonomic or thematic choice.

$p < .001$). These results indicate that people do not generally perceive the polysemous senses as forming a category, in spite of sharing a word form. The results also provide a manipulation check that the taxonomic and thematic alternatives were in fact related to the target phrase. Both taxonomic and thematic alternatives were preferred to the polysemous senses, and in fact are preferred equally. No doubt the fairly strong thematic relations used were partly responsible for their high level of selection. However, the same preference was shown for superordinate category relations, such as paper and cloth (both materials) or tin and basket (both containers), which we have pointed out are a weak level of taxonomic category. Furthermore, if one believes that adults should shun thematic categories, as has traditionally been argued (Inhelder & Piaget, 1964), it is very surprising that the polysemic senses could beat them only 20% of the time.

One question is to what degree the average results reflect individual differences in responding. If a few subjects did a lot of polysemous sorting, but most subjects did not, the presence of those who preferred polysemous sorting would be masked. In fact, Lin and Murphy (2001) found that subjects usually had very strong preferences for one type of categorization, in their case taxonomic or thematic, but the averaged results hid these preferences. An examination of our results showed no such trend. Only 2 of 32 subjects chose the polysemous choice greater than 50% of the time, with no clear differences between the taxonomic and thematic alternatives. The dramatically different response preferences that Lin and Murphy found were not apparent in these data.

Experiment 2

Experiment 1 demonstrated that people prefer not to group different senses of a polysemous word together. There are several uninteresting explanations for this result. The instructions may somehow have encouraged subjects to choose the taxonomic or thematic alternative choice. Perhaps subjects thought that the items that shared a word were "too obvious" a choice. Another possibility is that because taxonomic categories are more traditional, subjects find taxonomic choices so compelling that they would never group two polysemous phrases together, although this obviously does not explain the equivalent preference for the thematic choices.

To test these hypotheses, Experiment 2 used the same instructions, alternatives, and task as Experiment 1 did, changing only the critical phrases. In this experiment, the target and the polysemous choice used the same sense of the polysemous word. For example, previously, the target *wrapping paper* (which uses the material sense of *paper*) had been paired with *liberal paper* (which uses the news source sense of *paper*). In the present experiment, *wrapping paper* was paired with *shredded paper* (both using the material sense of *paper*), and *daily paper* was paired with *liberal paper*. Note that exactly the same phrases were used as in Experiment 1—they were just re-paired so that each triad had phrases using words in the same sense. The taxonomic and thematic choices were unchanged. If the previous results were due to same-word avoidance, preference for traditional categories, or the instructions, the same results should be obtained here.

Method

The method used was identical to that in Experiment 1, except that items were re-paired so that in each triad, the target and the polysemous choice used the same sense. Thirty-two native English speakers attending the University of Illinois who had not been in the previous experiment participated in this experiment for the partial fulfillment of class requirements.

Results and discussion

For each target phrase, subjects could choose either the phrase containing the same word (now used in the same sense) or the alternative choice (either taxonomic or thematic). In this study, when the polysemous choice was used in the same sense as the target, subjects chose it about 70% of the time, in contrast to 20% of the time in Experiment 1. The preference for the polysemous choice was now greater than chance by both subjects and items for both the taxonomic ($M = 73.1\%$; $t(31) = 8.81$, $p < .001$; $t(29) = 6.54$, $p < .001$) and thematic alternative choices ($M = 69.4\%$; $t(31) = 6.87$, $p < .001$; $t(29) = 5.58$, $p < .001$). Recall that in Experiment 1, such choices were made reliably less than chance. As in Experiment 1, the thematic and taxonomic choices were chosen about equally often.

These results indicate that the results of Experiment 1 were not due to any repeated word bias, preference for traditional categories, or effects of

instructions, since these were equally present in Experiment 2. Rather, they can be attributed to the independence of the different senses.

Similarity analysis

Very different results were found for the same-sense phrases in Experiment 2 than for the different-sense phrases in Experiment 1. We have divided the items up into "same-sense" and "different-sense" pairs, but there is no doubt a continuum of similarity of the meanings of a given word in different contexts. One might wonder, then, whether similarity has an effect within the different-sense items. For example, perhaps the material and news source sense of *paper* are not very similar, whereas the tree and material senses of *oak* are. This could in turn affect the perception of categorical relations between them.

To test this, similarity ratings of the phrase pairs in Experiment 1, collected in the stimulus selection phase of Klein and Murphy (2001), were examined. In this task, 20 subjects rated the similarity of the phrases using the same polysemous word on a scale of 1–7. We correlated these ratings with their sorting results from Experiment 1. That is, for each pair, we correlated the similarity of the different-sense phrases with how often those two phrases were sorted together. That correlation was reliable, both for the taxonomic and thematic alternatives, both r's $= .42$, $p < .05$. To illustrate this effect, the items from Experiment 1 were divided into five equal-sized groups, based on the similarity of the two senses, as shown in Table 3. An examination of the rate of polysemous sorting for each group shows the expected trend: the items generally rated the least similar were sorted together the least, and as similarity increased, so did polysemous sorting. However, note that even the top quintile of items were grouped together less than 30% of the time. Thus, even for the most similar senses tested, the different senses of a word were not usually perceived as forming a category.

Table 3
Mean percent of time the polysemous phrase was chosen in Experiment 1 as a function of similarity (on a scale of 1–7, with low similarity corresponding to low numbers)

Rated similarity	Choice alternative	
	Taxonomic	Thematic
1.6–1.9	7.3	6.3
2.0–2.25	16.0	9.0
2.3–2.65	25.0	26.3
2.7–3.1	25.9	26.8
3.2–4.3	27.1	28.1

This reanalysis of Experiment 1, combined with the findings of Experiment 2, demonstrates that polysemy forms a continuum. Same-sense pairs are often but by no means always perceived as forming a category. Uses of a word in different senses are seldom categorized together, but senses that are more similar were also sorted together more often in the sorting task. Thus, polysemy is not all-or-nothing, but rather reflects a gradation in closeness of the different senses. This predictable variation in categorization also shows that the sorting task is sensitive to semantic overlap and therefore reinforces the surprising finding that different uses of the same word are usually not categorized together.

Experiment 3

Although subjects in Experiment 1 sorted the different senses of a polysemous word together far below both taxonomic and thematic sorting, polysemous sorting was still found in Experiment 1, around 20% of the time. Is this amount meaningful? We have been arguing that this is a very small amount, but it seems likely that completely unrelated phrases would seldom be grouped together, and the fact that polysemous senses are sometimes grouped together might indicate a weak but nonzero relationship. However, this number could also reflect some tendency to sort phrases containing the same word together regardless of the actual meanings of the phrases, as well as simple error or inattention.

To answer this question requires us to establish a baseline of how often people would sort together unrelated phrases containing the same word. This was accomplished through the use of homonyms, which have unrelated meanings. A phrase involving one meaning was used as a target, and a phrase involving another meaning acted as one of the choices. The other choice used a word that was related thematically to the target meaning of the homonym. Thematic choices were used because roughly equal preference for taxonomic and thematic choices had been found in both Experiments 1 and 2 and because we could equate their relatedness in the polysemous and homonym items. An example of a homonym trial is:

national BANK

(1) river BANK (2) checking ACCOUNT

The question of interest was how often subjects chose the phrase with the same word (river BANK), even though its meaning was unrelated to the target phrase. For purposes of comparison, the polysemous items from Experiment 1 were also included.

Equivalent levels of polysemous choice and homonym choice sorting would show that the tendency to group different polysemous senses together is due only to the sharing of a word form. This would support a model in which there are virtually no perceived connections between the senses of a polysemous word. A finding of less grouping together of homonym meanings would indicate that subjects recognize some conceptual overlap of the two polysemous senses.

A different way to conceive of this experiment is as a way of assessing more accurately just how related polysemous senses are to one another. If sense relatedness lies on a continuum, the homonym condition can be thought of as establishing the lower bound of that continuum. The results will position polysemous senses relative to that lower bound.

Method

Subjects. Thirty-two new native English speakers attending the University of Illinois participated in this experiment for the partial fulfillment of class requirements.

Materials. The 30 polysemous items from Experiment 1 in their thematic-alternative triads were used in this experiment. In addition, 24 homonym items were also created, along with corresponding thematic choices. Homonyms were found in the lexical access literature, and their homonym status was independently verified by checking that their meanings had separate listings in the dictionary. They included often-tested homonyms such as *calf*, *match*, and *bank* and are listed in the Appendix of Klein and Murphy (2001). Both interpretations of the polysemous and homonym items were used as targets. The thematic alternatives were pretested to ensure that they were equally related for the two kinds of words. The thematic choices were equally close for the polysemous (rated 4.11 out of 7) and homonym (rated 4.37) items, $t(8) = 1.84,\ p > .10$; $t(52) = 1.04,\ p > .30$. Therefore, if the results find differences between the two types of items, they cannot be due to differences in relatedness to their thematic alternatives. Subjects saw each word in only one triad, and different lists were used to present all the words in all conditions.

Procedure. The task was identical to that used in Experiment 1, namely a forced-choice categorization task with 54 trials.

Results and discussion

Subjects could choose either the thematic alternative or the same-word (polysemous or homonym) choice. Results showed that people chose the polysemous choice 14.1% of the time, but selected the homonym choice only 6.6% of the time, and this difference was reliable by both subjects and items, $t(31) = 3.66,\ p < .001$; $t(52) = 3.47, p < .005$.

When the data were examined more closely, three homonym items were found whose meanings may not have been well distinguished in their phrases, yielding relatively high levels of homonym sorting (19–28%). These items appeared to be outliers, and when they were removed, the difference between polysemous and homonym items was even greater, increasing to 9.7%. However, the results do point out that even unrelated uses of a word will be sorted together some nonzero amount of the time, and so one does need a baseline as a comparison for the polysemous items. Whether the 6.6% homonym choices are due to "error" or legitimate judgments of relatedness is not known (and was not the focus of the experiment).

The results show that there is an advantage for the polysemous words over the homonyms, and thus that the polysemous items share more connections than just a repeated word. The polysemous senses have more conceptual coherence than the different homonym meanings do. However, the results also reveal that these connections are not at all strong. The grouping of polysemous phrases was only 20% in Experiment 1 and only 14% in the present experiment. Although this was significantly greater than an appropriate baseline, it is clearly quite low.

In short, the present results are reassuring in two respects. First, they replicate the very low amount of sorting of polysemous senses found in Experiment 1. However, they also verify the intuition that these senses are more related than the meanings of homonyms are, and they confirm that the sorting task was sensitive enough to reveal even small differences in relatedness.

Experiment 4

The results so far suggest that polysemous senses are related, but not very strongly. The

findings help to explain the lack of positive transfer from interpreting a word in one sense to interpreting it in another sense (Klein & Murphy, 2001). Experiments 4 and 5 attempt to characterize the nature of the relationship between the different senses of a word. On the single sense view, people generate most of the senses from a stored core meaning, and so the question of how those senses are related in memory does not arise. On the separate sense view, people store the familiar uses of a word separately. How are these different uses represented and coordinated in memory? One possibility is that they are simply created and stored independently, as they are encountered (e.g., see discussion in Kawamoto, 1993; Klein & Murphy, 2001). Another possibility is that the senses are connected by labeled links of some kind. To explain why this is a possibility, we need to briefly review some of the linguistic literature on polysemy.

As explained in the Introduction, polysemous words often follow familiar patterns. Within a semantic field, a number of words may have a consistent form of polysemy. For example, words referring to animals can also be used (as mass nouns) to refer to the meat of that animal:

(4) I saw a/I ate some... *chicken*, *horse*, *fish*, *rat*, *squid*.

Almost all nouns can be used to refer to an individual member of a category and the entire category, as in (5).

(5) a. The dog has been domesticated for millennia. [the class of dogs]
b. The dog is drooling on the baby. [an individual member of that class]

Within the domain of information sources, most nouns that refer to an information container can be used polysemously to refer to the object itself and the information it contains, as shown in (2) and (3) earlier. There are other productive patterns within specific domains, as discussed by Lehrer (1990), Nunberg (1979), and Sweetser (1990), among others.

The use of familiar patterns of polysemy could influence the representation and processing of polysemous words. Consider first the single sense view. On this view, most senses are derived via these familiar rules of extending meanings (e.g., Caramazza & Grober, 1976). If this process is like most other mental processes, applying the pattern in one case should affect later uses of the same pattern. For example, if one applies the animal-meat extension in one case, this might speed the use of that extension in a subsequent case, but it would have little effect on applying the object-content extension.

Similarly, suppose that senses are prestored in labeled links. It is possible that traversing the link from an animal to a meat sense in one word would benefit the access of the same link for another word. Here, the prediction is not as certain, since different links would be involved (e.g., two senses of *chicken* in one case and the two senses of *salmon* in another). However, if one assumes that the familiar patterns of polysemy are explicitly represented, then it seems likely that choosing one pattern should give an identically labeled pattern an advantage.

Experiment 4 investigated this possibility by asking whether a task that serves to prime information about the relation between the different senses will yield more polysemous sorting. More specifically, reading a paragraph that uses two different senses of a polysemous word may lead subjects to see that the senses are linked and how one might have been extended from the other. This specific information about a relation among senses might prime a similar relation in the categorization task, leading the polysemous senses to be grouped together at a higher rate. The single sense view clearly makes such a prediction, and the labeled-link notion would certainly be consistent with such a finding. Thus, this experiment should help to reveal the nature of the stored relations, if any, among polysemic senses.

Each trial had three parts. First, subjects read a short paragraph that used two senses of a new polysemous word. Second, they then wrote down the commonalities between the two senses. This was to ensure that they fully understood the two senses of the word and had thought about the links between them. Third, immediately afterward, they did a forced-choice task similar to that used in the prior experiments. The relation between the first paragraph and the sorting task was the independent variable. The two senses of the polysemous word used in the paragraph could either parallel or not parallel the relationship between the two senses in the forced-choice task. This word was matched to one of the experimental words in having the same relation between the different senses. For example, *videotape* can be used in material (*disintegrating videotape*) or information source (*boring videotape*) senses, which parallel the material and news source senses of *paper*. If people represent and use such patterns, then reading a paragraph

involving this pattern could lead to a preference for sorting the analogous senses of *paper* together in a forced-choice task.

However, it is possible that being forced to consider the same word in two uses could act as a demand characteristic causing increased selection of the polysemous choice, independent of the particular preceding paragraph. That is, simply repeating the word *videotape* might draw attention to the repeated word *paper* in the test. To rule this out, a control paragraph was used that also contained two senses of a polysemous word but that did not involve the specific relationship between the tested senses. For example, the paragraph could use the word *wrap* in both the clothing and culinary senses, which do not bear the same relation as the two senses of *paper* do. This paragraph should therefore not lead to a preference for grouping the two senses of *paper*.

The third part of the trial was identical to the forced-choice task used in Experiment 1, using only the taxonomic alternatives. (No difference between thematic and taxonomic alternatives had been found in Experiments 1 and 2. In addition, because of the nature of the priming task, it was possible that thematic links could be inadvertently evoked by the preceding paragraph, which would disrupt the manipulation.)

For example, subjects would read one of the following paragraphs.

Parallel with paper: The videotape they were watching was very boring. Worse, the quality was poor and the videotape was disintegrating as they played it.

Nonparallel with paper: Rebecca pulled her wrap around her shoulders as they walked into the restaurant. Inside, she ordered a tortilla wrap with black beans.

Then, after writing the commonalities of the two senses of either *videotape* (using the same polysemic pattern as used by *paper*) or *wrap* (using a different pattern), subjects saw the same triads for *paper* as in Experiment 1 and performed the same task. If polysemic patterns are explicitly represented, sorting of the different senses of *paper* would be predicted to be higher following the parallel paragraph compared to the nonparallel paragraph.

Method

Subjects. Twenty-four native English speakers attending the University of Illinois at Urbana-Champaign who had not been in the previous experiments participated in this experiment.

Materials. New polysemous words were found that matched the original set in having the same relation between the senses. They were then embedded in short paragraphs that used both senses. The test items were identical to the taxonomic triads in Experiment 1. Half of the sorting trials were preceded by a parallel and half by a nonparallel paragraph, counterbalanced across subjects.

Procedure. As described above, there were three interleaved parts to each trial in this experiment. The first part was to read a short paragraph. Then subjects wrote down commonalities between the two senses of the polysemous word used in the paragraph. The third part was the forced-choice categorization task used in prior experiments. So, the forced-choice triad was immediately preceded by its priming context. Subjects were informed in the instructions that these tasks were not related and were interleaved "to avoid boredom." Informal questioning after the experiment revealed that subjects accepted this explanation. There were 30 trials.

Results

The dependent variable in the experiment was the percent of polysemous sorting following either a parallel or nonparallel paragraph. Results showed that people chose the polysemous choice only about 18% of the time, regardless of the preceding context, which was about the same proportion as in Experiments 1 and 3. There was no difference between the percent of polysemous sorting when preceded by the parallel (17.4%) or nonparallel (18.0%) paragraph (all t's < 1).

There are several possible reasons for why the manipulation had no effect. One is that the relations used in the priming paragraphs were not similar enough to the tested polysemous word. To make the relationship between the two parts of the task more explicit and make the priming of the sense relations more exact, an experiment was conducted which used the identical polysemous word in the paragraph as in the forced-choice task. We postpone discussion until those results have been reported.

Experiment 5

In an attempt to "hit the subjects over the head" in order to obtain an effect, we repeated the previous experiment, only using priming

paragraphs that used the same word as was tested. This ensured that the same senses and sense relations were involved in the priming paragraph and sorting task. Using the same words also addressed the possibility that the relationship between a videotape and its content (see example above) is not exactly the same as that between paper and its content. If concrete information about the connections between senses is stored, activating that connection should yield more priming than activating the analogous relation of a different word.

The control paragraphs used in Experiment 4 could not be used here, because they used different words (e.g., *videotape* instead of *paper*), in contrast to the priming paragraphs. Therefore, we contrasted paragraphs that used the target word twice, in a way that was parallel or nonparallel to the following triad. The parallel condition used the word in two different senses, whereas the nonparallel (control) condition used the word in the same sense twice and so would not prime the relation between the two senses. For example, subjects would see one of the following paragraphs before doing a sorting task with *paper*.

Parallel: Bill finished reading the morning paper and went to go get ready for the birthday party. He got out some shiny paper to cover the gift in.

Nonparallel (material sense): The night before Christmas, Bill finished rolling the mug in paper and moved on to the next gift. He got out some shiny paper to cover the toy in.

Nonparallel (news source sense): Bill finished reading the morning paper and went to go get ready for work. He got out yesterday's paper for the train. He hadn't finished the crossword puzzle. The nonparallel paragraphs controlled for number of exposures of the polysemous word, while not priming the relation between the senses. If there is an established representation of the semantic relation between the senses, the fact that the relation is processed in the parallel paragraphs but not in the nonparallel paragraphs should lead to an increase in polysemous choices in the former.

Method

The procedure was identical to that of Experiment 4, in that subjects read short paragraphs containing two phrases using a polysemous word, followed by the sorting task. However, now the two phrases used either the same senses or different senses of the target word. The nonparallel paragraph used the same sense as the choice item in the following triad; across subjects, the different senses were used as nonparallel primes equally often. There were 32 new subjects from the same population.

Results

People chose the polysemous choice 15.2% of the time when they received a paragraph using both senses, and 11.5% of the time when the preceding paragraph used only the target sense, which was not a reliable difference, $t(31) = 1.47$, $p = .15$; $t(29) = 2.01$, $p < .06$. The difference between the paragraph types approached reliability in the item analysis, but given the strength of the manipulation, a much larger difference had been expected. The difference is small and the absolute numbers are quite low, indicating that the preceding paragraph task did not have a strong effect on sorting. For example, the polysemous sorting in the primed condition is less in absolute terms than that found in Experiments 1 and 4. Even if the difference were to become reliable with more subjects, the very small effect size gives little confidence in the existence of labeled relations linking senses. Indeed, this null effect is quite surprising given that subjects may have consciously noticed the relationship between the priming paragraph and test phrases.

Discussion

The first three experiments showed that people do not see strong links between the different senses of a polysemous word, although the connections were stronger than for the different meanings of homonyms. Experiments 4 and 5 looked at the more specific question of the form of those sense relations. If the senses were actively constructed by applying a polysemic pattern to the core meaning of words, there should have been priming of relations when those patterns were repeated. However, there was no significant priming, which is inconsistent with the notion that senses are constructed from a core. Thus, this result is consistent with Klein and Murphy's (2001) conclusion that senses are separately represented.

For prestored senses, it would still be possible for links to be labeled by the form of polysemy that generates them. Certainly, links of this type have been popular in theories of semantic memory since Collins and Quillian (1969). If something

like an animal-meat link is involved in specifying the relevant senses of *chicken*, then one might expect that using the same type of link in the same or even another word would show priming. One might question whether the present task had the potential to show such priming, but similar manipulations in other domains have shown analogous priming. In particular, reading or performing a task with a conceptual combination (e.g., *snake smile* meaning a smile at seeing a snake) influences the interpretation of a subsequent potentially ambiguous combination (e.g., *dog smile*) (Gagné, 2001; Gerrig & Murphy, 1992; Klein & Shoben, 1998). As the interpretation of the combination has to do with priming a particular kind of relation (a smile in response to a snake), this seems analogous to the present attempt to prime the relation between senses of a word (animal-meat; object-content) (see also Murphy, 1990). Most likely the difference is that the different senses of these polysemous words are already stored in semantic memory, as opposed to being constructed online, as novel combinations are. Thus, the fact that one gets such effects in conceptual combination suggests that an online construction process is not the way that people process familiar senses.

As with any null result, it is of course possible that future work could find evidence for the existence of stored sense relations, perhaps using a different task. The marginal significance in the item analysis of Experiment 5 suggests that if there is such an effect, it is tied to a particular lexical item rather than being a general one. That is, using the animal-meat polysemy pattern for one word apparently does not affect the same pattern in another word. However, using *chicken* in both senses may activate the relation between the two senses within this word. If this is true, it could suggest that the animal-meat relation (for example) is not as consistent across lexical items as linguists have assumed; alternatively, it could reflect a processing limitation in which a labeled relation in one word does not affect the use of the same relation in another word. However, speculation on this matter is premature, given the tiny effect found in Experiment 5, which did not reach conventional significance levels.

Experiment 6

We have found thus far that people usually do not perceive different senses of a polysemous word as belonging to the same category, even when relations between those senses are primed. However, the sorting task used in the previous studies depends on conscious decisions about what constitutes a category: Subjects were instructed to think about the phrases and then group them together based on their intuitions. It is possible that relations among the different senses are present in semantic structure itself, even if subjects do not believe that they form "a category." Thus, we used a different task with a different kind of decision to attempt to replicate the results of the categorization judgment, namely induction.

The ability to make inferences from one category member to another is an important function of categorization. Knowing that dogs bear live young helps one to infer that other mammals, such as mice, also bear live young. It is possible that the relationship between the different senses of a polysemous word, while not able to aid in a sorting task, could be found in an induction task.

Rips (1975) was perhaps the first to show that category membership can serve as the basis for induction, a finding confirmed in numerous later studies (e.g., Gelman & Markman, 1986; Osherson et al., 1990). Medin, Lynch, Coley, and Atran (1997) examined the role of expertise in category-based inference about types of trees. They found that landscapers used a folk taxonomy for a sorting task but a scientific taxonomy as the basis for an induction task. For the landscapers, the scientific taxonomy had more inductive power than their own folk categories, at least for the properties tested. This raises the possibility that people may use different information as the basis for induction than they do in a sorting task. In the present case, perhaps the underlying relation of different senses can serve as the basis for fairly strong inferences, even if it is not used for grouping (see also Proffitt, Coley, & Medin, 2000).

Lin and Murphy (2001) demonstrated that inferences can be drawn from thematic categories, indicating that items do not have to fall into a taxonomic category in order to serve as the basis for inferences. They suggested that this induction was possible because of co-occurrences in their thematic categories. If subjects were told that a cat had a kind of bacteria and asked if kitty litter or a lion were more likely to also have that bacteria, people might select kitty litter because they could imagine a chain of contagion. In contrast, there is little contact between lions and housecats. Of course, this finding depends in part on the particular property tested, which is a well-documented

aspect of category-based induction (Heit & Rubinstein, 1994; Kalish & Gelman, 1992; Ross & Murphy, 1999), and so our experiment also addressed the effect of different properties.

Using a paradigm similar to other studies of induction, subjects in Experiment 6 were told about the presence of a property in an item described in a phrase and were asked to draw inferences to either a polysemous or homonym phrase which shared a word with the first item. Ideally, the property would not depend on connections between the two items, and therefore would be "blank" (Smith, Shafir, & Osherson, 1993). Unfortunately, no real property is truly a blank one, as subjects may have expectations about even quite abstract properties (e.g., only living things can have a disease; only artifacts are made out of metal). For purposes of generalizability, we used three different properties: biological, cost increase, and liking. Although something like the biological property has been used in most past research, the other two properties may be more appropriate for inanimate stimuli or could possibly show a different pattern of induction.

The properties were used to form induction problems using phrases from the earlier experiments, in which a word was used in two different senses. Following is an example of the biological property with polysemous phrases:

Suppose that scientists find the biotin bacteria in wrapping PAPER
Type in the probability (out of 100) that the biotin bacteria will also be in liberal PAPER

For cost increase, the premise was "Suppose there is an increase in the cost of...," and for liking, it was "Suppose that someone likes..." Because subjects probably had prior expectations about the prevalence of these properties, we obtained baseline measures to estimate the base rates of each property in the target category.

Experiment 6 examined whether subjects use information about the relatedness of the different polysemous senses in making inferences, that is, whether the links between the senses play a role in induction. If sense relations influence induction, subjects should use the information given about one sense to draw conclusions about the other sense. Knowing something about *wrapping paper* should therefore affect the way they judge *liberal paper*. Homonyms were included as control materials, following the same reasoning as used in the sorting experiments: The homonym phrases also share a word, but their meanings are unrelated. Thus, they provide a baseline measure of induction just based on sharing a word form.

Method

Baseline measure. The main experiment examined the strength of induction from one kind of object described by a phrase to another kind described by a phrase using the same word. We were not interested in the overall base rate of the property in the target phrase, but rather in how much the attribution of that property varied based on the first phrase. One concern was that differences between polysemous and homonym items could arise for reasons having nothing to do with induction per se: The polysemous items might just happen to be intrinsically more likely to have bacteria in them, be more prone to rising cost, or be more likable. We evaluated this by obtaining measures of the base rates of the properties for the polysemous and homonym phrases.

A baseline experiment was therefore performed on the target phrases, in which subjects were simply asked to estimate the probability that someone would like *liberal PAPER*, for example—or that its price would increase or that it would have a bacteria. Thirty-two subjects performed the baseline judgments.

Subjects. Subjects in the main experiment were 32 new native English speakers at the University of Illinois.

Materials. The 30 polysemous and 24 homonym items used previously were paired with each of three of the properties. We eliminated a few combinations of item and property that did not make sense or seemed inapplicable, using a conservative criterion. For example, it does not seem to make sense to ask if a loud ring has a bacteria. For the increased cost property, four phrase pairs were excluded, and for the bacteria property, six were excluded.[2] Each trial consisted of an induction from one phrase to another that shared the same word. The trial asked subjects to type in a probability as their answer (see example above). Only different-sense or different-meaning pairs of phrases were used. Subjects only saw an item once

[2] This did not reduce the number of items in the item analyses by the same number, because an *item* there was a set of phrases, as described in Experiment 1. If both senses of a word were deleted for a predicate, it was deleted from the analyses; otherwise it was retained.

for each property. Trials were blocked by property, and at the beginning of each block, subjects were given instructions related to that particular property and had a chance to ask any questions to ensure they understood it.

Procedure. Before the experiment, subjects were given instructions about induction and explicitly told to make their judgments assuming that the property was true of the first item, regardless of whether or not they agreed with that. To help them accept that the predicates might indeed be true of the items, subjects were told that they should imagine this was happening in another country with different practices. The instructions provided a rating scale from 0 to 100, where 0 meant that the second statement was impossible, 100 that it was certain, and 50 "means that there is an equal chance that the second statement is true or false, based on the first statement."

Results and discussion

Table 4 shows the mean induction ratings for the two phrase types and three properties. As can be seen, the induction ratings to polysemous words were 15 points higher than the ratings of homonyms, a highly significant difference, $F(1,31) = 70.38,\ p < .001$; $F(1,50) = 23.00,\ p < .0001$. The three properties differed in their overall probability ratings (which was not of interest), and there was also a reliable interaction of property and phrase type, $F(2,62) = 13.11,\ p < .001$; $F(2,100) = 4.56,\ p < .02$. However, the polysemous phrases had higher ratings for all three properties.

Recall that one concern we had was that the base rates of each property might be different for the polysemous and homonym items. However, as Table 4 reveals (second line of each section), the baseline measures were very similar for the two phrase types, differing by less than 4%. To ensure that the observed effects were not due to differences in base rates of the features, for each trial we subtracted the baseline strength of the property for that item from the subject's induction estimate. This will be referred to as the *corrected score*. As shown in the third line of each section in Table 4, many of these scores were in fact negative. The reason for this seems to be that subjects were giving extremely low ratings for items that they felt had no inductive strength, even if the property had some substantial independent probability of occurring. For example, subjects in the baseline study might have given a reasonably high probability that a savings bank will have a cost increase. However, when subjects in the actual induction experiment were asked to judge the probability of a savings bank increasing in cost given that a sandy bank had increased in cost, they might have given a very low probability to indicate the low inductive strength of this argument. Normatively, subjects should have simply given the base rate probability when the items were unrelated. The observed strategy leads to negative induction strengths in many cases and is reminiscent of other findings in the induction literature (Sloman, 1994).

Although these negative values may seem strange, note that the absolute size of the induction measure is not at issue—the important comparison is between the amount of induction for the polysemous items as compared to the homonyms, regardless of whether either is positive or negative. The results, shown in Table 4, revealed that people did draw stronger inferences (an advantage of 12 points) to the polysemous items than to the homonyms in the corrected scores, and this was reliable by both subjects and items,

Table 4
Mean induction ratings in Experiment 6

		Property			*M*
		Cost increase	Biological	Liking	
Polysemous words	Induction	39.24	39.62	41.82	40.23
	Baseline	47.18	30.46	49.55	42.40
	Corrected score	−7.94	9.17	−7.74	−2.17
Homonyms	Induction	23.10	21.12	30.54	24.92
	Baseline	40.15	31.33	44.72	38.73
	Corrected score	−17.05	−10.21	−14.19	−13.81
Polysemous minus homonym corrected score		9.11	19.38	6.45	11.65

Note. Corrected score = Induction rating − baseline rating.

$F(1,31) = 51.72,\ p < .001;\ F(1,50) = 9.56,\ p < .005$. The different properties yielded different inductive strengths, $F(2,62) = 18.47,\ p < .001$; $F(2,100) = 12.94,\ p < .001$. The biological property led to the strongest inferences, the cost increase property was next strongest and liking led to the weakest inferences. In addition, there was also an interaction between the factors of word type (polysemous or homonym) and property, significant by subjects and items, $F(2,62) = 35.67, p < .001;\ F(2,100) = 3.52, p < .05$. However, all three properties again showed the advantage for the polysemous items.

In short, polysemous phrases yielded more induction than the homonym phrases did in both the raw probability ratings and the scores corrected for base-rate differences.

The reanalysis of the data from Experiment 1 showed that different senses that were more similar were also sorted together more. If relative sense closeness is an important factor in how the senses are processed, it should also be correlated with induction. Different senses that are more similar should yield stronger inferences than those senses that are less similar (Osherson et al., 1990; Sloman, 1993). To test this, the similarity ratings of polysemous phrases were correlated with the corrected inference strength scores from this study. A positive relationship between similarity and induction was found for the items using all three properties: cost increase, $r = .55$, $p < .005$; biological, $r = .51$, $p < .005$; and liking, $r = .55$, $p < .005$. (Obviously, the analysis was not carried out for homonyms, which were highly dissimilar.) The closer two senses were, the stronger the induction from one to the other, and this was true for all three properties.

The results for induction are consistent with the results found in the sorting task. First, the absolute level of induction for polysemous items was very low. As Table 4 shows, there was no overall increase in subjects' ratings as a result of being provided the property for a different-sense phrase (i.e., the difference between the induction and baseline measures was virtually 0). Thus, the absolute strength of this induction was quite low, though one should not put too much emphasis on the absolute numbers of such a scale. The second finding, that polysemous items had stronger induction than homonyms, is not subject to this concern. The fact that polysemous items had a higher induction judgment shows again that subjects perceived them as being somewhat related compared to an unrelated control. However, that effect was again rather small—only 11.6% in the corrected scores. This is in the same ballpark as the categorization difference found in Experiment 3, viz., about 6%. Third, both sorting and induction were moderately correlated with the similarity of the phrases. Thus, these very different techniques converge on the conclusion that polysemous senses are perceived as being more related than homonyms are, but only slightly more so. It is significant that this result was found in a task that did not ask subjects to make a meta-conceptual judgment of whether the items were in the same category. Instead, subjects had to use whatever similarity or relation they knew between the items to make an inference about shared properties, which is the kind of judgment that everyday category membership is used for.

General discussion

Before discussing the theoretical implications of these results, we note that the experiments have fulfilled one of the main goals of this research, namely to explain the somewhat puzzling results of Klein and Murphy (2001). In those experiments, we found that there was no priming from one use of a polysemous word to a use involving a different sense: In memory experiments, a word used in one sense did not serve as a cue to recall the word used in a different sense; in semantic judgment tasks, the different senses appeared to interfere with one another. These results flout the intuition that the senses of a polysemous word are highly similar, which is why they receive the same name. If the senses are not similar, then why are they picked out by the same name? And if they are similar, then why isn't there stronger priming between them?

The present results help to explain the earlier RT and memory experiments. First, the results showed that the amount of conceptual similarity of the polysemic senses is indeed quite low. Subjects did not group together phrases containing different senses of a word, although they did categorize together phrases that used the words in the same sense. Neither did subjects form strong inductions from an item described by the word in one sense to an item described by the word in a different sense. Knowing that one kind of paper had a property did not raise subjects' estimates that another kind would have it very much. In short, Klein and Murphy's (2001) proposal for the lack of priming in their experiments has received considerable support in these experiments: The senses of poly-

semous words are not very similar, which would explain why they are not useful as memory cues or as primes for one another.

This discovery does not yet explain why it is, then, that the different senses have the same name. If they are so different, why aren't they lexically distinguished? We have two answers. The first is that although the overlap of polysemous senses is small, it is not zero. In both the categorization and induction tasks, we found a small but consistent advantage for polysemous items over unrelated homonyms, suggesting that there is at least some conceptual overlap, which could partly explain the historical process by which words are extended to new meanings.

The second answer is that words may be extended via relations other than similarity, and these relations may not provide a processing advantage. Two uses of a word may have an obvious relation, but those uses may nonetheless have very different contents. As pointed out before, if one is looking for paper to write on, the concept of a newspaper publishing company will not be helpful in directing the search; if one needs new glasses to aid in reading, buying champagne glasses will not be very useful; and if one is looking for a tin to put cookies into, knowledge of the 50th element is superfluous. Historically, however, the fact that containers of a certain kind were made out of tin seemed to be a good enough reason to call them *tins*, and people could readily work out this connection and understand why these objects received this name. Pustejovsky (1995) describes in some detail the way in which such relations may cause polysemy. He argues that our knowledge of a complicated object such as a newspaper includes a variety of information, such as its being produced by a set of editors and writers, who are paid by a publishing company, with the product itself being printed on paper. He proposes that a process of construal allows the name for this whole complex of information to be extended to certain critical parts of it, such as the publishing company or the content of the newspaper.[3] However, there is nothing in that process that ensures that the different uses of this word will refer to conceptually similar entities. In fact, the opposite is often the case, as important components of an object, such as its substance, function, or maker, are usually not the same kind of thing as the object itself.

In short, although polysemous words may well gain their senses by extension from an existing sense to a novel, related sense, that process does not result in a set of similar senses, but instead in a set of senses that are pairwise related. The processing results of earlier research combined with the categorization and induction results of the present experiments converge on this explanation.

In contrast, the notion that there is a substantial core meaning that is constant across senses (Caramazza & Grober, 1976; Ruhl, 1989; Schreuder & Flores d'Arcais, 1989) does not pass muster as a theory of polysemy. The present experiments give strong evidence against this view. First, the very low rate of categorizing together the different senses of a single word show that the senses are even less related than superordinate-level taxonomic categories, a level with a low functional amount of similarity (Rosch et al., 1976). Second, the weakness of induction across senses verifies this conclusion using a different measure. Third, the core meaning theory must claim that individual senses are derived by a productive semantic process when each word is encountered. However, we could not find any evidence of this process applied to different words (Experiment 4) or even the same word in different phrases (Experiment 5). We are not denying that people are able to productively use the animal-meat or object-content patterns of polysemy, for example. However, the results suggest that the active use of such a rule may be confined to novel or very infrequent senses of a word, such as those tested by Murphy (1997). If a new form of information storage called a *nanodisc* comes into being, people will no doubt use this pattern to say things like, "I saw a great nanodisc last night." However, if they say such things a few hundred times, they will store the different senses (the object and its content) separately, so that both senses may be retrieved rather than having to be constructed from a core meaning.

The structure of polysemous categories

In the Introduction, we discussed the fact that there are different sorts of categorical relations,

[3] If we take this process as intending to describe the online derivation of senses during word use, then Pustejovsky's view would be a form of the single- sense view and so would be subject to the criticisms we have made of that position, here and in Klein and Murphy (2001). However, one could take this construal process as happening historically and/or during language acquisition, which would not conflict with our claims about lexical representation.

and polysemous senses might fit any of them. Although our experiments did not attempt a content analysis of these relations, our results still speak to the different possibilities that we raised earlier. We have concluded that different senses of a word are probably related but are not generally similar. How does this fit with the different forms of category structure discovered in past research?

Taxonomic categories are the most prominent form of category relation, containing objects that are similar to one another, to at least some degree (see Lin & Murphy, 2001, for discussion). Unlike taxonomic category members, polysemous senses can refer to very different kinds of things. This claim is strongly supported by our own sorting and inference data, which found little strength to the polysemous categories. As taxonomic categories provide a basis for both sorting (as shown by the taxonomic choices in Experiments 1, 4, and 5, as well as innumerable categorization studies) and inference (Osherson et al., 1990), polysemous senses do not appear to form taxonomic categories.

Perhaps polysemous senses are organized thematically. After all, there are many polysemous extensions that are thematic in nature, such as *paper* being used to refer to either a newspaper or the company that publishes a newspaper. The problem, however, is that the entire set of a word's senses do not appear to form a coherent thematic category. One cannot unify wood pulp, textual meaning, wall covering, a publisher, and an oral presentation (see Table 1) in a single thematic relation. Instead, most senses seem to share a thematic relation to one or two of the other senses, and these thematic relations are themselves different (e.g., the sheets of paper are made from the wood pulp, but the publisher is not made out of the publication). Thus, it does not seem correct to call polysemous senses a thematic category. A very similar argument applies to ad hoc categories, as different senses do not appear to share a common goal or purpose.

Polysemous categories seem more similar to the chained categories discussed by Lakoff (1987) and Malt et al. (1999); and (see also Heine, 1992). The members of chained, or radial, categories can be linked in very different ways. Polysemous relations are also very diverse (see Nunberg, 1979), and the senses of one polysemous word can extend in many different directions. It seems probable that new uses for polysemous words are created by a process of chaining from a known sense (Cruse, 1986; Murphy, 1997; Sweetser, 1990). The different members of the chain are related in different degrees and by a variety of different relations. So far, there is no evidence that the links connecting each usage are stored (see Experiments 4 and 5). The different senses of a polysemous word appear to develop historically through various relations, but those relations may not themselves be lexically represented.

In summary, the different senses of a polysemous word do not appear to correspond to a unified taxonomic, thematic, or ad hoc category. This is not to say that they are totally unconstrained and unstructured, but that the structure appears to be somewhat weak and to apply on a pairwise basis rather than to the entire set of senses. Such a proposal is similar to the idea of a radial category (Lakoff, 1987), but one might question whether radial categories are categories in any real sense. If the members themselves are not similar (e.g., wood pulp, talks, and publishers), and if they are not unified by a common relation, then they do not appear to fit the basic requirement of category membership, that members have properties in common (Murphy, 2002; Smith & Medin, 1981). If senses accrue over time by chaining to a previously existing sense, the degree to which any two senses of a word are related and the nature of their relation would depend greatly on the particular pair of senses. On this view, it is not surprising that different senses appear to be represented separately.

This conclusion may lead one to question whether the conceptual approach we have taken is in fact ultimately helpful. If the different senses do not form a single coherent category, then why should one consider the conceptual relations among the senses? Part of the answer to this is empirical. We found that both sorting and priming were predicted by the similarity of the different senses. Thus, even though the entire set of senses of a word may be diverse, the conceptual similarity of pairs of those senses may determine aspects of their use. Furthermore, although many senses may not be similar, all theories seem to agree that the extension of a word from one sense to a novel one requires some kind of conceptual similarity, analogy, or relation. Without that constraint, the senses of a word could be completely random.

What is surprising about polysemy as a general phenomenon is that the great diversity of senses does not impair fluent understanding of everyday language. It doesn't seem to bother listeners that the word *paper* might refer to a company, some

text, an oral presentation, or some stuff. Why it doesn't is an important question for models of comprehension, as we shall discuss next.

Implications for process models of word meaning

Research on polysemy has implications for more general issues of how word meaning is represented. Perhaps the most basic assumption in the field is that each word is linked to "its meaning." In many cases, the assumption of a unitary meaning is unstated, but the possibility of a word having numerous different senses is simply not considered. For example, pictorial representations of lexical structure often depict each word as being connected to its single meaning (e.g., Levelt, Roelofs, & Meyer, 1999, p. 4). Although such depictions are clearly simplifications of a theory, they foster the assumption that most words have "a" meaning and hide the problem of coordinating multiple related senses.

The main question that arises is, when a word is read and its meaning activated, what is actually activated? If the word were completely unambiguous, then its entire meaning could be activated. However, as most frequent words are polysemous, we must decide whether all the different senses are activated, or only one or two. Traditional linguistic and lexicographic models of polysemy would have assumed that a basic or core meaning is activated, with other senses derived as necessary. In their studies of eye movements while reading sentences with polysemous words, Frisson and Pickering (1999) and (Pickering & Frisson (2001)) suggest that polysemous words initially activate an "underspecified" sense that contains elements common to the different senses. Then, "Once readers have used this underspecified meaning to assign a (rather abstract) semantic value to an expression, they can home in on the intended sense by instantiating any underspecified features" via context (Frisson & Pickering, 1999, p. 1379). Frazier and Rayner (1990) drew a similar conclusion.

However, our results show that if there is a core, it has minimal content. As we have mentioned many times, there is little semantic overlap between some senses, and subjects' performance in classification and inference suggests that they see little commonality across senses. As a result, it is not clear what the underspecified meaning could be. If it was intended to be a substantive, informative semantic representation, then the present results (along with those of Klein & Murphy, 2001) would be strongly inconsistent.[4] There are other possibilities of what an underspecified meaning could be, however, that might be more consistent with our own results.

One possibility, *radical under specification* is that a kind of neutral placeholder is activated until disambiguating information is encountered. For example, if someone reads "The paper...," the initial representation may be little more than "something called paper," with no commitment to any particular interpretation of the word. As later context emerges ("The paper reported in yesterday's edition...") then one sense would be selected. Radical underspecification would be consistent with our own claim that there is little overlap across senses. What is unclear, however, is why this should happen with polysemous words and not with homonyms. Current thinking on homonyms suggests that readers quickly activate both meanings or select one based on frequency and context (e.g., Duffy, Morris, & Rayner, 1988), as discussed by Pickering and Frisson (2001). Why should readers select one meaning of homonyms but not do so for polysemous words, whose senses are more related than homonym meanings?

A different version of underspecification is a more promiscuous approach in which many different aspects of a word's meaning are weakly activated (in the absence of disambiguating context). This diffuse activation would not select any single sense but instead prepare all of them for later selection by context. For example, for *paper*, information about the substance, writing material, text, newspaper, and oral presentations could all be weakly activated, perhaps proportionally to their frequency. This activation would be underspecified in the sense that no particular sense is picked out, but it would not be a core meaning. This kind of processing is reminiscent of what has been proposed for the comprehension of nonliteral language: Connotative and metaphoric aspects of

[4] For example, Klein and Murphy (2001) found that processing a word used in one sense slowed the processing of the word in a different sense. Although there are a number of possible ways to account for such an effect, most of them seem to require that readers have representations of the separate sense, which are then activated, inhibited, preserved in working memory, or whatever, as a result of priming. Furthermore, an underspecified representation of polysemous words would run into the problem identified by linguists of idiosyncratic and unpredictable senses, which require explicit, detailed listing, since they cannot be derived from general principles (Lehrer, 1990; Rice, 1992).

a word's meaning may be initially weakly activated by spreading activation (especially in the right hemisphere), and they are selected if other words in the context activate the same properties (see Beeman, 1998; Brownell, 2000).

Our experiments did not address the question of which senses are activated, in what order, and to what degree, and so we cannot address this question further. Nevertheless, we have argued that our results do speak to proposals for how different senses are processed, and in particular seem to cause problems for the view that a core or common sense is initially activated. However, we should note that our procedures and materials are very different from the few studies of polysemic processing. For example, our online measure in Klein and Murphy (2001) was priming in a sensicality task, whereas Pickering and Frisson (2001) and Frazier and Rayner (1990) used eye tracking during reading as their measure. Thus, making comparisons and drawing conclusions across studies is difficult at this point.

What *is* clear, however, is that the field needs to address issues of polysemy more vigorously. For example, note that the prototypical example of a homonym, *bank*, is also systematically polysemous. The financial meaning of *bank* is used to refer both to an institution (*The bank is broke*) and a building (*The bank burned down*). The physical meaning of *bank* can refer either to the sloping side of a river or to a piled up mass (*The security guard viewed a bank of monitors*). In experiments on homonyms, subjects are not asked to distinguish the different senses of *bank*, but in real life, listeners must not only choose the correct meaning, but determine which sense is intended within that meaning. Although homonyms have been very extensively studied, the field has not yet developed a complete proposal for how the full meaning of a word is derived. More generally, what is most needed now is a better understanding of which aspects of word meanings are stored and which are constructed on the basis of context and pragmatic inference (Gerrig, 1986).

The study of polysemy requires a unified account of the detailed lexical structure of polysemous words, the conceptual relations of the different senses, the learning processes involved in acquiring and distinguishing those senses, and the online processes of accessing and constructing word meaning from that structure when the word is used in a particular context. Our own investigations have made a dent in the questions of conceptual relations and online processing, but much more has yet to be learned about how polysemous senses are acquired and represented.

The polysemy–homonymy distinction

We noted at the beginning of this article that it is important to keep polysemy distinct from homonymy. However, one might ask whether our own results cast doubt on this claim. Given the small differences found between polysemous and homonymous phrases and the lack of priming across polysemic senses (Klein & Murphy, 2001), perhaps there is no reason to distinguish them.

This question has both a terminological and a theoretical aspect. As to terminology, let us immediately say that it is necessary to maintain the distinction if only because it is a widely held one in linguistics and lexicology. Furthermore, our results speak to some aspects of the distinction but not to others, and those others may require that the distinction be preserved. For example, Ravin and Leacock (2000, p. 2) argue, "The distinction between polysemy and homonymy is important because it separates the principled from the accidental and poses the following questions: If different senses of polysemous words are systematically related, how do they derive from each other, and how should they be organized to reflect this regularity?" One could easily have predicted in advance that the same word, *CD*, would be the name for a format, a disk, and the content of that disk, but the fact that *bank* refers to a financial institution and the side of a river could not have been predicted. Although the senses of polysemous words can be dissimilar, they develop through predictable processes, and those processes may be known and used by individual speakers (Murphy, 1997). Thus, even if homonyms and polysemous words are represented similarly, they are distinct in their historical development and perhaps in the way language learners acquire them.

Furthermore, it is important not to exaggerate the separation of polysemous senses in our results. For example, we found that more similar senses were sorted together and served as a basis for induction. In addition, we chose polysemous senses that were clearly distinct in meaning. For example, we did not use type-token polysemy (see (5) above), which naive subjects might not even identify as being different senses. Nor did we use subtle differences that Cruse (2000) calls *ways of seeing*, in which different aspects of the same word are emphasized depending on the perspective of the speaker (and see Anderson & Ortony, 1975).

It is widely recognized that polysemous senses range from nearly identitical to nearly unrelated, and it seems likely that the latter are processed much like homonyms are, but the former are not. Therefore, the study of polysemy will likely continue to distinguish the phenomenon from homonymy but also acknowledge that the line between the two is not always clear.

References

Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles—a problem of polysemy. *Cognitive Psychology, 7*, 167–180.

Annett, M. (1959). The classification of instances of four common class concepts by children and adults. *British Journal of Educational Psychology, 29*, 223–236.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 3*, 211–227.

Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 1–64). San Diego: Academic Press.

Beeman, M. (1998). Coarse semantic coding and discourse comprehension. In M. Beeman & C. Chiarello (Eds.), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience*. Mahwah, NJ: Erlbaum.

Brownell, H. (2000). Right hemisphere contributions to understanding lexical connotation and metaphor. In Y. Grodzinsky, L. P. Shapiro, & D. Swinney (Eds.), *Language and the brain* (pp. 185–201). San Diego: Academic Press.

Caramazza, A., & Grober, E. (1976). Polysemy and the structure of the subjective lexicon. In C. Rameh (Ed.), *Semantics: Theory and application. Georgetown University Round Table on languages and linguistics* (pp. 181–206). Washington, DC: Georgetown University Press.

Clark, E. V. (1983). Meaning and concepts. In J. H. Flavell & E. M. Markman (Eds.), *Manual of child psychology: Cognitive development* (Vol. 3, pp. 787–840). New York: Wiley.

Clark, E., & Clark, H. (1979). When nouns surface as verbs. *Language, 55*, 767–811.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, & Computers, 25*, 257–271.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 241–248.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.

Cruse, D. A. (2000). Aspects of the micro-structure of word meanings. In Y. Ravin & C. Leacock (Eds.), *Polysemy: Theoretical and computational approaches* (pp. 30–51). Oxford: Oxford University Press.

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production system. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 318–344.

Deane, P. D. (1988). Polysemy and cognition. *Lingua, 75*, 325–361.

Dickens, C. (1983). *The Pickwick papers*. New York: Bantam Books [Original work published 1867–1868].

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*, 429–446.

Fellbaum, C. (2000). Autotroponomy. In Y. Ravin & C. Leacock (Eds.), *Polysemy: Theoretical and computational approaches* (pp. 52–67). Oxford: Oxford University Press.

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language, 29*, 181–200.

Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1366–1383.

Gagné, C. L. (2001). Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 236–254.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183–209.

Gerrig, R. J. (1986). Process and products of lexical access. *Language and Cognitive Processes, 1*, 187–195.

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

Heine, B. (1992). Grammaticalization chains. *Studies in Language, 16*, 335–368.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 411–422.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child*. New York: W.W. Norton.

Kalish, C. W., & Gelman, S. A. (1992). On wooden pillows: Multiple classification and children's category-based inductions. *Child Development, 63*, 1536–1557.

Kawamoto, A. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language, 32*, 474–516.

Kay, P. (1992). At least. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization* (pp. 309–331). Hillsdale, NJ: Erlbaum.

Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language, 45*, 259–282.

Klein, D.E., & Shoben, E.J. (May, 1998). *The effects of context on conceptual combination.* Paper presented at the Midwestern Psychological Association Conference, Chicago, IL.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

Langacker, R. W. (1987). *Foundations of cognitive grammar* (Vols. 1 & 2). Stanford, CA: Stanford University Press.

Lehrer, A. (1990). Polysemy, conventionality, and the structure of the lexicon. *Cognitive Linguistics, 1*, 207–246.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts and categorization. *Journal of Experimental Psychology: General, 130*, 3–28.

Luria, A. R. (1976). *Cognitive development: Its cultural and social foundations*. Cambridge, MA: MIT Press.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing vs. naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language, 40*, 230–262.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology, 32*, 49–96.

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

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

Murphy, G. L. (1991). Meaning and concepts. In P. Schwanenflugel (Ed.), *The psychology of word meaning* (pp. 11–35). Hillsdale, NJ: Erlbaum.

Murphy, G. L. (1997). Polysemy and the creation of novel word meanings. In T. B. Ward, S. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 235–265). Washington, DC: American Psychological Association.

Murphy, G. L. (2001). Causes of taxonomic sorting by adults: A test of the thematic-to-taxonomic shift. *Psychonomic Bulletin & Review, 8*, 834–839.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Murphy, G. L., & Lassaline, M. E. (1997). Hierarchical structure in concepts and the basic level of categorization. In K. Lamberts & D. R. Shanks (Eds.), *Knowledge, concepts and categories* (pp. 93–131). Cambridge, MA: MIT Press.

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy, 3*, 143–184.

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

Pickering, M. J., & Frisson, S. (2001). Processing ambiguous verbs: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 556–573.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 811–828.

Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.

Ravin, Y., & Leacock, C. (2000). Polysemy: An overview. In Y. Ravin & C. Leacock (Eds.), *Polysemy: Theoretical and computational approaches* (pp. 1–29). Oxford: Oxford University Press.

Rice, S. A. (1992). Polysemy and lexical representation: The case of three English prepositions. In *Proceedings of the fourteenth annual conference of the Cognitive Science Society* (pp. 89–94). Hillsdale, NJ: Erlbaum.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14*, 665–681.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology, 38*, 495–553.

Ruhl, C. (1989). *On monosemy: A study in linguistic semantics*. Albany: State University of New York Press.

Schreuder, R., & Flores d'Arcais, G. (1989). Psycholinguistic issues in the lexical representation of meaning. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 409–436). Cambridge: MIT Press.

Sharp, D., Cole, M., & Lave, C. (1979). Education and cognitive development: The evidence from experimental research. *Monographs of the Society for Research in Child Development serial #148, 44*(1–2).

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25*, 231–280.

Sloman, S. A. (1994). When explanations compete: The role of explanatory coherence on judgements of likelihood. *Cognition, 52*, 1–21.

Smiley, S. S., & Brown, A. L. (1979). Conceptual preference for thematic or taxonomic relations: A non-monotonic age trend from preschool to old age. *Journal of Experimental Child Psychology, 28*, 249–257.

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

Smith, E. E., Shafir, E., & Osherson, D. (1993). Similarity, plausibility, and judgments of probability. *Cognition, 49*, 67–96.

Sweetser, E. E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.

Tuggy, D. (1993). Ambiguity, polysemy, and vagueness. *Cognitive Linguistics, 4*, 273–290.