Prototypicality in Sentence Production

Kristine H. Onishi,
McGill University

Gregory L. Murphy, and
New York University

Kathryn Bock
University of Illinois at Urbana-Champaign

Abstract

Three cued-recall experiments examined the effect of category typicality on the ordering of words in sentence production. Past research has found that typical items tend to be mentioned before atypical items in a phrase—a pattern usually associated with lexical variables (like word frequency), and yet typicality is a conceptual variable. Experiment 1 revealed that an appropriate conceptual framework was necessary to yield the typicality effect. Experiment 2 tested ad-hoc categories that do not have prior representations in long-term memory and yielded no typicality effect. Experiment 3 used carefully matched sentences in which two category members appeared in the same or in different phrases. Typicality affected word order only when the two words appeared in the same phrase. These results are consistent with an account in which typicality has its origin in conceptual structure, which leads to differences in lexical accessibility in appropriate contexts.

When people talk, the order of their words matters: *Lee inspired Pat* does not mean the same thing as *Pat inspired Lee*. How do people decide which words to say first? The syntax of English provides one constraint, as some words need to be the subject or object, given the meaning to be communicated. But English sometimes provides alternative ways to convey approximately the same meaning. For example, English allows us to say *Lee inspired Pat* and *Pat was inspired by Lee*. Passive sentences differ in structure and word order, but they provide the same basic information as the corresponding actives. Options may also differ in word order alone. Saying either *Please bring some apples and kiwis to the party* or *Please bring some kiwis and apples to the party* would probably lead to the same outcome (two sorts of fruit at the event), even though the order of words is different. In this article, we examine the factors that affect word order when order matters to sentence structure and whether those factors are the same when order does not greatly affect structure.

One principle determining word order is that things that are easier to say tend to get said earlier (Bock, 1982). There are at least two ways in which things can be easier to say: It can be easier to access the concept, or it can be easier to access the word that refers to the concept (Bock, 1987a; Clark & Clark, 1977). Both of these factors, called respectively *conceptual* and *lexical accessibility,* have been shown to affect the ordering of words in sentences (Bock, 1977; 1986; 1987a,b; Bock & Irwin, 1980; Bock & Warren, 1985; Kelly, Bock, & Keil, 1986;

McDonald, Bock, & Kelly, 1993; Prat-Sala & Branigan, 2000). Conceptual accessibility refers to the fact that certain kinds of concepts have prominence within the conceptual system and therefore references to them are more likely to occur as subjects than as direct objects (McDonald et al., 1993) and more likely to occur as direct or first objects than as indirect or second objects (Bock & Warren, 1985). Properties that make a concept more prominent include recent mention, concreteness, and animacy (Bock, Loebell, & Morey, 1992; Bock & Warren, 1985; Clark & Begun, 1971; McDonald et al., 1993). Concrete, animate entities reside in a rich conceptual network. People are more likely to talk about them, their representation is more easily activated, and they are more likely to end up as the subjects of sentences, where speakers put the things they are talking about.

Lexical accessibility has to do with the ease of retrieving a word. It is affected by frequency, with frequent words being accessed more rapidly than less frequent ones (Oldfield & Wingfield, 1965). Similarly, shorter words are generally produced more quickly than longer ones (Roelofs, 2002; Santiago, MacKay, & Palma, 2002). All things being equal, words that are more accessible tend be placed earlier within phrases (Bock, 1982; Kelly, 1986). This effect can be seen in common expressions in which people prefer to say shorter and more frequent words first, as in *salt and pepper* compared to *pepper and salt* (Cooper & Ross, 1975; Fenk-Oczlon, 1989; Pinker & Birdsong, 1979). High lexical accessibility may enable a word to be said earlier within a phonological planning unit, a phrase, without causing the restructuring of a sentence, presumably because the major determinants of sentence structure (specification of subjects, direct objects, and so on) have already been set by the time the word forms are selected (Bock & Warren, 1985). An entity with an accessible label is not necessarily more animate, interesting, salient, or relevant, and so it is not necessarily mentioned in the subject position (Bock & Warren, 1985; McDonald et al., 1993). In short, lexical variables associated with accessibility appear to influence word order within phrases more readily than the order of major sentence constituents.

One particular variable that has been shown to influence word order is category typicality. Although typicality is generally thought of as a conceptual variable (Murphy, 1991; Rosch & Mervis, 1975), there is some reason to believe that it affects language production through lexical accessibility (Kelly et al., 1986). In the present experiments, we investigated the nature of typicality's influence on word order. The results of the experiments speak to issues of conceptual and lexical representation, as well as to mechanisms determining word order in sentence production. We begin by briefly reviewing the phenomenon of category typicality and prior explanations of how it affects word order. We then describe three experiments and several follow-ups that investigated when and how typicality affects both the order of words in phrases and the structure of sentences.

Category Typicality

Category typicality is a measure of the goodness of category membership (Rosch, 1975). Some things are judged better (more typical) members of a particular category than others, even if both are judged to be category members. For example, apples are judged to be more typical fruit than lemons are. Typicality does not simply reflect intuitions about category members, as it influences virtually every task involving categories, including classification, induction, language comprehension, and category learning (see Murphy, 2002, ch. 2 for review). Goodness of membership is presumed to be about concepts—mental representations about classes of entities in the world—and not about the words that refer to them. For example, typicality effects are found in artificial categories even when category members do not have lexical labels (e.g., Medin & Schaffer, 1978; Rosch & Mervis, 1975), and learning of novel categories is easier with typical examples (Mervis & Pani, 1980). Typicality also influences

processing in preverbal infants (Bomba & Siqueland, 1983). Thus, typicality strongly influences performance even outside of linguistic contexts.

Judgments of typicality (Barsalou, 1983, 1985, 1987, 1991; Roth & Shoben, 1983) and the effects of typicality on inferences made during reading (Garrod & Sanford, 1977; McKoon & Ratcliff, 1989; Roth & Shoben, 1983) are affected by context. That is, both explicit (judgment) and implicit (reading) measures of typicality are influenced by how the entity in question is thought of, even when its label remains constant. Thus, typicality in these instances cannot be represented as a property of particular lexical entries, suggesting that typicality has to do with the underlying conceptual representation.

Finally, the accepted explanation of typicality effects has to do with the relations of conceptual representations. A robin is a typical bird because it possesses the properties found in other birds and does not share properties with members of contrasting concepts. Atypical birds like penguins lack many properties found in birds (flies, perches on trees, migrates, small size) and possess properties that are usually found in other categories (eats fish, swims, lives in arctic climate, wears tuxedo). Typical items generally have high family resemblance of this sort (Barsalou, 1985; Rosch & Mervis, 1975). Another determinant of typicality is how well the item fits the goal or ideal associated with the category, if there is one (Barsalou, 1985; Proffitt, Coley, & Medin, 2000). Both of these determinants of typicality have to do with how well an item fits the representation of the category in semantic memory. They do not involve linguistic or lexical properties of a category's name. Thus, it is widely assumed that typicality effects in tasks using words (reading, comprehension, ratings, induction, and so on) derive from relations among the conceptual representations associated with those words, rather than from specifically linguistic representations (see Murphy, 2002, ch. 10 for a complete argument).

However, some evidence suggests that typicality has a lexical component that affects language production. Kelly et al. (1986) investigated typicality effects in sentence production. During a study phase, people heard questions and associated answers containing category members, one highly typical and one less typical (e.g., apples and lemons). In a test phase, they heard just the questions and wrote the associated answers. Some sentences contained references to the category members within a single phrase embedded in a sentence (e.g., *The child's errand was to buy an apple and a lemon at the fruit stand*). The other sentences contained the references to a pair of category members presented in different major constituents. The entities were presented either in active (*Sears Roebuck reported that shirts outsold hats in their clothing department*) or passive (*Sears Roebuck reported that hats were outsold by shirts in their clothing department*) sentences. The highly typical things (apple, shirts) occurred either first or second in the sentence.

In producing the remembered sentences, when participants changed the word order, they tended to put the highly typical entities earlier when the entities had been presented in a single constituent, but did not do so when the entities were in different major constituents. For example, a phrase referring to an apple and lemon would be more likely to be produced as *an apple and a lemon* than *a lemon and an apple*, but when the entities were presented in different major constituents (e.g., *hats outsold shirts* or *shirts were outsold by hats*), there was no favored order of production. So, typicality affected the order of words within a phrase but did not have a reliable effect on the reordering of major sentence constituents. This pattern of results is associated with lexical variables such as frequency and length (e.g., Bock & Warren, 1985) rather than more conceptual variables such as concreteness and animacy (McDonald et al., 1993). Thus, Kelly et al. (1986) argued that typicality was having its effect through lexical accessibility.

Based on the concepts literature and Kelly et al.'s (1986) results, we can identify two very different hypotheses about how typicality could be influencing word order. The first (from Kelly et al.) is that typicality has some direct effect on the linguistic representation of the word, such as its inherent activation level or prominence. Indeed, words such as *robin* could simply be tagged as "typical," which could affect order during production. This hypothesis, the *lexical account*, argues that words referring to typical and atypical entities are differentially accessible by virtue of their typicality. In contrast, the *conceptual account* says that concepts underlying typical words are more similar to the concepts of their categories (e.g., robin-bird) than are concepts of atypical words (e.g., penguin-bird), and that this indirectly influences word order.

These two accounts have complementary problems in explaining past data. The lexical account can explain Kelly et al.'s (1986) data on typicality and word order but is inconsistent with explanations of typicality itself, which cast it as conceptual rather than lexical. The conceptual account is consistent with that past literature, but it has no ready explanation for why typicality's effect on word order is like those of lexical variables such as frequency and length.

There is a compromise account. Perhaps the apparent lexical nature of typicality's effect on the ordering of words comes about indirectly, through the activation of the superordinate concept which in turn activates the lexical information for category members. That is, perhaps both accounts are partly correct. For example, when one talks about apples and lemons, the superordinate concept of fruit is likely to be activated (Barsalou, 1982). In Kelly et al.'s experiment, the superordinate was explicitly mentioned in the target sentence, ensuring that it was readily available. Perhaps during the test, people recalled that the sentence had to do with someone buying fruit at the stand. With fruit as a cue, highly typical apples are likely to come to mind faster than less typical lemons. Once the concept of apples is activated, the word *apple* would become accessible for insertion into the sentence, leading the speaker to be more likely to say *an apple and a lemon* than the reverse. Figure 1 illustrates this effect. On this explanation, the effect of typicality is driven by processing at the conceptual level, which in turn causes differences in lexical activation, leading to within-phrase order differences. (Note that Figure 1 glosses over different levels of lexical representation, representing lexical entries as undifferentiated wholes. This is an oversimplification of lexical structure that we use only to highlight the conceptual underpinnings of the compromise account.) We discuss this proposal in more detail after presenting the experiments.

However, a less interesting explanation for the apparent lexical effects in Kelly et al. (1986) refers to their method of measuring typicality. They used production norms (Battig & Montague, 1969), which are themselves a language production measure. As Kelly et al.'s dependent measure was also production, there is a danger of circularity in using production frequency as an independent variable to predict output order.

The Present Study

Determining the locus at which category typicality influences word order in sentences will illuminate the interaction of linguistic and nonlinguistic information in language use and provide important information about what aspects of categories are represented conceptually and lexically. We performed three studies examining the nature of typicality effects on word order, attempting to gain independent evidence for their lexical or conceptual bases. Furthermore, we re-examined the finding that typicality has an influence on word order within phrases but not in determining sentence structure, since this was a critical piece of data in the earlier argument.

In Experiment 1, we looked at whether the typicality effect required evoking an appropriate conceptual framework. Since typicality at the conceptual level is always computed relative to

some concept (a robin is a typical bird but not a typical pet), the conceptual account predicts that typicality should affect word order only in an appropriate conceptual framework. In contrast, if the words referring to typical entities like apples and robins are simply more accessible, then they should be produced early whether or not there is an appropriate conceptual framework. In this experiment, we looked only within phrases, since typicality effects had previously been found there. In Experiment 2, we sought typicality effects for items that are not associated with familiar categories in long-term memory, which would not have a pre-existing lexical advantage. Here the conceptual account predicts typicality effects but the lexical account does not.

The final study examined whether typicality affects the ordering of major sentence constituents. Kelly et al. (1986) found that typicality did not affect constituent reordering when this reordering depended on the change from passive to active sentences. However, passives and actives differ in more than just the ordering of their major constituents. Furthermore, Kelly et al. used entirely different materials in their comparison of ordering within conjunctions and ordering of sentence constituents. We used structures that were more similar than those used in previous work but still required a change in constituent order for reordering of the target words. A finding that typicality affects the ordering of major constituents would provide evidence of typicality affecting conceptual accessibility, thereby supporting the conceptual account. We also directly measured typicality without using production norms, reducing the potential for circularity in explaining production order as well as reducing the possibility of effects due to factors other than typicality.

Experiment 1

The goal of Experiment 1 was to contrast the lexical and conceptual accounts of typicality effects in sentence production. We measured word order within phrases, since that is the level at which it would be possible to see lexical accessibility effects, and we manipulated the relevance of superordinate categories in order to vary conceptual information.

Participants heard sentences that contained two category members, one typical and one less typical, and then recalled the sentences. In one condition, the two items were members of the same category (e.g., both fruit or both clothing), and in the other condition, they were from different categories (e.g., one fruit and one item of clothing). Our assumption was that mentioning two fruits would make the category of fruit quite salient, whereas mentioning one fruit and one article of clothing would not. (Barsalou & Ross, 1986, and others have suggested that superordinate category membership is automatically activated by basic-level category names. However, our intuition was that presenting items from two different categories in a single phrase would not cause both their superordinates to be encoded into the sentence's memory representation, because neither superordinate would describe the entire phrase. The results will serve as a test of this intuition.) On the conceptual account, there should be an interaction between the tendency to put typical things early and whether the category is evoked: Typical things should go earlier in the phrase only when the category is evoked. When one is not thinking of fruit in general, the typicality of apples should not give the word *apple* any advantage, because the similarity of the apple concept to the fruit concept is not computed. Under a purely lexical account, there should be no such interaction: Because typicality has its effect on the lexical representation, the highly typical item should go earlier in the phrase regardless of whether the category is evoked.

The prediction of the conceptual approach raises the issue that previous demonstrations of the typicality effect have not relied on such interactions but rather on main-effect differences in which typical items tend to be placed before less typical ones. The problem with such comparisons is that they are correlational with respect to the critical typicality variable: Items are identified as being typical and less typical, and their relative positioning in the sentence is

measured—typicality of the item is not directly manipulated. Kelly et al. (1986) did carefully control other variables that they believed would influence positioning of the words such as length, frequency, and prosody (which we also controlled). Nonetheless, the conclusion of such studies relies on the assumption that typicality is the only variable that varies systematically across conditions, and (as with any correlational study) it is impossible to be certain that this was the case. It is unknown if there is some other factor that makes people prefer apples over lemons, for example.

Therefore, the present experiment makes the additional contribution of experimentally varying typicality while keeping the items constant. In the same-category condition, if conceptual information is necessary, *apple* will tend to be placed first, whereas in the different-category condition, this tendency should be less apparent. This prediction does not rely on the overall preference for typical over less typical items but directly manipulates a conceptual variable underlying typicality differences. As such, the results will greatly strengthen the evidence for typicality as a determinant of word order over that provided by prior studies. (Of course, the prediction of the lexical account *is* the main-effect advantage of typical items, and so it will always suffer from the correlational weakness.)

Another potential concern with Kelly et al. (1986) is that they used production norms (Battig & Montague, 1969) to determine the typicality of their items. The frequency of production of a word depends on many factors besides the typicality of the entity to which the word refers (e.g., frequency, familiarity), though there is a correlation between production and rating norms (Mervis, Catlin, & Rosch, 1976). When Kelly et al. (1986) found that typicality as measured by production frequency affected word order, it is not clear that the effect was caused only by category typicality. In the current study we used typicality ratings, thus avoiding some potential confounds. If we also find effects of typicality on word order, this will provide additional support for the findings of Kelly et al. (1986).

In summary, this experiment investigated whether the typicality effect on word order depends on the concepts evoked in a sentence, supporting the conceptual approach, or whether typical items tend to appear first regardless of the framework, which would support the lexical approach.

Method

Participants: Eighty participants were recruited from the University of Illinois at Urbana-Champaign Introductory Psychology participant pool or from the university community. They received class credit or cash remuneration. All reported speaking English as their first language.

Materials: We were interested in the effect of typicality on the ordering of words in sentences and whether typicality effects would occur only when the relevant category was evoked. To examine the effect of typicality, we constructed noun phrases that referred either to entities highly typical in their taxonomic category (e.g., *pants* for clothing) or to entities that were less typical (e.g., *scarf*). The phrases were put into critical conjunctions which contained two noun phrases from the same taxonomic category (e.g., *some pants and a scarf*) or two noun phrases from different taxonomic categories (e.g., *some pants and a rug*) connected by the word *and*. The critical conjunctions were embedded in sentence frames (e.g., *Nancy made some pants and a scarf out of denim*) that served as the answers to questions (e.g., *What did Nancy make out of denim?*). Questions served as cues for the cued recall of the answers. Table 1 outlines the materials and procedure.

Materials: Critical conjunctions: Thirty-two same-category conjunctions were constructed using 64 noun phrases from eight of the standard taxonomic categories listed by Rosch (1975). Table 2 shows the normative properties of the noun phrases. Half were highly typical

in their taxonomic category according to Rosch's ratings. They were conjoined with noun phrases from the same taxonomic category that were less typical of the category but had the same number of syllables and stress pattern. Table 2 also shows the mean difference in the typicality ratings within each pair. The high and low typicality items did not differ in raw or log frequency (Francis & Kucera, 1982). Highly typical items were more frequent than their less typical match for 15 out of 32 pairs, and 1 pair was tied.

Because the same noun phrases were used both in same- and different-category critical conjunctions, it was necessary to control typicality, frequency, and stress for the different-category conjunctions as well. The 32 same-category conjunctions (described above) were paired to make 16 complete items (called *quadruplets*). A quadruplet consisted of four noun phrases, two from each of two taxonomic categories (e.g., *pants* and *scarf* from clothing, *couch* and *rug* from furniture), all matched on syllable number and stress. Different-category conjunctions were made from the quadruplets by joining the highly typical noun phrase from one taxonomic category (e.g., *pants, couch*) with the less typical noun phrase from the other taxonomic category (e.g., *rug, scarf*). From each quadruplet, two different-category conjunctions were constructed (e.g., *some pants and a rug, a couch and a scarf*). The highly typical and less typical noun phrases were the same as in the same-category phrases described above, so the range and averages of the typicalities were the same. The difference in typicality ratings for the different-category pairs ranged from 1.12–4.26. The frequency difference of the less typical item minus the more typical item ranged from −62 to 29. Highly typical items were more frequent than their less typical match for 15 out of 32 different-category pairings, and 2 pairs were tied.

Noun phrases were presented in conjunctions in two different orders: The highly typical item could be first and the less typical item second, called *typical first* (e.g., *some pants and a scarf*) or the less typical item could be first, called *typical last* (e.g., *a scarf and some pants*). The order of presentation of the noun phrases (typical first or last) was crossed with the type of category factor (same or different category) to make four types of conjunctions.

Materials: Sentence frames: Each quadruplet required two answer frames with associated questions, one for each of the two same-category conjunctions. Sentence frames were written so as not to strongly evoke any particular category, since they had to serve as the frame for both same- and different-category phrases. For example, the frame for *Nancy made some pants and a scarf out of denim* accommodated both the same-category (*pants* and *scarf*) and the different-category (*pants* and *rug*) members of the quadruplet, as did its matched frame *A couch and a rug were needed before the play could open*. The omission of the superordinate category name was different from Kelly et al.'s (1986) method, in which the relevant superordinate category was always included in the sentence.

Each of the two frames associated with a quadruplet was used for one same- and one different-category conjunction. To make the different-category answer, the noun phrase referring to the highly typical entity remained in its frame, and the noun phrase referring to the less typical entity was placed in the other frame (e.g., *A scarf and a couch were needed before the play could open*), for different participants.

Questions that served as cues for each of the 32 sentence frames contained all the information from the target answer except the conjunction (e.g., *What was needed before the play could open?* for the answer *A couch and a rug were needed before the play could open.*).

Materials: Audiotapes: Questions and answers were presented to participants on audiotape. To make the tapes, answer frames (containing the different-category conjunctions in both typical-first and typical-last orders) and questions were read aloud and digitized using

Macromedia's SoundEdit 16 version 2. Recordings were sampled at 44.1 kHz, 16 bits per sample. A question and answer pair was treated as four units: the question, the answer frame, the first noun phrase, and the *and* plus the second noun phrase, which were considered a single unit. A single recording was chosen for each of the four units, then these units were pasted together to create the tokens that the participants heard. This resulted in each answer frame and each question being used four times (same- or different-category crossed with typical-first or typical-last order). Each selected noun-phrase recording was used twice, once in a same-category conjunction (either typical first or typical last) and once in the different-category phrase (again either typical first or typical last). The selected noun-phrase recordings were not used four times (as the answer frames and questions) because of the difficulty of separating the *and* from the following noun phrase. The splicing resulted in fluent, natural-sounding sentences, while ensuring that there were no prosodic differences between the same- and different-category conditions. Fillers were also spliced. The reassembled question and answer tokens were recorded onto audiotape.

Each participant heard eight instances of each of the four types of items. The 32 instances were presented in four blocks, each block containing two instances of each item type and seven fillers (in fixed positions in each block). The 28 fillers did not contain noun phrases from any of the taxonomic categories from which the target phrases were constructed. They did not contain two noun phrases connected by *and*. Their associated questions did not necessarily probe for noun phrases. Half of the fillers contained more than one noun phrase from a single taxonomic category. Each participant heard only one version of each of the answer frames and its associated question. Each noun phrase and each category instance was heard once by each participant.

Four lists were constructed. All the lists had the question-answer pairs in a single, fixed random order, but differed as to the type of conjunction in the answer frame. Participants were randomly assigned to one of the four lists.

Procedure: Participants were tested individually or in groups of up to 10. The experiment used cued recall, in which questions served as the cue for the recall of the target sentence containing a conjunction (Table 1) as a proxy for spontaneous sentence production (see Bock, 1982). After reading the instructions, participants listened to a list of 15 question-answer pairs (the study phase; see Table 1), played on a cassette tape recorder. There was 1.5 s after the question before its associated answer and 3.5 s after an answer before the next question. Next, participants heard each question (in the same order as in the study section) and had 18 s to write the answer "as completely and accurately" as they could (test phase). Two types of instruction were used. Half the participants were told to try "to remember the answers," and half were told to try to reproduce these answers verbatim. As there were no reliable main effects nor interactions involving instruction type, this variable will be ignored. The experiment took about 50 minutes to complete.

Results

Scoring: Written responses were scored if the participant included both the target noun phrases connected by the word *and*. Near synonyms were accepted if it was reasonable to assume typicality judgments would not be affected (e.g., *taxicab* for *taxi*), and changes of number and definiteness were accepted. The number of scorable responses (containing the two target noun phrases and an *and*) indicates how well people remembered the answers' contents in different conditions. Of the scorable responses, a *switch* was counted when the noun phrases were recalled in the reverse order from which they had been presented. *Switch proportion* was defined as the number of switches divided by the number of scorable responses. Within the same- and different-category conditions, switch proportions were calculated separately for

typical-first and typical-last conditions. High switch proportions in the typical-last conditions (the pattern found by Kelly et al., 1986) indicate a tendency to put highly typical items earlier. High switch proportions in typical-first conditions indicate a tendency to put less typical items earlier in the conjunction. We calculated the difference between the switch proportions in the typical-last and typical-first conditions. Barring an inherent bias for the ordering of the words unrelated to typicality, we predicted more switching in the typical-last condition (to put highly typical items earlier), so the *typical-leader effect* was calculated as the switch proportion for typical-last minus the switch proportion for typical-first items. A positive typical-leader effect indicates a tendency to put highly typical entities early in the sentence. A negative typical-leader effect indicates a tendency to put less typical entities early. An effect close to 0 indicates no preference for the ordering of the noun phrases based on typicality. Table 3 shows the switch proportions in each condition and the typical leader effects for this and the following experiments.

Scorable item analysis: Analyses were conducted using both participants (F_1) and items (F_2) as the random factor (Clark, 1973) to examine generalizability across each. Because these errors are sometimes sparsely distributed, item effects were not always significant when the participant effects were. We interpret these differences by participant as meaningful, although they clearly do not provide as strong evidence as when both are reliable.

There was a main effect of category type in which same-category pairs were slightly more difficult to remember ($M = 3.9$ pairs recalled out of 8) than different-category pairs ($M = 4.2$; $F_1(1, 79) = 5.95$, $p < .02$, $F_2(1, 15) = 2.24$, $p < .16$), but no effect of noun phrase order (typical first: $M = 4.0$; typical last: $M = 4.1$; $Fs < 1$) and no interaction ($Fs < 1$) in a category type (2) x noun phrase order (2) within-subject analysis of variance (ANOVA). So, different-category pairs were slightly easier to remember, but crucially it was no more difficult to remember pairs presented in typical-first or typical-last order, and there was no interaction of order with the category type. As the dependent measure is based on the difference in switch rates for the typical-first and typical-last items, it is particularly important that the typical-first and typical-last conditions did not differ greatly in how easy they were to remember.

Typical-leader effect: Figure 2 shows the switch proportions for the typical-first and typical-last conjunctions. There was a difference in the typical-leader effect in same- and different-category conditions: There was a greater tendency to put the typical item first in the same-category condition (a difference of .129 in the switch proportions between the typical-first and typical-last phrases) than in the different-category items (a difference of .002 in the switch proportions) ($F_1(1, 79) = 6.36$, $p < .02$; $F_2(1, 15) = 33.87$, $p < .0001$). These differences have a 95% confidence interval of .071 for planned pairwise comparisons, as shown in Figure 2. For the same-category condition, the typical-leader effect of .129 was reliably different from 0 ($t_1(79) = 3.45$, $p < .001$, $t_2(15) = 6.53$, $p < .001$), but the different-category condition difference of .002 was not ($ts < 1$). That is, when both noun phrases were from the same taxonomic category, errors tended to result in putting the highly typical item earlier in the conjunction. On the other hand, when the two noun phrases were not from the same taxonomic category, errors were just as likely to put highly typical things earlier as later.

Discussion—Experiment 1 examined whether the effect of typicality on word order in sentences depends on the relevant conceptual framework being evoked. Previous work in sentence production suggested that typicality had its effect when words were assigned to positions within the same syntactic constituent. If conceptual information plays no role, references to typical things should occur earlier in a phrase regardless of whether the category was evoked, since the lexical representation of typical items should be more accessible. In contrast, if conceptual information is involved in typicality effects, entities should only be

mentioned earlier in the phrase when the superordinate category was activated, since conceptual typicality is a function of the similarity of a concept to its superordinate.

The results show clearly that typicality influences word order only when the conceptual information was in fact salient. A positive typical-first effect was found only in the same-category condition (i.e., *pants and scarf* was produced more often than *scarf and pants*, but *pants and rug* and *rug and pants* were equally likely). Only when the superordinate category was activated did people tend to put words referring to highly typical entities earlier within phrases, supporting the claim that typicality effects depend on conceptual processing.

Experiment 2

Experiment 1's results suggest that typicality effects on word order stem from conceptual information. One potential contribution to this effect is that something about the *words* referring to concepts may be associated to one another and to the superordinate concept. As we mentioned above, thinking of fruit might activate the word *apple* (as in Figure 1), because apples are prototypical fruit. Thus, it is possible that the effects of conceptual relations found in Experiment 1 work through lexical associations of some kind. Depending on one's theory of the lexicon, these links could be at the level of word meaning (the lexical concept, the lowest level of conceptual structure in Figure 1), at the level of the lexical entry or lemma, or at the level of word form. For present purposes we are agnostic about the precise nature of the lexical representation; what is important is that the relevant representation is lexically specific.

An alternative explanation is that the results are purely conceptual, reflecting semantic congruity. For example, imagine that a person's memory representation of a sentence includes the fact that two fruits were mentioned, as well as specific information about which fruits they were (see General Discussion for more about this assumption). During the sentence construction process, a typical item would be more congruent with the fruit memory than would an atypical item—the properties of apples would overlap the properties of fruit more than the properties of lemons would. Therefore, the superordinate category would be a better retrieval cue for the typical item, or speakers might feel more confident that the typical fruit was in fact mentioned in the sentence, leading them to produce it first. One could construct a variety of explanations of this sort, but their critical property is that it is not pre-existing associations between the superordinate and category member and their lexical representations that are determining word order, but rather their semantic congruence or similarity.

One way to distinguish these possibilities is to use categories that are unlikely to have lexical-level links, because they are not well established in memory, but that would still have this semantic congruity. Ad hoc categories, which are organized around a goal instead of a prototype (Barsalou, 1983, 1985), fit this description. Like standard taxonomic categories, they show typicality gradients. That is, people generally agree that some things are good members and some things less good members of these categories. For example, in the ad hoc category of things to take out of a burning house, it is generally agreed that photos are a good member, whereas blankets are less good (Barsalou, 1985).

For our purposes, what is useful about ad hoc categories is that they are not as well established in memory as standard taxonomic categories are (Barsalou, 1983). This makes it less likely that words referring to members of an ad hoc category have direct lexical links to the category label or to each other. Typicality effects with ad hoc categories would be difficult to explain as a lexical level effect, even indirectly, for at least two reasons. First, Barsalou's ad hoc categories should not be strongly associated with their superordinate categories because some of them were activities labeled by verbs. It has been argued that activities/verbs are less likely to be organized into taxonomies in semantic memory than are objects/nouns (Miller & Fellbaum, 1991; Morris & Murphy, 1990). Thus, they would lack the long-term memory

associations of object categories illustrated in Figure 1. (However, since activities like eating are very typical of picnics, for example, any effect due to semantic congruity of the items should still have an effect.) Second, the word *photos* surely does not have prominence in the lexicon because photos are typical things to carry out of a burning house. But if conceptual accessibility or congruity drives the typicality effect, then ad hoc categories should act in a manner similar to standard taxonomic categories. That is, typicality effects should be seen when the ad hoc category is evoked, and not when it is not, because activation of a relevant category should aid retrieval or verification of highly typical entities. Thus, Experiment 2 had the goal of further investigating whether typicality affects word order only when the relevant category is evoked, further testing the lexical account.

Experiments 1 and 2 also differed in that in Experiment 1 the condition in which the category was evoked (same-category condition) and the condition in which the category was not evoked (different-category condition) had different pairings of category members. In Experiment 2, however, the pairs of words referring to category members were the same in both the category evoking (*category*) and comparison (*no-category*) conditions. Instead of varying the words in the conjunction, context sentences were varied. This was possible because members of ad hoc categories do not automatically evoke that category (Barsalou, 1982). This design provided an even more stringent control of lexical information, since there was no difference whatsoever between the target sentences in the category and no-category conditions.

Method

Participants: Ninety-one participants were recruited for Experiment 2 from the same population as Experiment 1. Sixty-two additional participants normed items and 96 served in three post-tests. None of them had taken part in Experiment 1.

Materials: Critical phrases: A full item consisted of two phrases from a single ad hoc category (e.g., *photos* and *blankets*, from things to take from a burning house) embedded in an answer frame with an associated question (described below). To construct the 36 pairs of phrases, 30 participants provided typicality ratings for entities in nine ad hoc categories. All the phrases in Barsalou's (1985) appendix (with a few lexical changes) plus a few additional phrases were rated. Participants were given a booklet with nine pages. At the top of each page the ad hoc category was mentioned. Participants were to circle a number from 1–7 for each item indicating that the entity was a good (7) or a poor (1) member of the category. Items within an ad hoc category were presented in a single random order or its reverse, with half the participants getting each order. The nine categories were presented in different random orders in each of the 30 booklets.

For each of the nine ad hoc categories, four pairs of phrases were selected such that the pairs were matched for number of syllables and stress patterns, for a total of 36 pairs. In each pair, one entity was rated as highly typical of its ad hoc category, and the other was rated as less typical. Table 2 gives the items' typicalities and word frequencies.

Three of the nine ad hoc categories had members that also fell into the same taxonomic category. For two of these, ratings (from Rosch, 1975) were available. For the category of things to wear in the snow, items were chosen such that the taxonomic typicalities did not follow the same pattern as the ad hoc typicalities, as much as possible. For example, for the items *scarf* and *vest*, the more typical item in the ad hoc category (*scarf*) was less typical in the standard taxonomic category. For the ad hoc category of long-distance transportation (taxonomic category of vehicles), it was not possible to unconfound the typicality ratings in the taxonomic and the ad hoc categories. However, any associated effect should work against the context manipulation in the present study. If typicality in taxonomic categories influences

word order, this influence would be present both when the ad hoc category is evoked and when it is not.

As in Experiment 1, category members occurred in both orders: either typical first (*photos and blankets*) or typical last (*blankets and photos*).

Materials: Sentence frames: Each category member pair had a single answer frame and an associated question. Answer frames were written so as not to evoke any particular category, since they served as frames for both category and no-category conditions. For example, an answer frame might be *Franklin ran in and grabbed…*, for the items photos and blankets. As before, cue questions contained all the information from the target answer except the conjunction.

In addition to the answer frame, each item was presented with a context sentence that either evoked a particular ad hoc category or did not evoke that category. To be sure that the context sentences evoked their intended ad hoc categories (or not), 32 additional participants rated how much the context sentences made them think of particular categories. They read context sentences followed by the answer sentence frame with the phrase *X and Y* substituted for the conjunction. They were to circle a number from 1–7 indicating how much the pair of sentences made them think of the indicated ad hoc category (where 1 = very likely to evoke the category). Based on these ratings, four context sentence pairs (ad hoc category evoking, non ad hoc evoking) were chosen for each of the nine ad hoc categories, yielding 36 context pairs. The contexts that should evoke the ad hoc category (category condition) were rated as more likely to evoke that category (range 1.1–3.3 out of 7, $M = 1.8$) than the contexts that should not evoke that category (no-category condition; range 2.7–6.8, $M = 5.2$). Across all 36 pairs of contexts, the difference in ratings for the category and no-category context sentence pairs ranged from 1.2–5.3 with an average of 3.4.

Materials: Audiotapes: Questions and answers were again presented to participants on audiotape. To make the tapes, questions, context sentences, and answers were read aloud twice, once with each ordering of the conjunction, into SoundEdit 16, as in Experiment 1. A question and answer pair was treated as four units: the question, the context sentence, the answer frame, and the entire conjunction. A single recording token was chosen for each of the four units, and then these units were pasted together to create the materials, as in Experiment 1. Because sentences were spoken somewhat slowly to facilitate the cutting and pasting, after final versions of question-answer pairings were assembled, all items were speeded up by 10%. These questions and answers were recorded onto audiotape.

Each participant heard nine instances of each of the four types of items (category vs. no-category X typical first vs. typical last). The 36 items were presented in four blocks, each block containing two or three items of each type and five fillers (in fixed positions in each block) that were similar to those used in Experiment 1. Phrases referring to category members were heard once by each participant.

Four lists were constructed. All the lists had the question-answer pairs in a single fixed random order but differed as to the type of conjunction and context sentence in the question-answer frame. Participants were randomly assigned to one of the four lists.

Procedure: The procedure was the same as Experiment 1 with the following exceptions. Participants were tested individually or in groups of up to 12. In the study phases, after a question there was 1 s before its answer was presented and 2.5 s between an answer and the next question. In the test phases, participants had 16 s to write each response. All participants received the instructions to "write the answer as completely and accurately" as possible.

Results

Scoring: Responses were scored if they contained the word *and* (*or* and comma were accepted as substitutes) and both category members were mentioned in the same phrase. Rough synonyms were accepted if they did not seem to affect the typicality in the ad hoc category, and changes in number and definiteness were accepted. Switch proportion and typical-leader effect were defined as before. Table 3 gives the switch proportions along with the differences between them (i.e., the typical-leader effects).

One of the items was discovered not to be matched in syllables (*mittens*, *shorts*) and was subsequently removed from all analyses. As a result, analyses by participant involved different numbers of items in the four conditions, and therefore scorable items analyses are presented for proportions of items completely recalled rather than raw number recalled (as in Experiments 1 and 3). Also, since we were interested in the creation of sentences, nine participants who did not write complete sentences on at least 35 of the 36 target trials were not included in the analyses. Two other participants were randomly excluded to ensure counterbalancing, leaving 80 participants.

Scorable item analysis: People were more likely to recall both category members in a phrase with *and* when the ad hoc category was evoked than when it was not (Ms =.70 and .59; $F_1(1, 79) = 43.58$, $p < .001$; $F_2(1, 34) = 20.48$, $p < .001$), but more important, it was no more difficult to remember pairs presented in typical-first or typical-last order (Ms = .64 and .65; Fs < 1), and there was no interaction of order and category availability (Fs < 1).

Typical-leader effect: There was a marginal difference in the typical-leader effect for category and no-category items: There was a tendency to put the *less* typical item first in the no-category condition ($M = -.06$), but no such effect in the category condition ($M = .01$; $F_1(1, 79) = 3.57$, $p < .07$; $F_2(1, 34) = 1.45$, $p < .24$); these effects have a 95% confidence interval of .05 for planned pairwise comparisons, calculated from the subjects analysis. This tendency was reliably greater than chance in the no-category condition ($t_1(79) = -2.28$, $p < .03$; $t_2(34) = -1.14$, $p < .27$), opposite to the expected direction. In the category condition, the typical-leader effect was not reliably different from 0 (ts < 1). So, when the ad hoc category was not evoked, errors tended to result in putting a less typical item earlier in the conjunction. On the other hand, when the context evoked a particular ad hoc category, this tendency to put the less typical item first was overridden, yielding no preference for either phrase (highly or less typical) to go earlier in the conjunction. To investigate the sources of the unexpected no-category effect, we carried out three post-tests.

Post-test 1: To find out if there were uncontrolled item differences that made the typical-last order generally preferred, nine raters participated in a forced-choice preference task in which the ad hoc category label was not provided. Raters were asked to "circle the phrase with the wording you prefer" and were given the choice between two phrases with conjunctions, one with the typical entity first and the other with the less typical entity first (e.g., *photos and blankets* or *blankets and photos*). The order of presentation of the phrase pairs was counterbalanced across raters. Across the 36 pairs, raters chose the typical-first phrase 49% of the time, suggesting that in the absence of a sentence context there was little inherent preference for less typical items to go earlier in the phrases. These results provide a contrast with taxonomic categories, which do show a preference for phrases with the typical item first (Kelly et al., 1986.)

Post-test 2: An additional 21 participants performed a cued-recall test identical to the ad hoc category experiment except that context sentences were not presented. The data from one participant who fell asleep during the experiment were not included in the analysis. Participants

heard the question and the target answer that contained the conjunction but did not hear the context sentences. The conjunctions were presented in both typical-first and typical-last orders. Because the context sentence alone was what differentiated the category and no-category conditions, eliminating this sentence left half as many conditions, requiring only two lists. Experimental procedures were the same as in the main experiment.

Without the context sentences, there was no tendency to put the entity that was highly typical of the ad hoc category later (i.e., there was no negative typicality effect like the one found in the no-category condition of the main experiment; $M = .007$; ns). There was also no tendency for the typical-first to have more scorable items than the typical-last condition (Ms = .53 and .52; Fs < 1). That is, there was no evidence for an inherent lexical bias in our items that could explain the unexpected typical-last effect in the no-category condition of the main experiment. Therefore, the nearly reliable difference between the category and no-category conditions in Experiment 2 cannot be explained in terms of the category-condition contexts allowing an inherent bias in the stimuli to be overridden.

Post-test 3: Although our focus here is sentence production, our hypotheses depend on memory and conceptual processes that are not unique to language production. In particular, we have argued that the typicality effect of Experiment 1 might be due to pre-existing associations between superordinate categories and their members, which could cause an advantage in memory retrieval. Given that we found no evidence for typical ad hoc category members being produced first, it is appropriate to ask whether the underlying memory advantage for typical category members is found in a simpler recall situation. If not, this would confirm our earlier result that there is no inherent, lexically based typicality advantage in sentence production. Thus, we performed a follow-up study in which category labels (instead of full sentence questions) served as cues for the recall of the conjunctions. We included taxonomic categories as a comparison group and predicted that there would be a typical-first preference for taxonomic but not ad-hoc categories.

Sixty-six participants heard a category label followed by a conjunction. We used 24 of the taxonomic conjunctions from Experiment 1 and all of the ad hoc conjunctions from Experiment 2. (Some taxonomic conjunctions were omitted because each block could have at most one of each category label, and there were more than four conjunctions for some taxonomic categories.) Participants in this post-test might hear *things to carry out of a burning house* followed by *photos and blankets*. The conjunctions were presented in both typical-first and typical-last orders, across participants. There were again only two lists (since the category was always relevant), and the procedures were the same as in Post-test 2.

When we used category labels as cues for recall of the conjunction, there was no tendency to put the entity that was highly typical of the ad hoc category later (or earlier; $M = -.0004$; Fs < 1). In contrast, the taxonomic category items from Experiment 1 had a reliable typical-leader effect ($M = .07$; $F_1(1, 65) = 6.19$, $p < .02$; $F_2(1, 23) = 3.44$, $p < .08$) which was reliably different from the typical-leader effect for the ad hoc category items ($F_1(1, 65) = 4.53$, $p < .04$; $F_2(1, 58) = 4.33$, $p < .05$). In short, ad hoc categories do not seem to lead to earlier recall of highly typical entities, though taxonomic categories do.

Combined analysis—We hope to draw inferences about the mechanism underlying typicality's effect on word order based on the lack of any order effect with ad hoc categories. However, it is difficult to interpret null results, and readers may wonder whether the results of Experiment 2 are really different from those of Experiment 1. In order to reach a more positive conclusion about the difference between the two experiments, we analyzed together the switch proportions of their experimental conditions (the within-category condition of Experiment 1 and the category condition of Experiment 2). (We are claiming that speakers switch words

when the category is evoked in Experiment 1 but not in Experiment 2; we are not claiming any difference between the control conditions of the two experiments when no category is evoked, and so testing these conditions would not address our hypothesis.) This analysis yielded a reliable interaction of word order (typical-first or typical-last) and experiment, $F_1(1, 158) = 7.91$, $p < .005$; $F_2(1, 50) = 11.80$, $p < .002$. That is, there was .129 more switching in the atypical-typical order than the typical-atypical order in Experiment 1, but only .009 in the corresponding orders in Experiment 2, and this difference was reliable. In combination with the similar interaction in Post-Test 3, these results show that ad hoc categories have a consistently different pattern of results than do taxonomic categories.

Discussion—Experiment 2 examined whether the typicality effects found in Experiment 1 with taxonomic categories would replicate with ad hoc categories, and whether the typical-leader effect depends only on the activation of a relevant conceptual framework. Typicality effects with ad hoc categories are unlikely to be due to purely lexical associations since words for category members are not generally linked in memory. That is, although photos and blankets are both things to take from a burning house, Barsalou (1983) showed that there is no direct link between the words *photo* and *blanket* and *things to take from a burning house*. The results suggested that typicality within ad hoc categories did not affect the ordering of words, either in sentence production or in simpler memory tests. People were no more likely to produce *photos and blankets* than *blankets and photos*, when the category of things to take from a burning house was relevant (or when it wasn't). (We are inclined to view the opposite effect found in the main experiment as a fluke, given that it was not replicated in three post-tests. Possibly there is something about sentence production that causes a reverse typicality ordering for ad hoc categories, but it is difficult to imagine what that might be.) In the face of consistent null effects of ad hoc category evocation on the tendency to reorder words within phrases, it is important to note that the experimental manipulation did have a highly reliable effect on the memory for the entities: Memory for the critical phrases was better in the ad hoc category context than outside of that context. This is what would be expected if the category information provided a useful retrieval cue for the to-be-remembered entities, and it argues that the manipulation of context was sufficiently strong.

In line with the assumption that ad hoc category members are not represented together in memory, Experiment 2 found a different result than Experiment 1 for accurate recall of category members in sentences. In Experiment 1, fewer within-category than cross-category pairs were remembered correctly, consistent with extensive evidence for within-category semantic interference from many areas (including part-list cuing effects in memory and semantic interference effects in picture and word naming; e.g. Roediger, 1973; Slamecka, 1968; Wheeldon & Monsell, 1994). The contrasting pattern in Experiment 2, with better recall when the category was evoked, implies different memory structures for ad hoc than for taxonomic categories. Indeed, Crowder (1976, p. 348) suggests interitem interference as a test of whether items are linked in memory.

The results of Experiment 2 differ markedly from the results of Experiment 1. Experiment 1 demonstrated that the ordering of words in sentences depended on the available context, suggesting that ordering cannot be based on lexical factors alone. Experiment 2, eliminating lexical factors, found no such effect. This suggests that conceptually-determined typicality is not sufficient to cause reordering of words during production. Although our typical ad hoc category members were rated much more typical than the less typical ones, this did not change their production order, either in sentences or in a simple cued-recall task (Post-test 3).

One explanation of the absence of a typical-leader effect for ad hoc categories is that in familiar categories, there is a lexical link between superordinates and their members that influences word order within phrases. Ad hoc categories, lacking that lexical link, therefore do not show

a preference for typical members appearing first. How does this explanation fit with the other key finding of Kelly et al. (1986), that typicality did not influence choice of a constituent as subject or object? Given that syntactic, sentence-level ordering seems not to be strongly influenced by lexical form (Bock, 1986), the lexical account is consistent with their null result. However, null results may always be due to a failure to find the correct experimental conditions to reveal the effect. Therefore, in Experiment 3, we used better controlled stimuli and a different sentence construction to provide a more powerful test of their conclusion.

Experiment 3

Experiment 1 replicated and extended the finding of Kelly et al. (1986) that typicality influences the order in which words are produced within a phrase. All the conjunctions used two category members connected by *and*, so reordering the category members did not change the major sentence constituents (subject, direct object). Experiment 3 addressed whether typicality can influence the order of the same kinds of noun phrases serving as major sentence constituents. Using the active-passive alternation, Kelly et al. (1986) found no tendency to put highly typical category members earlier in sentences when switching meant changing from a passive to an active sentence. However, passives and actives differ on features besides word order that may have reduced the switching from one to another. For example, if participants believed that a stimulus sentence was primarily about scarves, they may have resisted making *hats* the subject, because that would have changed the sentence's perspective. Also, any memory participants had for the voice of the sentence would have inhibited switching.

We used pairs of sentences that shared as much lexical and perspective information as possible. They referred to two entities in the same taxonomic category, one that was highly typical of the category and one that was less typical. One sentence positioned the two entities in the same phrase (e.g., *A turkey and a robin took a walk around the lake*) with the same grammatical role (in this case, both are part of the subject). An effect of typicality, the same one observed in Experiment 1, would be demonstrated through a reordering of words within the phrase. The paired sentence used the two category members as the subject and prepositional object (e.g., *A turkey took a walk around the lake with a robin*), giving the entities different grammatical roles. For these sentences an effect of typicality would be manifest in a switching of the critical noun phrases while maintaining the overall sentence structure. The use of symmetrical verbs reduces the pragmatic changes associated with reordering of major constituents (Gleitman, Gleitman, Miller, & Ostrin, 1996) and does not require a change in the sentence's voice when constituents are switched. Most importantly, the same category names and verbs were used in the two conditions, in contrast to Kelly et al., who used different categories and unrelated sentences in the phrasal and sentence-ordering conditions.

Since both category members were always from the same taxonomic category (as in Experiment 1), when entities had the same grammatical role and were mentioned in the same phrase, we expected a tendency for highly typical things to be placed earlier. When the entities had different roles and were in different major constituents, there were two possibilities. First, the conceptual forces apparent in Experiment 1 could also cause highly typical things to exchange position (and, thereby, grammatical functions) with less typical things. That is, typicality could affect the order of words both within (as in Experiment 1) and across phrases, which would be evidence that typicality has effects on multiple mechanisms of sentence production, perhaps reflecting both conceptual and lexical influences. Alternatively, there might be no tendency for typicality to effect changes in grammatical roles, as Kelly et al. (1986) found. This would be consistent with typicality, like most lexical factors, affecting lexical selection rather than the assignment of grammatical functions.

Method

Participants: Sixty-six participants were recruited from the same population as the previous experiments. Thirty-two others participated in a study to norm items (three of these also participated in posttests of Experiment 2, which had no item overlap with this experiment). An additional 53 participated in a replication of the main result of Experiment 3.

Materials: As in Experiment 1, pairs of highly typical and less typical noun phrases from the same taxonomic category were used. These noun phrases appeared in two different sentence frames. One frame contained the two noun phrases in the same phrase connected by the word *and* (same-phrase condition). The other sentence placed the same two noun phrases in different major constituents (different-phrase condition). As in Experiments 1 and 2, the sentences served as the answers to questions which were used as cues in a cued-recall task, and noun phrases occurred in both orders (typical first or typical last).

Materials: Target noun phrases: Thirty-two critical sentence pairs were selected as described below. Each sentence contained two target noun phrases, one highly typical and one less typical of a standard taxonomic category. The 64 selected noun phrases came from eight of the standard taxonomic categories normed by Rosch (1975). Half of the noun phrases were highly typical in their taxonomic category (see Table 2 for typicality ratings from Rosch, 1975). These were paired with noun phrases from the same taxonomic category with the same number of syllables and stress pattern that were less typical of the category. The high and low typicality items did not differ in mean raw or log frequency (see Table 2). Highly typical items were more frequent than their less typical match for 11 out of 32 pairs, with 3 pairs tied.

Materials: Critical sentence frames: Each noun phrase pair was put into two answer frames sharing a single associated question. Answer frames were written so as not to evoke any particular categories. One of the sentence frames contained the word *and* connecting the two noun phrases in a single phrase (e.g., *A turkey and a robin took a walk around the lake,* the same-phrase condition). Its matching sentence frame was as similar in meaning as possible while placing the two noun phrases in different major constituents (e.g., *A turkey took a walk around the lake with a robin,* the different-phrase condition). The same-phrase condition was a replication and extension (with additional noun phrase pairs and new sentences) of the same-category condition of Experiment 1. As before, cue questions contained all the information from the target answer except the critical phrase.

These 32 pairs of answer frames with embedded noun phrases were selected from 80 potential answer frame pairs based on ratings from 32 participants. The raters read pairs of sentences and judged similarity of the meanings of the pairs, from 7 (meanings are the same) to 1 (meanings are very different). Participants were also permitted to select "nonsense" instead of a rating, in case a word was unknown to them. Each participant rated 40 items with noun phrases in the typical-first order and 40 items in the typical-last order, counterbalanced across lists. The range of meaning similarity ratings in the 32 selected sentences was 4.1–6.3 out of 7, with an average of 5.2. Thus, the two versions of the sentences had similar meanings.

Materials: Audiotapes: Questions and answers were again presented to participants on audiotape. To make the tapes, questions and the four versions of the answer were read into the Goldwave 4.02 recording and editing program with the same recording parameters as used in Experiments 1 and 2. Winamp 2.08 was used to construct play lists and to play the sound files for recording onto audiotape. A question and answer pair was treated as five units: the question, the two (same-phrase, different-phrase) answer frames, and the two target noun phrases. A single recorded token was chosen for each unit, and the units were then pasted together to create the tokens the participants would hear. Since there were two answer frames (same-phrase,

different-phrase), parts of the answer frames were used only twice (e.g., the word *and* in the same-phrase condition frame), and parts were used four times.

Each participant heard eight instances of each of the four types of items. The 32 items were presented in four blocks, each block containing two items of each type and seven fillers. The 28 fillers from Experiment 1 were used. Each participant heard one version of each of the answer frame pairs and its associated question, and each noun phrase pairing was heard just once by each participant.

Four list versions were constructed. All the lists had the question-answer pairs in the same fixed random order but differed as to the noun phrase order and the phrase type that went with each question. Participants were randomly assigned to one of the four lists.

Procedure: The procedure was the same as in Experiment 2.

Results

Scoring: One participant was eliminated for not writing complete sentences, two others were excluded for being nonnative speakers of English, and three more were randomly excluded to ensure counterbalancing, leaving 60 participants. Responses were included for scoring using two criteria, *presentation scoring* or *production scoring*. Presentation scoring was similar to scoring in Experiments 1 and 2. Here, items were scored only if they were produced using the same type of sentence frame in which they had been presented. That is, if the original sentence presented the noun phrases in the same phrase, responses were only scored if the participant produced both noun phrases in the same phrase with the word *and* (or an acceptable substitute). For sentences in the different-phrase condition, responses were scored only if the target noun phrases were in different phrases. They did not have to have the exact syntax of the target sentence, but this was the type of different-phrase sentence most commonly produced. So, a response was scored when the type of sentence produced was the same as the type presented and both target noun phrases (or acceptable substitutes) were included. Within the scorable responses, a switch was counted when the noun phrases were reversed from the order presented. Switch proportion, as before, was the number of switches divided by the number of scorable items, and the typical-leader effect was the switch proportion for the typical-last conditions minus the switch proportion for the typical-first conditions.

The *production scoring* method was to look only at what the participant produced, ignoring how the sentence had been presented. This was done because there was a very low number of scorable items under the presentation scoring criterion (on average, fewer than 2.4 scorable responses per participant per condition). So, if the participant produced the two nouns within a single constituent, this was included in the same-phrase condition, and if the nouns occurred in different constituents, it was included in the different-phrase condition—regardless of the original sentence's form. It turned out that people tended to produce sentences with the noun phrases in the same phrase regardless of how the sentence had been presented. (A similar preference for coordinated descriptions has been noted in spontaneous descriptions of moving shapes; Levelt & Maassen, 1981.) In one sense, this was positive. The confusion between the two types of sentences indicated that the intended manipulation of making the sentences in the two conditions as similar as possible worked, since people seemed not to distinguish them.

Since we were primarily interested in what happens when people produce (not recall) sentences, examining the data ignoring the condition of presentation was informative. However, because we are now allowing participants to decide how to produce each sentence rather than determining the form ourselves, there is some loss of control in this procedure. But unlike a pure observational study, we have controlled the sentence's content, the typicality and frequency of its words, its verb, and so on. So, it is unlikely that such extraneous factors can

explain our results. Furthermore, the results of the two analyses were extremely similar—the difference is that with more observations, the production scoring analyses have more power.

Table 3 shows the proportions of responses in each condition for both types of scoring and the corresponding typical leader effects.

Presentation scoring: Scorable item analysis: The number of scorable items for the same-phrase condition ($M = 3.74$) was higher than for the different-phrase condition ($M = 1.03$; $F_1(1, 59) = 178.72$, $p < .001$; $F_2(1, 31) = 115.83$, $p < .001$), with no effect of noun phrase order (typical first: $M = 2.43$; typical last: $M = 2.34$; $Fs < 1$). There were more scorable items in the same-phrase condition when the noun phrases were presented in the typical-first order ($M = 3.95$) than in the typical-last order ($M = 3.53$), but in the different-phrase condition the reverse was true (typical first: $M = .90$; typical last: $M = 1.15$) yielding a statistically significant interaction ($F_1(1, 59) = 4.77$, $p < .04$; $F_2(1, 31) = 4.84$, $p < .04$). So, people were more likely to recall the two target noun phrases when they were presented in the same phrase than when presented in different phrases, but crucially it was no more difficult to remember noun phrases presented in typical-first or typical-last order. Note, however, the small number of scorable items due to the fact that people tended to produce the noun phrases in the same phrase regardless of the structure of the presented sentence. Under the presentation scoring method, sentences heard in the different-phrase condition but produced with noun phrases in the same phrase were not scored.

Presentation scoring: Typical-leader effect: Although the effect was sizeable and in the expected direction, the typical-leader effect was not significantly different for the same-phrase ($M = .06$) and different-phrase ($M = -.04$) conditions ($F_1(1, 59) = 2.13$, $p < .16$, $F_2(1, 31) = .70$).

Production scoring: Scorable item analysis: With production coding, the number of scorable same-phrase responses ($M = 6.8$) was higher than that for different-phrase responses ($M = 1.7$; $F_1(1, 59) = 220.54$, $p < .001$; $F_2(1, 31) = 107.80$, $p < .001$), but there was no overall difference in number of scorable responses for typical-first ($M = 4.2$) and typical-last ($M = 4.2$) items ($Fs < 1$) and no reliable interaction ($Fs < 1$).

Production scoring: Typical-leader effect: Production scoring revealed a greater typical-leader effect for same-phrase responses (.08) than for different-phrase responses (−.07). The difference was significant by participants ($F_1(1, 59) = 6.78$, $p < .02$) but not by items ($F_2(1, 31) = 1.82$, $p < .19$); each effect had a 95% confidence interval of .079 for planned pairwise comparisons. The typical-leader effect for the same-phrase responses was different from 0 in both participant and item analyses ($t_1(59) = 2.86$, $p < .006$, $t_2(31) = 2.19$, $p < .04$). The negative typical-leader effect for the different-phrase responses did not differ from 0 ($t_1(59) = -1.37$, $p < .18$; $t_2(31) = -.38$, $p < .71$). This is consistent with the absence of an ordering effect when grammatical role changes are involved.

Replication study—A potential objection to this interpretation of the absent typical-leader effect in the different-phrase items in Experiment 3 is the small number of observations. Participants preferred to produce sentences with same- rather than different-phrase uses of the target phrase. In an effort to increase the likelihood of different-phrase responses, we conducted a replication study using only the different-phrase items. The rationale was that if participants did not hear any target sentences with the critical words in the same phrase, they would be less likely to produce the same-phrase responses. (We thank Scott Watter for suggesting this.) Since only the different-phrase items were used, there were only two lists. The experimental procedure and scoring were the same as in Experiment 3. After excluding the data from one nonnative speaker and from two other participants to restore counterbalancing, the data from

50 participants remained. We performed only production scoring, since it allowed comparison of the two sentence types.

Although participants heard only different-phrase items, the majority of their scorable responses were still in the same-phrase format ($M = 4.7$ in same phrases vs. $M = 3.2$ in different phrases), but the number of different-phrase constructions increased markedly compared to the main experiment. The typical-leader effect was numerically larger for the same-phrase responses ($M = .05$) than for the different-phrase responses ($M = .02$), but the difference was not reliable ($F_1(1,49) < 1$; $F_2(1,31) < 1$). The typical-leader effect for the same-phrase responses was reliably different from 0 by participants ($t_1(49) = 2.02$, $p < .05$) but not by items ($t_2(31) < 1$). The typical-leader effect for the different-phrase responses did not differ reliably from 0 ($t_1(49) < 1$; $t_2(31) = 1.62$, $p < .12$). This pattern of results is very similar to that of the main experiment.

Combined results: Experiment 3 and replication: Combining the data from Experiment 3 and its replication, and using production scoring, there was a greater typical-leader effect for same-phrase responses ($M = .06$) than for different-phrase responses ($M = -.03$) by participant ($F_1(1, 108) = 6.21$, $p < .02$) but not by item ($F_2(1, 31) < 1.0$). (The latter result is not surprising, given that combining experiments increased the number of subjects but not items.) The typical-leader effect for the same-phrase responses was different from 0 ($t_1(109) = 3.30$, $p < .002$; marginal by items, $t_2(31) = 1.83$, $p < .08$), but the negative typical-leader effect for the different-phrase responses did not approach significance ($t_1(109) = -1.0$, $p > .30$; $t_2(31) = -.45$, $p > .50$). There was no reliable effect of experiment ($Fs < 1$) nor interaction with experiment ($F_1(1, 108) = 2.33$, $p < .13$; $F_2(1, 31) = 2.76$, $p < .11$).

Detailed item analysis: Both Experiment 3 and its replication showed the expected interaction, significant by participants but not by items. As we have remarked, item analyses are weak in this design, because participants' switch errors are fairly rare, resulting in few switches on some items, leading to variability. However, we nonetheless were a bit surprised at the absence of an item effect for the critical interaction, especially given that there was a significant item effect in the same-phrase condition. A closer look at the materials suggested a possibly revealing reason for the weakness of these results. The different-phrase condition required the noun phrases to be parts of different major constituents. However, given the hierarchical nature of syntactic structure, some constituents are more closely related than others. In particular, some sentences separated the noun phrases into the subject and predicate (e.g., *A turkey took a walk around the lake with a robin*), and others separated the noun phrases into two constituents in the predicate, such as direct object and adjunct (e.g., *Andy confused the spinach with the mushrooms*). It seemed possible that switching would be less frequent in the first case than in the second. When planning the sentence subject, it is unlikely that speakers have already selected words that will appear after the verb, thereby making it difficult to switch subject and predicate nouns (as found by Kelly et al., 1986). But when a speaker is planning the structure of the predicate, lexical access of the two critical nouns within the predicate may have begun, allowing them to be switched in some cases, albeit less often than when the words are to be inserted into the same phrase.

To test this possibility, we carried out an exploratory analysis in which we examined the 21 items in which the critical phrases were separated in the subject and predicate in the different-phrase condition vs. the 11 items in which they were both in the predicate. To maximize power, we analyzed the combined Experiment 3 and replication data. For the subject-predicate items, the typical-leader effect was .125 greater in the same-phrase condition than in the different-phrase condition. When both phrases came from the predicate, the typical-leader effect was only .06 greater in the same-phrase than in the different-phrase condition. The finding that the expected effect is twice as large when the phrases are split between subject and predicate than

when both are in the predicate suggests that the weakness in the item analyses is principled rather than reflecting random variation.

We cannot make too much of this finding, because the experiment was not designed to explore this difference, and as a result, categories and sentence frames are not held constant across these conditions (and the numbers of items were too small to evaluate statistical significance). However, the results suggest that there may be a continuum of situations in which nouns can be exchanged. The smaller the planning unit the more likely lexically-based switching is to occur, because the nouns are more likely to both be active (Garrett, 1975). Complicating the picture, however, is evidence that conceptual variables like concreteness matter less to the ordering of nouns in conjunctions than to the ordering of direct and indirect objects (which can occur in immediate succession, as in *The old hermit left the university the property;* Bock & Warren, 1985). This implies that more is involved in the positioning of major sentence constituents than their proximity in sentences. Clearly, there are interesting questions here for future research to explore.

Discussion—Experiment 3 found that typicality influenced the ordering of nouns in the same phrase more than the nouns in different major constituents of a sentence. Kelly et al. (1986) obtained similar results using the active-passive alternation. Experiment 3 explored a different type of alternation to minimize the pragmatic differences between the same-phrase and different-phrase materials, using the same noun phrases in both types of sentences. Just as in Kelly et al. (1986) and as in Experiment 1, we found a typical-leader effect in sentences containing the two noun phrases in the same major constituent, but, in a new comparison, not in sentences containing the noun phrases in different major constituents. That is, errors tended to result in *A robin and a turkey walked* rather than *A turkey and a robin walked*, but there was no tendency to prefer *A robin walked with a turkey* over *A turkey walked with a robin*. So, typicality affected the ordering of words within phrases, but not across phrases, even though the meanings of the two versions of the sentences were rated as similar, and they had the same main verb, voice, and category names.

These results complement the findings from Experiments 1 and 2 by demonstrating that when the relevant well-established superordinate is evoked, typicality does not affect all aspects of word order. The findings are consistent with other evidence that typicality does not exert a discernible influence on which constituent serves as the subject of a sentence. Thus, Experiment 3's results are in accord with an account in which typicality acts more on mechanisms involved in the ordering of words than on the assignment of grammatical roles.

General Discussion

Three experiments examined the effect of category typicality on word order in sentence production. Experiments 1 and 3 found that typical things tended to be referred to earlier within phrases when the superordinate category was evoked, and not when the superordinate was not evoked. This supports the results of Kelly et al. (1986) in that when a superordinate was accessible, typicality affected word order. It extends their results to the situation in which the superordinate is not explicitly mentioned. But Kelly et al. (1986) did not examine cases in which the superordinate category was not relevant, which is a critical case for evaluating their conclusion that the typicality ordering effect arises from lexical properties.

Our Experiment 1 presented contexts that did or did not evoke the relevant superordinate, using the same sentence frames and the same lexical items (in different combinations) in both conditions. An effect of typicality was found when the relevant category was evoked, but not when it was not. If typicality were related only to lexical-level information, a word like *pants*, which is highly typical of its standard taxonomic category, should always be more accessible than a word like *rug*, which is not very typical of its standard taxonomic category.

Experiment 1 showed that this is not the case. The relative accessibility of *pants* depended on whether the category of clothing was relevant. Additionally, by measuring typicality using rating norms rather than production norms, we can be more certain that the observed order effects were due to typicality and not other factors related to production.

The effect of typicality on phrases with two items from the same category seems straightforward enough. But why didn't typicality affect phrases like *pants and rug* or *scarf and couch*? It is logically possible that people would think of these as some kind of clothing and furniture—that is, people might encode the two superordinates. If they did, then wouldn't the typical item (*pants, couch*) come to mind before the atypical item (*scarf, rug*) even in the between-category condition? Apparently people do not encode the items that way, since we did not find this effect. Perhaps when two items from the same category are mentioned, listeners take this as indicating that the category itself is relevant and therefore encode it (e.g., it is no coincidence that two items of clothing were needed for the play—perhaps the company has no costumes). Furthermore, when a phrase includes two diverse items from different categories, little is gained by categorizing those items, since the categories do not help to unify them. That is, it might be helpful to remember that some furniture was needed for the play, because furniture includes both items needed and could help the listener remember what they were. But remembering that some clothing and furniture were needed for the play does not unify the items or add anything to the memory structure beyond the items themselves.

Experiment 2 explored whether similar effects of typicality could be found with categories that are less well-established in memory. We found little evidence that typicality within ad hoc categories affected the ordering of words within phrases during sentence production. That is, there was no preference for *photos* to be placed earlier in a phrase than *blankets* regardless of the relevance of "things to take from a burning house." This negative result contrasts interestingly with the findings for taxonomic categories in Experiments 1 and 3. Though ad hoc categories share many of the features of standard taxonomic categories—most relevantly, graded membership—they did not have the same effect on word order in sentences. We suspect that the most important difference between the two kinds of categories is that taxonomic categories are familiar ones represented in long-term memory. Thus, thinking about fruit could well activate concepts of apples and oranges (but less so kiwis and lemons). In contrast, thinking about things to carry out of a burning house does not spontaneously activate typical subordinate categories. Post-Test 3 of Experiment 2 confirmed that these ad-hoc concepts did not yield a typicality ordering effect even in a simple memory experiment, whereas the taxonomic concepts did, supporting the conclusion that ad hoc and taxonomic categories have different long-term memory representations.

The result for ad hoc categories rules out the possibility that it is just semantic overlap or congruence that explains the typicality effect. For example, when participants were attempting to recall a sentence about Franklin grabbing things out of the burning house, they might begin to generate items that they would carry out of a burning house, and the typical ones might come to mind first. Or, as they begin to retrieve the semantic representation of the sentence from memory, perhaps the item more consistent with the category would be retrieved and verified faster, because it is more consistent with the category (surely people would save their family photos or pets). This did not occur, either in sentence production or the simpler memory test, suggesting that the item's pre-existing link to its superordinate is critical.

Experiment 3 reconfirmed that taxonomic categories affect the ordering of words within phrases, as in Experiment 1, but the same categories did not affect the order of major sentence constituents. There was no tendency for more typical entities to be placed earlier when the shift changed grammatical functions. So, despite the preference for producing *A robin and a turkey walked* over *A turkey and a robin walked*, there was no preference for *A robin walked with a*

turkey over *A turkey walked with a robin*, across two replications. This supports the findings of Kelly et al. (1986) regarding the absence of typicality effects on word order across major constituents and extends these findings to different sentence structures with smaller semantic and pragmatic differences than those used by Kelly et al. (1986). A priori, reducing the semantic and pragmatic differences would be expected to increase the likelihood of ordering effects. It did not.

Saying What We Mean: Implications for Sentence Production—Previous work on sentence production (e.g., Bock, 1982, 1987b, 1995) has argued that there are at least two different ways in which formulation processes can adjust word order in sentences without changing the underlying message. One occurs when elements of messages are mapped onto the structures of language, in the selection of grammatical functions. It is here that certain conceptual properties have their influence: Concrete things, animate things, and things already mentioned tend to become subjects. More formally, things that have these properties tend to get mapped to roles higher in the grammatical role hierarchy (i.e., subject > object > indirect object; Keenan & Comrie, 1977). These assignments launch the construction of a sentence frame. It is in spelling out the linkage between sentence frames and words that *lexical* accessibility plays a major role. If a word is easier to access, it can be placed earlier within the local pieces of a structure that is being worked on. But lexical accessibility at best weakly influences grammatical role assignments due to the normal transience of lexical activation in the run-up to speech (e.g., Schriefers, Meyer, & Levelt, 1990).

If the sentence production literature is right about the locus of the effects of lexical accessibility, and the category literature is right about the locus of the effects of typicality, how can we explain results in which typicality affects phrasal ordering only? Our explanation relies on a combination of conceptual and lexical structure, illustrated in Figure 1. Our proposal is that familiar superordinate concepts are associatively linked to their subordinate concepts, with the strength of that link determined by typicality (and likely other variables as well; Barsalou, 1985). Thus, the concept of fruit is more strongly linked to the concept of apple than it is to the concept of lemon (as shown by the thickness of the links in the figure). Word meaning is represented as a weighted linkage between words and conceptual structure (which is very schematically represented in Figure 1 as a single link—see Murphy, 2002, ch. 11 for more detail). Thus, the words *apple* and *lemon* are linked to the nonlinguistic concepts of apples and lemons, respectively, which allows those words to be produced when one is thinking about those concepts.

The right panel of Figure 1 shows a snapshot of the activation of these structures during production. When speakers are producing a sentence involving two instances of fruit, it seems likely that the concept of fruit is activated. Activation flows to related concepts. Since the apple concept is more strongly related to fruit than the lemon concept is, it receives more activation. Activation is then passed down to the lexical items that pick out these concepts, which occur in the sentence. Because of the greater activation of the concept of apples, the word *apple* in turn receives more activation than the word *lemon* and as a result is produced first. So, typicality differences in the conceptual structure lead to activation differences in the lexicon, which is where the ordering is determined.

We did not investigate exactly which components of a lexical entry are involved in typicality ordering. Our suspicion is that it is the lemma (see Levelt, 1989), because that level is directly connected to the lexical concepts that vary in their typicality. A potential problem with this conjecture is that lemmas are often assumed to take part in function assignment, and our account challenges this assumption. Another possibility is that the activation of word forms influences word order within phrases, and we cannot rule out involvement at that level. Our goal was not

to use typicality ordering to investigate structure internal to the lexicon but to illuminate interactions of the lexicon and conceptual structure.

This proposal resolves the contradiction between Kelly et al.'s (1986) identification of typicality as a lexical effect and concepts researchers' claim that it is a conceptual phenomenon. The proposal also resolves the apparent contradiction in results indicating that typicality requires a relevant conceptual framework (Experiment 1) and that it patterns with lexical variables (Experiments 2 and 3). First, our proposal is that the link between typical concepts and their superordinates is in fact conceptual and is (in large part) based on conceptual variables such as the overlap of properties or shared goals, as discussed in the Introduction. However, the associative link between superordinate and subordinate only applies to familiar categories that are represented in long-term memory. Adults have had thousands of exposures to fruit of different kinds and are very familiar with their properties. In American society, thinking of fruit immediately brings to mind apples, oranges, and bananas. However, that is not true of all concepts. In particular, Barsalou's ad hoc categories are (by definition) not stored and so are not strongly linked to their subordinates in semantic memory. When one thinks of things to carry out of a burning house, one must generate exemplars, perhaps through evoking that scenario (Barsalou, 1991), rather than just retrieving instances from memory. Thus, activation cannot quickly spread to the subordinate concept and then to the lexical item in order to influence production order, as it apparently can for familiar taxonomic categories.

Second, this proposal denies that lexical items referring to typical entities have some property that makes them inherently easy to access. That is, (once frequency is controlled for) in a neutral context, words for typical things like *pants* are accessed no faster than words for less typical clothing like *scarf*. In this respect, then, our conclusion is different from that of Kelly et al. (1986). However, the two explanations agree that it is indeed the time course of lexical activation that determines the typicality effect. That is, we agree it is the activation of the word *pants*, which is caused by the conceptual structure, that makes it more likely to be said first—not the conceptual structure alone. It is activation at the bottom level—not the middle level—of Figure 1 that is responsible for typicality's influence on word order.

Why do we need a two-part explanation involving typicality at the conceptual level and activation differences at the lexical level? A pure conceptual explanation does not seem possible, because of our failure to find an ordering effect in ad hoc categories (Experiment 2). Furthermore, the typicality effect is found only within phrases (Experiment 3), a result that is more readily accounted for by lexical activation. A pure lexical explanation is not possible either, because when the conceptual category was not evoked, the typical words were not produced prior to atypical words (different-category condition, Experiment 1). Thus, the typicality effect is an intriguing demonstration of the close link between conceptual structure and lexical items, and explaining it requires reference to the relationship between the two levels of representation.

The final question is why typicality does not influence sentence order if it is indeed a conceptual variable, as it seems to be. If concrete and animate things are more likely to be sentence subjects, why aren't typical things as well? We have two answers for this. The first is that when the two category names are in different sentence constituents, their superordinate categories are probably not as strongly encoded. For example, it would do the participant little good to remember that clothing outsold clothing at Sears Roebuck, because the roles that the sentence distinguishes would be blurred by the identical category name: The gist of the sentence would not be preserved (i.e., "X outsold X" does not make sense). However, the memory representation that two pieces of clothing were needed for the play would be helpful, because the category name corresponds to a single sentence role. In short, people may strongly encode the common category of two nouns when they play the same role in the sentence and not when

they play different roles. Therefore, when the words are in different major constituents, they are not thought of in terms of their superordinate category, and so their typicality in that category would not be computed.

This discussion raises the more general point that in order to explain ordering effects we need to understand the underlying memory structure of the thing being talked about. As speakers often talk about events or objects stored in their memories, the way that the objects are encoded in memory will obviously determine to some degree their descriptions. In this case, the categories used to represent the items mentioned in a sentence will partly determine the order in which things are mentioned.

The second explanation is related, having to do with the communicative reasons for putting things first, or at least earlier (see Bock, Irwin, & Davidson, 2004, for review). The kinds of conceptual factors associated with grammatical roles do not often involve low-level taxonomic categories but instead very broad classes such as animate vs. inanimate objects or humans vs. nonhumans. The importance of these variables can be explained by goals and biases of the speaker. For instance, animate things and concrete things have several natural advantages. They benefit from the ease with which they combine with verbs, compared to inanimate and abstract things (Bock et al., 1992), in simple active sentence structures. That is, the language offers more ways and simpler ways of saying what needs to be said about animate and concrete subjects. At the root of this bias in language is the fact that animates (especially people) are of greatest interest to most people, so that sentence subjects naturally tend to be human. In discourse, the precedence of already-mentioned items comes about through a combination of the maintenance of topic continuity plus the memory advantage in retrieving entities that have recently been mentioned.

In contrast, none of these communicative functions seems to strongly favor typical over atypical basic-level categories. Apples are no better subjects than lemons simply because they are more typical as fruit; photos do not become more interesting than blankets when you carry both out of a burning house. If anything, it is likely that atypical objects attract more interest than typical ones (Brewer, 2000). It would be a dull discourse in which the speakers reported all the typical things they had seen and done that day.

In short, the functions and mechanisms that underlie global sentence ordering do not necessarily apply to typicality. In this sense, it is something of a simplification to refer to all such variables as *conceptual*, in that not every conceptual distinction will necessarily lead to an ordering preference. One can draw conceptual distinctions based on size, morality, or price, but it seems unlikely that any of these will lead to sentence ordering effects. However, the fact that a basic force in many conceptual tasks, typicality, does not influence the ordering of sentence constituents is an important piece of information for framing an account of sentence formulation. Further work is needed to explore other conceptual features and structures in order to better understand the mechanisms underlying their influence—or lack of influence—on language use.

Conclusion—Much has been written on the relationship between word meaning and language, and recent attention has been given to the possibility that words in a language influence what concepts are formed (see e.g., essays in Gentner & Goldin-Meadow, 2003; Gumperz & Levinson, 1996), although the influence is not always as strong as one might expect (Malt, Sloman, Gennari, Shi, & Wang, 1999). Our experiments emphasize a different sort of influence—how conceptual structure and lexical representations interact during processing. How one thinks of a thing and the conceptual representation of that thing in long-term memory apparently influence the ease with which a name for it is retrieved. When one is thinking about fruit, it is easy to retrieve the word *apple*; but when one is not thinking about fruit, retrieving

the same word is not quite as easy. Our results suggest that lexical accessibility and conceptual accessibility alone cannot account for typicality's effects on word order, but that the links between concepts and words formed over years of language use are essential to explaining how typicality influences language production.

Acknowledgements

References

Barsalou LW. Context-independent and context-dependent information in concepts. Memory & Cognition 1982;10:82–93.

Barsalou LW. Ad hoc categories. Memory & Cognition 1983;11:211–227.

Barsalou LW. Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. Journal of Experimental Psychology: Learning, Memory, and Cognition 1985;11:629–654.

Barsalou, LW. The instability of graded structure: Implications for the nature of concepts. In: Neisser, U., editor. Concepts and conceptual development: Ecological and intellectual factors in categorization. Cambridge: Cambridge University Press; 1987. p. 101-140.

Barsalou, LW. Deriving categories to achieve goals. In: Bower, GH., editor. The psychology of learning and motivation, volume 27. San Diego: Academic Press; 1991. p. 1-64.

Barsalou LW, Ross BH. The roles of automatic and strategic processing in sensitivity to superordinate and property frequency. Journal of Experimental Psychology: Learning, Memory, and Cognition 1986;12:116–134.

Battig WF, Montague WE. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology 1969;80:1–46.

Bock JK. The effect of a pragmatic presupposition on syntactic structure in question answering. Journal of Verbal Learning and Verbal Behavior 1977;16:723–734.

Bock JK. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. Psychological Review 1982;89:1–47.

Bock JK. Meaning, sound, and syntax: Lexical priming in sentence production. Journal of Experimental Psychology: Learning, Memory, and Cognition 1986;12:575–586.

Bock JK. An effect of the accessibility of word forms on sentence structures. Journal of Memory and Language 1987a;26:119–137.

Bock, JK. Coordinating words and syntax in speech plans. In: Ellis, AW., editor. Progress in the psychology of language. London: Erlbaum; 1987b. p. 337-390.

Bock JK, Irwin DE. Syntactic effects of information availability in sentence production. Journal of Verbal Learning and Verbal Behavior 1980;19:467–484.

Bock, JK.; Irwin, DE.; Davidson, DJ. Putting first things first. In: Henderson, JM.; Ferreira, F., editors. The integration of language, vision, and action: Eye movements and the visual world. New York: Psychology Press; 2004. p. 249-278.

Bock JK, Warren RK. Conceptual accessibility and syntactic structure in sentence formulation. Cognition 1985;21:47–67. [PubMed: 4075761]

Bock, K. Sentence production: From mind to mouth. In: Miller, JL.; Eimas, PD., editors. Handbook of perception and cognition. Vol. II: Speech, language, and communication. Orlando, FL: Academic Press; 1995. p. 181-216.

Bock K, Loebell H, Morey R. From conceptual roles to structural relations: Bridging the syntactic cleft. Psychological Review 1992;99:150–171. [PubMed: 1546115]

Bomba PC, Siqueland ER. The nature and structure of infant form categories. Journal of Experimental Child Psychology 1983;35:294–328.

Brewer, WF. Bartlett's concept of the schema and its impact on theories of knowledge representation in contemporary cognitive psychology. In: Saito, A., editor. Bartlett, culture and cognition. London: Psychology Press; 2000. p. 68-98.

Clark HH. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior 1973;12:355–359.

Clark HH, Begun JS. The semantics of sentence subjects. Language and Speech 1971;14:34–46. [PubMed: 5580469]

Clark, HH.; Clark, EV. Psychology and language. New York: Harcourt Brace Jovanovich; 1977.

Cooper, WE.; Ross, JR. World order. In: Grossman, RE.; San, LJ.; Vance, TJ., editors. Papers from the parasession on functionalism. Chicago: Chicago Linguistics Society; 1975. p. 63-111.

Crowder, RG. Principles of learning and memory. Hillsdale, NJ: Erlbaum; 1976.

Fenk-Oczlon G. Word frequency and word order in freezes. Linguistics 1989;27:517–556.

Francis, WN.; Kucera, H. Frequency analysis of English usage: Lexicon and grammar. Boston: Houghton Mifflin; 1982.

Garrett, MF. The analysis of sentence production. In: Bower, GH., editor. The psychology of learning and motivation. 9. New York: Academic Press; 1975. p. 133-177.

Garrod S, Sanford A. Interpreting anaphoric relations: The integration of semantic information while reading. Journal of Verbal Learning and Verbal Behavior 1977;16:77–90.

Gentner, D.; Goldin-Meadow, S., editors. Language in mind: Advances in the study of language and thought. Cambridge, MA: MIT Press; 2003.

Gleitman LR, Gleitman H, Miller C, Ostrin R. Similar, and similar concepts. Cognition 1996;58:321–376. [PubMed: 8871343]

Gumperz, JJ.; Levinson, SC., editors. Rethinking linguistic relativity. Cambridge: Cambridge University Press; 1996.

Keenan EL, Comrie B. Noun phrase accessibility and universal grammar. Linguistic Inquiry 1977;8:63–99.

Kelly, MH. On the selection of linguistic options. Psychology Department; Cornell University: 1986. Unpublished PhD dissertation

Kelly MH, Bock JK, Keil FC. Prototypicality in a linguistic context: Effects on sentence structure. Journal of Memory and Language 1986;25:59–74.

Levelt, WJM. Speaking: From intention to articulation. Cambridge, MA: MIT Press; 1989.

Levelt, W.; Maassen, B. Lexical search and order of mention in sentence production. In: Klein, W.; Levelt, W., editors. Crossing the boundaries in linguistics. Dordrecht: D. Reidel; 1981. p. 221-252.

Malt BC, Sloman SA, Gennari S, Shi M, Wang Y. Knowing versus naming: Similarity and the linguistic categorization of artifacts. Journal of Memory and Language 1999;40:230–262.

McDonald JL, Bock K, Kelly MH. Word and world order: Semantic, phonological, and metrical determinants of serial position. Cognitive Psychology 1993;25:188–230. [PubMed: 8482072]

McKoon G, Ratcliff R. Inferences about contextually defined categories. Journal of Experimental Psychology: Learning, Memory, and Cognition 1989;15:1134–1146.

Medin DL, Schaffer MM. Context theory of classification learning. Psychological Review 1978;85:207–238.

Mervis CB, Catlin J, Rosch E. Relationships among goodness-of-example, category norms, and word frequency. Bulletin of the Psychonomics Society 1976;7:283–284.

Mervis CB, Pani JR. Acquisition of basic object categories. Cognitive Psychology 1980;12:496–522. [PubMed: 7418367]

Miller GA, Fellbaum C. Semantic networks of English. Cognition 1991;41:197–229. [PubMed: 1790654]

Morris MW, Murphy GL. Converging operations on a basic level in event taxonomies. Memory & Cognition 1990;18:407–418.

Murphy, GL. Meaning and concepts. In: Schwanenflugel, PJ., editor. The psychology of word meanings. Hillsdale, NJ: Erlbaum; 1991. p. 11-35.

Murphy, GL. The big book of concepts. Cambridge, MA: MIT Press; 2002.

Oldfield RC, Wingfield A. Response latencies in naming objects. Quarterly Journal of Experimental Psychology 1965;17:273–281. [PubMed: 5852918]

Pinker S, Birdsong D. Speakers' sensitivity to rules of frozen word order. Journal of Verbal Learning and Verbal Behavior 1979;18:497–508.

Prat-Sala M, Branigan HP. Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. Journal of Memory and Language 2000;42:168–182.

Proffitt JB, Coley JD, Medin DL. Expertise and category-based induction. Journal of Experimental Psychology: Learning, Memory, and Cognition 2000;26:811–828.

Roediger HL III. Inhibition in recall from cueing with recall targets. Journal of Memory and Language 1973;12:644–657.

Roelofs A. Syllable structure effects turn out to be word length effects: Comment on Santiago et al. (2000). Language and Cognitive Processes 2002;17:1–14.

Rosch E. Cognitive representations of semantic categories. Journal of Experimental Psychology: General 1975;104:192–233.

Rosch E, Mervis CB. Family resemblances: Studies in the internal structure of categories. Cognitive Psychology 1975;7:573–605.

Roth EM, Shoben EJ. The effect of context on the structure of categories. Cognitive Psychology 1983;15:346–378.

Santiago J, MacKay DG, Palma A. Length effects turn out to be syllable structure effects: Response to Roelofs (2002). Language and Cognitive Processes 2002;17:15–30.

Schriefers H, Meyer AS, Levelt WJM. Exploring the time course of lexical access in language production: Picture-word interference studies. Journal of Memory and Language 1990;29:86–102.

Slamecka NJ. An examination of trace storage in free recall. Journal of Experimental Psychology 1968;76:504–513. [PubMed: 5650563]

Wheeldon LR, Monsell S. Inhibition of spoken word production by priming a semantic competitor. Journal of Memory and Language 1994;33:332–356.

Appendices

In each appendix, the relevant category is listed in the first column. In the sentence frame, the two category members are listed in italics—highly typical item first and the less typical item second. In the experiments, the order of these noun phrases varied.

I. EXPERIMENT 1: taxonomically related/unrelated items in conjunctions

Same-category items are presented here. Different-category items were constructed by interchanging noun phrases of different categories, as explained in the text.

Category	Question	Sentence Frame
furniture	What was needed before the play could open?	*A couch* and *a rug* were needed before the play could open.
clothing	What did Nancy make out of denim?	Nancy made *some pants* and *a scarf* out of denim.
tools	What did Tim find on the floor in the closet?	Tim found *a nail* and *a bolt* on the floor in the closet.
clothing	What was needed to complete the costume?	With the addition of *a sock* and *a tie*, the costume was complete.
clothing	What was in the back of Linda's car?	*The skirt* and *the gloves* were in the back of Linda's car.
vehicles	What did the army dispose of after the inspection?	The army disposed of *the jeep* and *the sled* after the inspection.
tools	What was in the abandoned house?	The only things in the abandoned house were *a drill* and *an axe*.
clothing	What was neatly laid out on the table?	*A shirt* and *a belt* were neatly laid out on the table.
clothing	What remained after the tag sale?	Only *a dress* and *a hat* remained after the tag sale.
vehicles	What did Scott dream about?	Scott had a dream about *a bus* and *a skate*.
vehicles	What was locked in the warehouse?	*The van* and *the raft* were both locked in the warehouse.
vegetable	What did Greg bring into the house?	Greg brought *the squash* and *the yams* into the house.
birds	What was part of the display at the children's museum?	*A robin* and *a turkey* were part of the display at the children's museum.
clothing	What did Sarah donate to the Salvation Army?	Sarah donated *a nightgown* and *some sandals* to the Salvation Army.
birds	What was left outside during the storm?	*The sparrow* and *the ostrich* were left outside during the storm.
vehicles	What were big attractions at the fair this year?	Both *the trolley* and *the rowboat* were big attractions at the fair this year.

Category	Question	Sentence Frame
birds	What did Diana mumble about in her delirium?	In her delirium, Diana mumbled about *pigeons* and *penguins*.
tools	What had Eric written books about?	Eric had written books about *hacksaws* and *hatchets*.
furniture	What did the refugees load into the wagon?	The refugees loaded *a bureau* and *a mirror* into the wagon.
birds	What were the symbols of the secret society?	*An eagle* and *a chicken* were the symbols of the secret society.
clothing	What did Dan notice was dirty?	Dan noticed that *the sweatshirt* and *the earmuffs* were quite dirty.
tools	What did Anne discover under the sofa when she was cleaning?	Anne discovered *the hammer* and *the paintbrush* under the sofa when she was cleaning.
weapons	What indicated that something was amiss?	*A dagger* and *some poison* indicated that something was amiss.
clothing	What was shipped together in one box?	*A sweater* and *a slipper* were shipped together in one box.
vehicles	What can be dangerous?	*Taxis* and *rockets* can be dangerous.
fruit	What did the group want price reports for?	The group wanted price reports for *melons* and *pumpkins*.
vegetables	What did Andrew forget on the train?	Andrew forgot *the spinach* and *the mushrooms* on the train.
clothing	What was confiscated by the customs agent?	*The raincoat* and *the mittens* were confiscated by the customs agent.
fruit	What were traditional courtship gifts in Albania?	*Strawberries* and *coconuts* were traditional courtship gifts in Albania.
clothing	What are white?	*Undershirts* and *handkerchiefs* are white.
vegetables	What did Karen buy because they were on sale?	Karen bought some *broccoli* and *sauerkraut* because they were on sale.
tools	What ended up in the cabinet?	Some *screwdrivers* and *sandpaper* ended up in the cabinet.

II. EXPERIMENT 2: ad-hoc categories

Category	Ad Hoc Context	Control Context	Sentence
bad diet foods	The doctor told Jane she could stand to lose a few pounds, so at the buffet what did she eye enviously?	The restaurant was really crowded and Jane did not want to be pushy, so at the buffet what did she eye enviously?	At the buffet Jane eyed the *chocolate* and *celery* enviously.
bad diet foods	Zach's diet was going well, but the food at the party was tempting. Since he knew it was cheating, what did he eat quickly?	Zach's vacation was fun but he was glad to be back. Since he was hungry when he walked in, what did he eat quickly?	Zach ate the *sweets* and *bread* quickly.
bad diet foods	Losing weight was not easy for Diana, but sometimes she could resist eating things. What did she resist yesterday?	Saving money was not easy for Diana, but sometimes she could resist buying things. What did she resist yesterday?	Yesterday Diana resisted the *cookies* and *water*.
bad diet foods	Because he was overweight, people tried to tell Bill what he should and should not eat. What did he decide to have last week?	Because he was shy, people tried to tell Bill what he should and should not order. What did he decide to have last week?	Last week Bill decided to have *pizza* and *pasta* anyway.
birthday present	Anne had to send birthday presents to some friends. What was wrapped up and ready to go?	Anne had to mail a bunch of packages at the post office. What was wrapped up and ready to go?	The *jewelry* and *candy* were wrapped up and ready to go.
birthday present	Scott's coworkers bought some things for his birthday. What did he forget when he left work?	Scott had done some shopping during lunch. What did he forget when he left work?	Scott forgot the *card* and *tie* when he left work.
birthday present	Greg's family was really into birthdays and always got lots of presents. What was his favorite?	Greg's family was really into sketching and always had lots of pictures around. What was his favorite?	The *CD* and the *keychain* were Greg's favorite.
birthday present	Since it wasn't a holiday month, the store had a display for birthday presents. What looked good in the window?	Since it wasn't a holiday month, the store had a display without a theme. What looked good in the window?	The *cake* and *art* looked good in the window.
camping equipment	Camping really wasn't Nancy's idea of a good time. What did she decide to give away?	Spring cleaning wasn't Nancy's idea of a good time. What did she decide to give away?	Nancy decided to give away the *sleeping bag* and *opener* that had been in the basement for years.
camping equipment	Sam remembered that the last time he'd gone camping he'd forgotten what?	Sam remembered that the last time he'd gone to New Jersey he'd forgotten what?	Sam remembered that he'd forgotten the *matches* and the *swatter*.
camping equipment	Amanda didn't really expect the campers to unload the van carefully, so she wasn't surprised to see what on the ground?	Amanda didn't really expect the kids to unload the car carefully, so she wasn't surprised to see what on the ground?	Amanda wasn't surprised to see the *tent* and *pots* on the ground.

Category	Ad Hoc Context	Control Context	Sentence
camping equipment	The campsite was swarming with people. What were they searching for?	The store was swarming with people. What were they searching for?	The people were searching for *knives* and *fuel*.
clothes to wear in snow	On his way home from school Mark got stuck in a huge snow storm. He was glad to have on his what?	On his way home from school Mark got stuck in a huge traffic jam. He was glad to have on his what?	Mark was glad to have on his *scarf* and *vest*.
clothes to wear in snow	Karen worried about her first snowy Montana winter. She wondered if she would need some new what?	Karen worried about her first day on the new job. She wondered if she would need some new what?	Karen wondered if she would need some new *boots* and *skirts*.
clothes to wear in snow	Brandon was packing for his skiing vacation in Colorado. What went into the bag immediately?	Brandon was gathering things to give to the Salvation Army. What went into the bag immediately?	*Mittens* and *shorts* went into the bag immediately.
clothes to wear in snow	After it stopped snowing, Marcia was outside in the cold shoveling the driveway. What was she wearing?	After her guests left, Marcia was outside the house chatting with a neighbor. What was she wearing?	Marcia was wearing a *hat* and a *dress*.
personalities of non-friends	Patricia had strong opinions about what kinds of people she wanted as friends— definitely not what kind of people?	Patricia had strong opinions about what kinds of people would be successful in life—definitely not what kind of people?	Definitely not the *mean* and *loud* ones.
personalities of non-friends	Max was glad to leave his old school. There were some people he couldn't make friends with—the ones who were what?	Max watched the activities of people in his town. There were some people who were busier—the ones who were what?	The ones who were *phony* and *quiet*.
personalities of non-friends	It was difficult for Fred to make friends because he was so picky. Right away, he could spot people who were what?	For years Fred had been working in the public sector. Right away, he could spot people who were what?	Fred could spot people who were *unfriendly* and *sarcastic* right away.
personalities of non-friends	Anya started her new job at the library, but was afraid she wouldn't make any friends. Everyone seemed what?	Anya started her new job at the library and had been busy meeting all her new coworkers. Everyone seemed what?	Everyone seemed *obnoxious* and *forgiving*.
picnic activities	John's family picnic was an enjoyable event. What played a big role?	John's paintings were of unique subjects. What played a big role?	*Talking* and *sleeping* played a big role.
picnic activities	Since it was summer, Sarah organized a picnic outing. What was bound to happen?	Since it was for a class, Sarah watched the movie intently. What was bound to happen?	*Eating* and *tanning* were bound to happen.
picnic activities	Alicia always found the company picnics fun. Usually there was lots of what?	Alicia was trying to guess what would be on the news. Usually there was lots of what?	Usually there was lots of *frisbee* and *reading*.
picnic activities	After Dan and his pals had been at the picnic for several hours what did he suggest doing?	After Dan and his pals had been at the symphony for several hours what did he suggest doing?	Dan suggested doing some *barbecuing* and some *somersaulting*.
take out during fire	In the middle of the night, Laura's house burst into flames. What did she run out with?	A local charity dropped by Laura's house for donations. What did she run out with?	Laura ran out with her *pets* and *clothes*.
take out during fire	When Franklin got home, his house was on fire. What did he run in and grab?	As soon as Franklin got home, he decided to go back out again. What did he run in and grab?	Franklin ran in and grabbed *photos* and *blankets*.
take out during fire	When the smoke detector went off, what did Knut run around yelling?	When the alarm clock went off, what did Knut run around yelling?	Knut ran around yelling, "Where is my *family* and my *camera*?"
take out during fire	Carmen was terrified of a fire in her home. She imagined dire scenes in which she was carrying what?	Carmen was worried about moving to a new house. She imagined dire scenes in which she was carrying what?	Carmen imagined dire scenes in which she was carrying her *children* and *dishes*.
long-distance transport	Dave needed to get from Chicago to San Francisco for a conference. He contemplated going by what?	Dave thought of the short holiday that was coming up. He wanted to go on a trip. He contemplated going by what?	Dave contemplated going by *car* and *bike*.
long-distance transport	Andy and his friends discussed how to get from Seattle to Florida. Travel by what had come up?	Andy and his friends hated a good many things. Travel by what had come up?	Travel by *plane* and by *foot* had come up.
long-distance transport	Shannon wondered how she would get across the country next summer. What did her brother suggest?	Shannon, who was only 5, was trying to think of something that moved for a class project. What did her brother suggest?	Shannon's brother suggested *buses* and *horses*.
long-distance transport	For a wedding, Vicki wanted to go to Atlanta from her home in Connecticut. How did she think she might go?	For no particular reason, Vicki wanted to travel around and see some new places. How did she think she might go?	Vicki thought she might go by *train* and *boat*.
weekend entertainment	Some friends were coming for the weekend, so Nadia was thinking of exciting things to do. So far she'd come up with what?	She was teaching a class on retirement, so Nadia was thinking of some activities. So far she'd come up with what?	So far Nadia had come up with *movies* and *yard work*.

Category	Ad Hoc Context	Control Context	Sentence
weekend entertainment	Since he knew a lot, people often asked Olaf what they should do on the weekends. Often he suggested what?	Since he knew a lot, people often asked Olaf what they should write their papers on. Often he suggested what?	Often he suggested *partying* and *studying*.
weekend entertainment	Linda had been very busy lately, so she decided to give herself a break for the weekend. What did she look for in the paper?	Linda was doing research for a book about the history and culture of her town. What did she look for in the paper?	Linda looked for information on *concerts* and *TV* in the paper.
weekend entertainment	Michael had planned some really great things to do this weekend. What would he be doing?	Michael was overbooked at work for the next several weeks. What would he be doing?	Michael would be going *dancing* and *walking*.

One of the contexts was presented prior to the sentence in the study period. At recall, only the question portion of the context was presented (generally with the proper noun replacing the pronoun).

III. EXPERIMENT 3: taxonomically related items in same-phrase or different-phrase sentences

Category	Question	Same-Phrase Sentence	Different-Phrase Sentence
vehicles	What is it that became affiliated?	Somehow *trolleys* and *rowboats* became affiliated.	Somehow *trolleys* became affiliated with *rowboats*.
vegetables	What bordered the yard quite elegantly?	*Eggplants* and *garlic* bordered the yard quite elegantly.	*Eggplants* bordered the yard with *garlic* quite elegantly.
tools	When the truck went over a bump, what bounced out?	When the truck went over a bump, the *screwdriver* and the *sandpaper* bounced out.	When the truck went over a bump, the *screwdriver* bounced out with the *sandpaper*.
tools	On the right, what was clustered?	The *hacksaws* and the *hatchets* were clustered on the right.	The *hacksaws* were clustered with the *hatchets* on the right.
fruit	What was combined to make the secret sauce?	The secret sauce was made by combining *melons* and *pumpkins*.	The secret sauce was made by combining *melons* with *pumpkins*.
vegetables	In the recipe, what did Andy confuse?	In the recipe Andy confused the *spinach* and the *mushrooms*.	In the recipe Andy confused the *spinach* with the *mushrooms*.
vegetables	What was consolidated into bins?	Leftover *lettuce* and *pickles* were consolidated into bins.	Leftover *lettuce* was consolidated with *pickles* into bins.
weapons	What was the correlation that the spy noticed?	The spy noticed the correlation between *daggers* and *poison*.	The spy noticed the correlation of *daggers* with *poison*.
birds	What was it that dated?	The *eagle* and the *chicken* dated.	The *eagle* dated the *chicken*.
weapons	What is it that it is important to distinguish?	It is important to distinguish *hand grenades* and *razor blades*.	It is important to distinguish *hand grenades* from *razor blades*.
clothing	What was it that fluttered down to the ground mysteriously?	Mysteriously, a *dress* and a *hat* fluttered down to the ground.	Mysteriously, a *dress* fluttered down to the ground with a *hat*.
tools	What was it that Sam fused for his art project?	Sam fused some *sawhorses* and *scissors* for his art project.	Sam fused some *sawhorses* to *scissors* for his art project.
birds	In the zoo's plan, what was to be herded together?	In the zoo's plan, *pigeons* and *penguins* were to be herded together.	In the zoo's plan, *pigeons* were to be herded with *penguins*.
tools	What was it that Linda finally got to intersect?	Linda finally got the *T-square* and the *plumb-line* to intersect.	Linda finally got the *T-square* to intersect with the *plumb-line*.
furniture	In the back room, what was it that was jumbled?	*Bureaus* and *mirrors* were jumbled in the back room.	*Bureaus* were jumbled with *mirrors* in the back room.
clothing	During the wash cycle, what was it that got knotted?	During the wash cycle, the *shirts* and the *belts* got knotted.	During the wash cycle, the *shirts* got knotted with the *belts*.
clothing	What was it that Roberto lined up neatly?	Roberto lined up the *nightgowns* and the *sandals* neatly.	Roberto lined up the *nightgowns* with the *sandals* neatly.
clothing	Historically, the manufacture of what are linked?	Historically, the manufacture of *undershirts* and *handkerchiefs* are linked.	Historically, the manufacture of *undershirts* is linked with *handkerchiefs*.
clothing	During the move, what was it that Greg lumped on the table?	During the move, Greg lumped *skirts* and *gloves* on the table.	During the move, Greg lumped *skirts* with *gloves* on the table.
vehicles	What was massed at the landing zone?	The *taxis* and the *rockets* were massed at the landing zone.	The *taxis* were massed with the *rockets* at the landing zone.
clothing	What was usually packaged together?	*Sweaters* and *slippers* were usually packaged together.	*Sweaters* were usually packaged with *slippers*.
tools	After the earthquake, what was it that was scrambled?	After the earthquake, the *hammers* and the *paintbrushes* were scrambled.	After the earthquake, the *hammers* were scrambled with the *paintbrushes*.

| Category	Question	Same-Phrase Sentence	Different-Phrase Sentence
tools	What was sent by overnight mail?	*A drill* and *an axe* were sent by overnight mail.	*A drill* was sent with *an axe* by overnight mail.
clothing	What did Zoe have to separate?	Zoe had to separate the *raincoats* and the *mittens*.	Zoe had to separate the *raincoats* from the *mittens*.
clothing	During the reorganization, what was it that was shuffled up?	During the reorganization, *parkas* and *aprons* were shuffled up.	During the reorganization, *parkas* were shuffled up with *aprons*.
clothing	Later in the week, what would Emily have to split up?	Later in the week, Emily would have to split up the *socks* and the *ties*.	Later in the week, Emily would have to split up the *socks* from the *ties*.
clothing	What is it that supports the retail business?	*Sweatshirts* and *earmuffs* support the retail business.	Along with *sweatshirts*, *earmuffs* support the retail business.
clothing	What swayed in the breeze?	Some *pants* and a *scarf* swayed in the breeze.	Some *pants* swayed with a *scarf* in the breeze.
fruit	What did Audrey swirl in her bowl?	Audrey swirled the *strawberry* and the *coconut* in her bowl.	Audrey swirled the *strawberry* with the *coconut* in her bowl.
vehicles	In the garage, what was it that was touching?	In the garage the *van* and the *raft* were touching.	In the garage the *van* was touching the *raft*.
birds	What was it that took a walk around the lake?	A *robin* and a *turkey* took a walk around the lake.	A *robin* took a walk around the lake with a *turkey*.
birds	In the next match, what is it that will wrestle?	The next match will feature *a sparrow* and *an ostrich* wrestling.	The next match will feature *a sparrow* wrestling *an ostrich*.
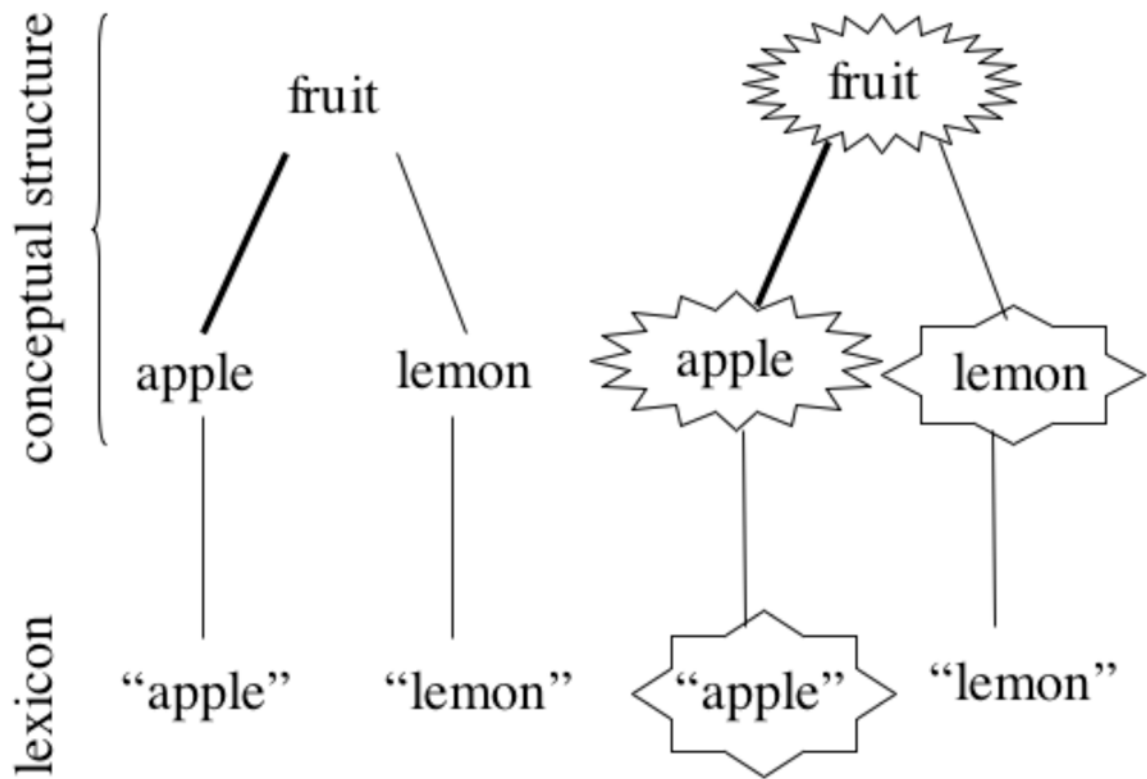
Figure 1.
The left panel shows the connections between concepts in the domain of fruit and their lexical counterparts. Internal details of the lexical entry are omitted. Note that the more typical fruit, apple, is more strongly connected to its superordinate (as indicated by the thicker line). The right panel shows the spreading of activation (amount of activation represented by the number of vertices in the star shape) when the concept of fruit is activated. Considerable activation flows to the apple concept, due to its typicality, and in turn to the word *apple*. Less activation flows to the lemon concept and therefore less to the word *lemon*. Thus, conceptual structure mediates lexical accessibility.
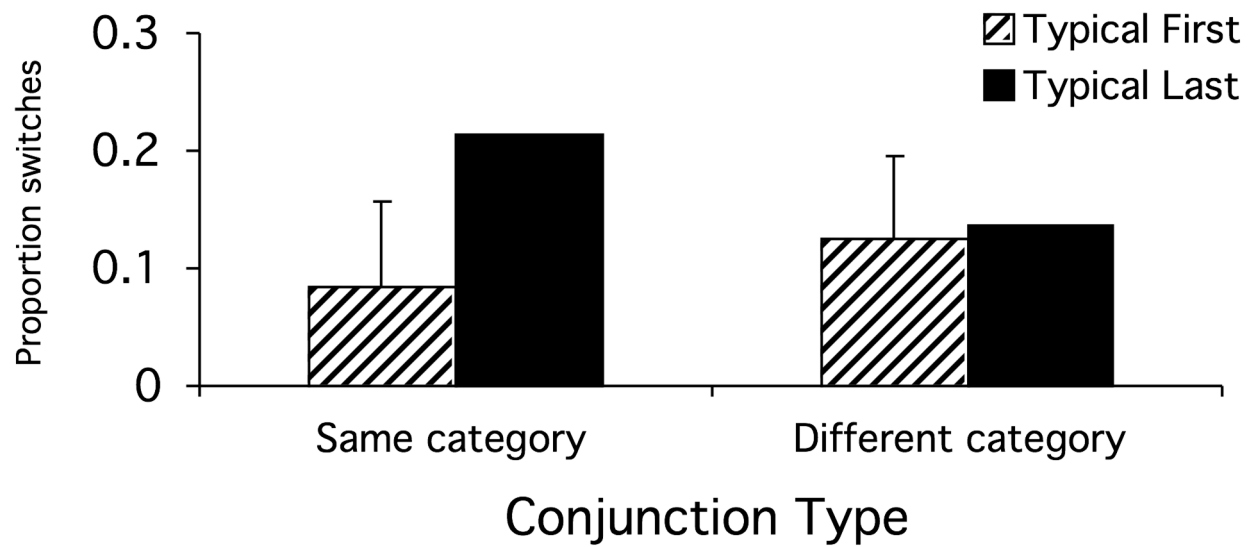
Figure 2.
Typical-leader effect, Experiment 1. Switch proportions for typical-first and typical-last conjunctions of same- and different-category nouns. (Error bars are halfwidths of the 95% confidence interval for pairwise planned contrasts of condition means.) The typical leader effect is the proportion switches in the typical-last (LH) condition minus that in the typical-first (HL) condition.

Table 1

Sample Materials and Procedure in Experiments 1–3

STUDY PHASE: listen to questions and related answers, played from audiotape
  Q: What did Andrew forget on the train?
    A: Andrew forgot the spinach and mushrooms on the train.
  Q: What was needed before the play could open?
    A: Some pants and a rug were needed before the play could open.
  … (for 15 question-answer pairs within each block)
TEST PHASE: hear question, recall answer in writing
  Q: What did Andrew forget on the train?
  (18 s for response)
  Q: What was needed before the play could open?
  (18 s for response)
  … (for 15 answers within each block)
STUDY and TEST PHASES for remaining blocks

Note. The number of items per block and time parameters varied slightly across experiments. The values shown are for Experiment 1. See text for details.

Table 2

Normative Properties of Noun Phrases in Experiments 1–3

Noun Phrase Type	Mean (and range)			
	Typicality rating	Typicality Difference Within Pairs [*]	Frequency	Frequency Difference Within Pairs (High – Low)
Experiment 1				
Highly typical	1.64 (1.0 – 2.4)	2.7 (1.2 – 3.9)	12 (0 – 63)	0 (−37 – 41)
Less typical	4.35 (3.3 – 5.9)		12 (0 – 71)	
Experiment 2				
Highly typical	6.24 (5.3 – 7.0)	3.4 (1.1 – 5.9)	78 (0 – 620)	6 (−476 – 584)
Low typical	2.83 (1.0 – 4.9)		72 (0 – 486)	
Experiment 3				
Highly typical	1.75 (1.0 – 2.4)	2.6 (1.2 – 3.6)	9 (0 – 63)	−2 (−37 – 17)
Low typical	4.39 (3.6 – 5.9)		11 (0 – 71)	

[*] In Experiment 2, the typicality scale ran from 1 (atypical) to 7 (typical), whereas in the other experiments, the scale (matching Rosch, 1975) was reversed. For ease of comparison, the typicality difference scores in each experiment are calculated so as to be positive.

Table 3

Proportions of Switches During Recall to Alternative Forms when Typical Items Occurred First or Last in Presented Sentences (Experiments 1, 2, and 3, including post-tests and replications)

Experiment	Condition	Switch proportions		Typical-leader effect
		Typical last	**Typical first**	
Experiment 1 (taxonomic categories)	Same category	.214	.085	.129
	Different category	.136	.134	.002
Experiment 2 (ad hoc categories)	Ad hoc category evoked	.106	.097	.009
	No category evoked	.120	.175	−.055
	Post-test on NP-order preference	.510[*]	.490[*]	.020
	Post-test on no-context recall	.231	.224	.007
	Post-test on taxonomic category cued recall	.160	.086	.073
	Post-test on ad hoc category cued recall	.110	.110	−.0004
Experiment 3 (major phrases in sentences)	Same phrase (presentation scoring)	.229	.175	.055
	Different phrase (presentation scoring)	.089	.125	−.036
	Same phrase (production scoring)	.224	.149	.076
	Different phrase (production scoring)	.117	.187	−.070
Experiment 3 replication	Same phrase (production scoring)	.160	.109	.051
	Different phrase (production scoring)	.073	.057	.016

[*] These numbers are not switch proportions, but proportion choosing a particular order.

Note. Typical-leader effect is not always identical to the difference of the typical-first and typical-last columns due to rounding error.