Prototypicality in Sentence Production

Kristine H. Onishi
McGill University

Gregory L. Murphy
New York University

Kathryn Bock
University of Illinois at Urbana-Champaign

Please address all correspondence to:
  Gregory L. Murphy
  Department of Psychology
  New York University
  6 Washington Place, 8th floor
  New York, NY 10003
  gregory.murphy@nyu.edu
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 but have no effect on determining the order of major constituents. This is a pattern usually associated with lexical variables (like word frequency), and yet typicality is usually thought of as a conceptual feature. Experiment 1 examined whether an appropriate conceptual framework was necessary to yield the typicality effect, and it was. 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 phrase or in different constituents (e.g., *The turkey and the robin danced* vs. *The turkey danced with the robin*). Typicality affected word order only when the two words appeared in the same phrase. This constellation of results is consistent with an account in which typicality has an 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 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 indirect 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 *Pat was inspired by Lee* and *Lee was inspired by Pat*. The 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 structure and meaning, and whether those factors are the same when order does not make a difference to meaning.

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 that concept (Bock, 1987; 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, 1987b; Bock & Irwin, 1980; Bock & Warren, 1985; Kelly, Bock, & Keil, 1986; McDonald, Bock, & Kelly, 1993; Prat-Sala & Branigan, 2000). Conceptual accessibility is the ease of bringing a concept to mind. For example, previous mention and concreteness both increase conceptual accessibility. References to conceptually accessible entities are likely occur in subject position (McDonald et al. 1993), followed by object (Bock & Warren, 1985) and adjunct positions. Subjecthood is also correlated with other properties that lead to conceptual
accessibility. For example, concrete entities (Bock & Warren, 1985) and animate entities (Bock, Loebell, & Morey, 1992; McDonald et al., 1993) tend to appear in subject position more often than abstract and inanimate entities do (Clark & Begun, 1971). If something is conceptually accessible, it resides in a rich conceptual network. People are more likely to talk about it, the representation is more easily activated, and it is more likely to end up in the subject position of a sentence, where speakers put the things they are talking about.

Lexical accessibility has to do with the ease of retrieving the word form. 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 faster 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 sentence structure has already been set by the time the word forms are being 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, word form variables associated with lexical accessibility appear to influence word order within phrases but not 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 (Rosch & Mervis, 1975;
Typicality and word order

Murphy, 1991), there is some reason to believe that it affects language production through lexical accessibility (Kelly, Bock, & Keil, 1986). In the experiments below, we investigated whether its influence is likewise due to conceptual accessibility. 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 some prior explanations of how it affects word order. We describe three experiments 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 done with categories, including categorization, 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 the words that refer to them. One piece of evidence for this is that typicality effects are found in artificial categories even when category members do not have lexical labels (e.g., Medin & Schaffer, 1978; Rosch & Mervis, 1975). It also influences the processing of preverbal infants (Bomba & Siqueeland, 1983). Learning of new categories is easier with typical examples (Mervis & Pani, 1980). Thus, typicality strongly influences performance 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 prestored with particular lexical items, which suggests 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 associated to the bird concept and does not have properties associated to contrasting concepts. In contrast, a penguin lacks many properties associated to bird (flies, perches on trees, migrates, small size) and possesses 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, 2002). 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 related to word forms. Thus, it is widely assumed that the explanation of typicality effects in tasks using words (comprehension, ratings, induction, and so on) derives from relations among the conceptual representations associated to 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. For example, Kelly et al. (1986) looked at 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. So 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. Sentences were presented with the highly typical things (apple or shirt) either first or second.

In producing the remembered sentences, speakers tended to change the word order 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 is the pattern of results attributed to lexical variables such as frequency and length (e.g., Bock & Warren, 1985) in contrast to 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 effect on the linguistic representation of the word, such as its resting activation level or prominence. Indeed, words such as *robin* could simply be tagged as “typical,” which could affect order during production. It is not clear how the lexical representation of typical words differs from atypical words, but the essence of this hypothesis, the *lexical account*, is that typical and atypical words come to be differentially accessible. 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 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 past 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 frequency and length.

Perhaps the apparent lexical level effect of typicality on the ordering of words arose through the activation of the superordinate concept which in turn activated 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 a fruit at the store. 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. On this
Typicality and word order

explanation, the apparent lexical effect of typicality is driven by effects at the conceptual level, which have implications for the lexical level.

However, another potential explanation for the apparent lexical effects in Kelly et al. (1986) could be the method of measuring typicality. They used production norms (Battig & Montague, 1969) which tend to correlate with word frequency (Mervis, Catlin, & Rosch, 1976). Indeed, 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 the word order effect should only appear in an appropriate conceptual framework. In contrast, if something about typical words like apple and robin makes them 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. For example, *Shirts outsold hats* is more about shirts than about hats and so is not uniformly interchangeable with *Hats were outsold by shirts*. 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 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: Established Categories**

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 to vary conceptual information.

Participants read 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. 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, apples’ typicality 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. That is, the word *apple* is more accessible than the word *scarf*, and so it should appear first, regardless of what superordinate categories are 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 the conceptual variable that is said to account for the preference in word order. 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 et al., 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, noun phrases that differed in typicality were identified. The noun phrases 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 (Table 2) but had the same number of syllables and same 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; see Table 2). 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).

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 noun phrases in these 32 different-category conjunctions were matched on syllable number and stress. 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 with 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, typical last) was crossed with the type of category factor (same- or different-category) to make four types of items.

**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 same-category frame for the pants and scarf pair would
be about Nancy (*Nancy made some pants and a scarf out of denim*), but the same-category answer frame for the other items in the quadruplet might be: *A couch and a rug were needed before the play could open.* The lack of mention of the superordinate category was different from Kelly et al.’s (1986) method, in which the relevant superordinate category was always mentioned. In addition to allowing us to use the same sentence frames for the same- and different-category conditions, it reduced the likelihood that participants would use the superordinate category as a strategic cue to recalling the category members in the sentence. That is, it reduced the chance of finding a typicality effect due to an overt guessing strategy that has little to do with normal sentence production.

Each of these two frames was also used for 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 (e.g., *scarf, rug*) was placed in the other frame. This resulted in the Nancy frame also containing *pants and rug* (or *rug and pants*) and the frame about the opening of the play also containing *couch and scarf* (or *scarf and couch*).

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 44.100 kHz, 16 bits per sample. A question and answer pair was treated as 4 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 would
hear. 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 used a Macromedia paste function to yield 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 items were presented in four blocks, each block containing two items of each 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 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” (loose instructions) and half were told to try to produce these answers verbatim (verbatim instructions). 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, we predicted more switching in the typical-last condition (to put highly typical items earlier), so the typical-leader effect was calculated as switch proportion for typical-last minus 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.

Scores item analysis. Analyses were conducted using both participants (F₁) and items (F₂) as the random factor (Clark, 1973) to examine generalizability across each. 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 out of 8); (F₁(1,79) = 5.95, p < .02, F₂(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 analyses of variance. 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 no interaction of
order with the category type. Ideally, there would be no difference in number of scorable items across the four conditions, since that would indicate all types of items were equally difficult to remember. Differences in memorability of the answers to the questions may affect how likely people are to switch the target noun phrases during recall. 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 not differ greatly in how easy they were to remember.

*Typical-leader effect.* Figure 1 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) \((F_1(1,79) = 6.36, p < .02; F_2(1,15) = 33.87, p < .0001)\). For the same-category condition the typical-first effect was reliably different from 0 \((t_1(79) = 3.45, p < .001, t_2(15) = 6.53, p < .001)\), but the different-category condition 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 end up putting 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 active 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 was necessary for 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 preferred to 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 rely on conceptual information. One potential contribution to this effect is that the words referring to concepts may be associated to one another and to the superordinate concept. As we mentioned above, thinking of the concept fruit might activate the word apple, 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. One way to examine this possibility is to use categories that are unlikely to have lexical level links because they are not well established in memory. Ad hoc categories, which are centered 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. If conceptual accessibility drives the typicality effect, 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 of highly typical entities. This should then increase activation for words referring to those entities. Thus, Experiment 2 had the goal of further investigating whether typicality effects on word order would occur only when the relevant category was 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 101 carried out in three post
hoc studies. 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. See Table 2 for the items’ typicalities and word frequencies.

When entities in the ad hoc category also fell into the same taxonomic category and ratings (from Rosch, 1975) were available, items were chosen (if possible) such that the taxonomic typicalities did not follow the same pattern as the ad hoc typicalities. For example, entities in the ad hoc category of things to wear in the snow were all members of the taxonomic category clothing. We tried to select items such as scarf and vest for which one (scarf) was more typical than the other (vest) in the ad hoc category, but less typical in the standard taxonomic category. It was not always possible to do this, because some taxonomic categories did not have ratings (food) and for others (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 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 a particular ad hoc 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 16.2, as in Experiment 1. A question and answer pair was treated as 4 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-first effect were defined as before.

In one of the test lists an item was inadvertently omitted. This item (mittens, shorts) was subsequently removed from all analyses. Due to this error, 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 greater tendency to put the less typical item first in the no-category condition ($M = -.06$) than in the category condition ($M = .01; F_1(1,79) = 3.57, p < .07; F_2(1,34) = 1.45, p < .24$). 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. Although not the reason for performing this posttest, its 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; $M = .006; ts < 1$). There was also no tendency for the typical-first ($M = .53$) to have more scorable items than the typical-last condition ($M = .52; Fs < 1$). 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 typicality ordering of ad hoc category members, it is appropriate to ask whether the underlying memory advantage for typical category members is found for them 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.

Sixty-six participants heard a category label followed by the conjunctions from both Experiment 2 and Experiment 1 (as comparison items). For example, they 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. 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; \( M = -.0004; F(1,65) < 1 \)), again suggesting that there was no inherent lexical bias in the items in Experiment 2. For the taxonomic category items from Experiment 1, there was a reliable typical-first effect (\( M = .07; F(1,65) = 6.19, p < .02 \)) which was reliably different from the typical-first effect for the ad hoc category items (\( F(1,65) = 4.53, p < .04 \)). In short, ad hoc categories do not seem to lead to earlier recall of highly typical entities, though taxonomic categories do.

**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 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, it is unlikely that there is a direct link between the words *photo* and *blanket* or between either and *things to take from a burning house*, since the relevant category is infrequent. The results suggested that typicality within ad hoc categories did not affect the ordering of words in sentence production. People were no more likely to produce *photos and blankets* than *blankets and photos* whether the category of things to take from a burning house was relevant or not. In the face of the null effect of ad hoc category evocation on the tendency to reorder words within phrases, it is important to note that the experimental manipulation did have an 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 argues that the
manipulation of context was sufficiently strong.

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. If a lexical linkage is critical to this ordering, but the lexical linkages do not establish the conceptual conditions for the assignment of grammatical functions, it would help to explain the earlier finding by Kelly et al. (1986) that typicality did not influence the choice of a constituent as subject or object. Since that choice seems to be driven by discourse or conceptual factors such as animacy and agency, the lexical link would not influence subjecthood. However, the Kelly et al. finding is a null effect, and it is possible that better controlled stimuli would reveal an effect of typicality after all. Experiment 3 attempts to provide a more powerful test of this possibility.

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. Experiment 3 addressed whether typicality can influence the order of the same kinds of 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., *The turkey and the robin danced*) with the same grammatical role (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., *The turkey danced with the robin*), giving the entities different grammatical roles. For these sentences an effect of typicality would be manifest in a reordering of these major sentence constituents 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).

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.
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 affecting sentence production during processes that are not involved in the assignment of grammatical functions, like most lexical factors.

**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 were used (always from the same taxonomic category). These noun phrases were placed into 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 from 80 candidate sentences. 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 which had 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., The turkey and the robin danced, the same-phrase condition) and its matching sentence frame was as similar in meaning as possible while placing the two noun phrases in different major constituents (e.g., The turkey danced with the 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, since noun phrase pairs always came from the same taxonomic category. 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 the 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 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 separately. A single recorded token was chosen for each unit, and the units were then pasted together to create the tokens the participants would hear. Where possible, pieces of the answer frame were used in four final items. 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).

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 list versions had the question-answer pairs in the same fixed random order, with the conditions that were instantiated by the items within the lists varying systematically over versions. 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 scored in two ways. Strict scoring was similar to scoring in Experiments 1 and 2. Under a strict scoring criterion, items were scored only if they were produced using the same type of sentence frame in which they had been presented. That is, if a 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. In 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 second method of scoring 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 strict scoring criterion (on average, fewer than 2.4 scorable responses per participant per condition). 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. (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 sentences (not recall), examining the data ignoring condition of presentation was informative.

Under the production scoring method, an item was scored as long as it contained both the target noun phrases or an acceptable substitution (as before). Scorable responses were categorized as same-phrase responses if the target noun phrases were in the same phrase and connected by and (or and comma were acceptable substitutes) or different-phrase responses if the target noun phrases were not in the same major sentence constituent.

**Strict 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: $M = 2.34$; $F$’s < 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 strict scoring method, sentences heard in the different-phrase condition but produced with noun phrases in the same phrase were not scored.

**Strict scoring: Typical-leader effect.** Although the effect was in the expected direction, the
typical-leader effect was not significantly different for the same-phase \( (M = .06) \) and different-phase \( (M = -.04) \) 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-phase responses \( (M = 6.8) \) was higher than that for different-phase 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-first effect for same-phase responses \( (M = .08) \) than for different-phase responses \( (M = -.07) \). The difference was significant by participants \( (F_1(1,59) = 6.78, p < .02) \) but not by items \( (F_2(1,31) = 1.82, p < .19) \). The typical-first effect for the same-phase responses was different from 0 in both participant and item analyses \( (t_1(59) = 2.86, p < .006, t_2(1,31) = 2.19, p < .04) \). The negative typical-first effect for the different-phase responses did not differ from 0 \( (t_1(59) = -1.37, p < .18; t_2(1,31) = -.38, p < .71) \). 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-first effect in the different-phase items in Experiment 3 is the small number of observations. Participants preferred to produce sentences with same- rather than different-phase uses of the target phrase. In an effort to increase the likelihood of different-phase responses, we conducted a replication study using only the different-phase 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-phase responses.\(^1\) Since only the different-phase items were used, there were only two lists. The
experimental procedure and scoring were identical to that of Experiment 3. After excluding the data from one nonnative speaker and from two other participants to restore counterbalancing, the data from 50 participants remained.

*Production scoring: Scorable item analysis.* Each participant heard 32 critical phases in the different-phrase format. Of these, an average of 15.9 per participant were scorable. Although participants heard only different-phrase items, the majority of their scorable responses were still in the same-phrase format ($M = 3.2$ in different phrases vs. $M = 4.7$ in same phrases; $F_1(1, 49) = 12.07, p < .002; F_2(1, 31) = 4.75, p < .04$). There was no difference for typical-first ($M = 4.0$) and typical-last ($M = 4.0$) items ($F$s < 1) with a marginal interaction between order and phrase type produced in the subject analysis only ($F_1(1, 49) = 3.92, p < .06; F_2(1, 31) = 2.13, p > .15$) (the larger number of same-phrase items was more marked when the sentences presented the typical item first than when the less typical item was first).

*Production scoring: Typical-first effect.* The typical-first 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-first effect for the same-phrase responses was reliably different from 0 by participants ($t_1(49) = 2.02, p < .05$) but not by items ($t_2(1, 31) < 1$). The typical-first effect for the different-phrase responses did not differ reliably from 0 ($t_1(49) < 1; t_2(1, 31) = 1.62, p < .12$). 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-first effect for same-phrase responses than for different-phrase responses by participant ($M$ difference = -.03; $F_1(1, 108) = 6.21, p < .02$) but not by item ($F(1, 31) < 1.0$). The typical-first effect for the same-phrase responses was also different from 0 ($t_1(109) = 3.30, p < .002$), but the negative typical-first effect for the different-
phrase responses was not ($t_1(109) = -1.0, p > .30$). There was no reliable effect of experiment ($F$s < 1) nor interaction with experiment ($F_1(1,108) = 2.33, p < .13; F_2(1,31) = 2.76, p < .11).

Discussion

Experiment 3 disclosed no tendency for typicality to influence the ordering of major constituents of a sentence even in the combined analyses of over 100 participants. 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 within-phrase and across-phrase materials, using the same 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 phrase, but, in a new comparison, not in sentences containing the noun phrases in different phrases. That is, errors tended to result in *Robins and turkeys danced* rather than *Turkeys and robins danced*, but there was no tendency to prefer *Robins danced with turkeys* over *Turkeys danced with robins*. 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.

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 jibe with an account of typicality in which it acts on mechanisms involved more with the ordering of words than with the assignment of grammatical roles. We address this further in the General Discussion.

General Discussion
Three experiments examined the effect of category typicality on word order in sentence production. Experiments 1 and 3 demonstrated 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 activated, 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. This is consistent with the finding that category information is context sensitive (Barsalou, 1982; Roth & Shoben, 1983) and that features evoked by particular words depend on contextual information (Anderson & Ortony, 1975; Anderson, Pichert, Goetz, Schallert, Stevens, & Trollip, 1976; Barclay, Bransford, Franks, McCarrell, & Nitsch, 1974). If typicality were related only to lexical-level information, a word like pants, which is highly typical in its standard taxonomic category, should always be more accessible than a word like rug, which is not very typical in its standard taxonomic category. Experiment 1 demonstrated 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 frequency.

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 (e.g., graded membership), they do not seem to 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 may not spontaneously activate typical subordinate categories. Post-Test 3 for 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. This supports our hypothesis about the long-term memory representation of ad hoc and taxonomic categories.

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 *Robins and turkeys danced* over *Turkeys and robins danced*, there was no preference for *Robins danced with turkeys* over *Turkeys danced with robins*, 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.

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 conceptual accessibility has its impact: Things that we have something to say about become subjects. More formally, things that are high in conceptual accessibility 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 and the linkage of words to positions in the frame. It is here that lexical accessibility plays a role. If a word form is easier to access, it can be placed earlier in the structures that are being worked on. But lexical accessibility cannot influence the ordering of those larger structures.

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 our results? Our explanation relies on a combination of conceptual and lexical structure, illustrated in Figure 2. Our proposal is that familiar superordinate categories are associatively linked to their subordinate categories, with the strength of that link determined by typicality (and likely other variables as well; Barsalou, 1985). Thus, the concept of apples is more strongly linked to the concept of fruit than is the concept of lemons (as shown by the thickness of the links in the figure). Word meaning is represented as a linkage between word forms and conceptual structure (which is only very schematically represented in Figure 2 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 2 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.

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 apples, oranges, and bananas to mind. However, that is not true of all concepts. In particular, Barsalou’s ad hoc categories are (by definition) not stored and so 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 aid recall of a learned sentence, as it apparently can for familiar taxonomic categories.

Second, this proposal denies that typical lexical items have some property that makes them inherently easy to access. That is, once frequency is controlled for, words for typical items like pants are not accessed faster than less typical clothing words 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 lexical activation that determines the typicality effect. That is, we agree it is the accessibility 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. That is, it is activation at the bottom level—not the middle level—of Figure 2 that is responsible for word order.

Why do we need a two-part explanation involving typicality at the conceptual level and accessibility 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 variables. 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.

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 think (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 an item and the conceptual representation of that item 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 so 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.
References


Typicality and word order


*Journal of Verbal Learning and Verbal Behavior, 18*, 497-508.


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 experiment, the order of these two items varied.

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

<table>
<thead>
<tr>
<th>Category</th>
<th>Question</th>
<th>Sentence Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>furniture</td>
<td>What was needed before the play could open? A couch and a rug were needed before the play could open.</td>
<td>A couch and a rug were needed before the play could open.</td>
</tr>
<tr>
<td>clothing</td>
<td>What did Nancy make out of denim?</td>
<td>Nancy made some pants and a scarf out of denim.</td>
</tr>
<tr>
<td>tools</td>
<td>What did Tim find on the floor in the closet?</td>
<td>Tim found a nail and a bolt on the floor in the closet.</td>
</tr>
<tr>
<td>clothing</td>
<td>What was needed to complete the costume?</td>
<td>With the addition of a sock and a tie, the costume was complete.</td>
</tr>
<tr>
<td>clothing</td>
<td>What was in the back of Linda's car?</td>
<td>The skirt and the gloves were in the back of Linda's car.</td>
</tr>
<tr>
<td>vehicles</td>
<td>What did the army dispose of after the inspection?</td>
<td>The army disposed of the jeep and the sled after the inspection.</td>
</tr>
<tr>
<td>tools</td>
<td>What was in the abandoned house?</td>
<td>The only things in the abandoned house were a drill and an axe.</td>
</tr>
<tr>
<td>Category</td>
<td>Question</td>
<td>Answer</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>clothing</td>
<td>What was neatly laid out on the table?</td>
<td><em>A shirt and a belt</em> were neatly laid out on the table.</td>
</tr>
<tr>
<td>clothing</td>
<td>What remained after the tag sale?</td>
<td><em>Only a dress and a hat</em> remained after the tag sale.</td>
</tr>
<tr>
<td>vehicles</td>
<td>What did Scott dream about?</td>
<td><em>Scott had a dream about a bus and a skate.</em></td>
</tr>
<tr>
<td>vehicles</td>
<td>What was locked in the warehouse?</td>
<td><em>The van and the raft</em> were both locked in the warehouse.</td>
</tr>
<tr>
<td>vegetable</td>
<td>What did Greg bring into the house?</td>
<td><em>Greg brought the squash and the yams</em> into the house.</td>
</tr>
<tr>
<td>birds</td>
<td>What was part of the display at the children's museum?</td>
<td><em>A robin and a turkey</em> were part of the display at the children's museum.*</td>
</tr>
<tr>
<td>clothing</td>
<td>What did Sarah donate to the Salvation Army?</td>
<td><em>Sarah donated a nightgown and some sandals</em> to the Salvation Army.</td>
</tr>
<tr>
<td>birds</td>
<td>What was left outside during the storm?</td>
<td><em>The sparrow and the ostrich</em> were left outside during the storm.</td>
</tr>
<tr>
<td>vehicles</td>
<td>What were big attractions at the fair this year?</td>
<td><em>Both the trolley and the rowboat</em> were big attractions at the fair this year.</td>
</tr>
<tr>
<td>birds</td>
<td>What did Diana mumble about in her delirium?</td>
<td><em>In her delirium, Diana mumbled about pigeons and penguins.</em></td>
</tr>
<tr>
<td>tools</td>
<td>What had Eric written books about?</td>
<td><em>Eric had written books about hatchets and hacksaws.</em></td>
</tr>
<tr>
<td>Typicality and word order</td>
<td>Page 52</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td><strong>furniture</strong></td>
<td>What did the refugees load into the wagon?</td>
<td>The refugees loaded a <em>bureau</em> and a <em>mirror</em> into the wagon.</td>
</tr>
<tr>
<td><strong>birds</strong></td>
<td>What were the symbols of the secret society?</td>
<td><em>An eagle</em> and a <em>chicken</em> were the symbols of the secret society.</td>
</tr>
<tr>
<td><strong>clothing</strong></td>
<td>What did Dan notice was dirty?</td>
<td>Dan noticed that the <em>sweatshirt</em> and the <em>earmuffs</em> were quite dirty.</td>
</tr>
<tr>
<td><strong>tools</strong></td>
<td>What did Anne discover under the sofa when she was cleaning?</td>
<td>Anne discovered the <em>hammer</em> and the <em>paintbrush</em> under the sofa when she was cleaning.</td>
</tr>
<tr>
<td><strong>weapons</strong></td>
<td>What indicated that something was amiss?</td>
<td>A <em>dagger</em> and <em>some poison</em> indicated that something was amiss.</td>
</tr>
<tr>
<td><strong>clothing</strong></td>
<td>What was shipped together in one box?</td>
<td>A <em>sweater</em> and a <em>slipper</em> were shipped together in one box.</td>
</tr>
<tr>
<td><strong>fruit</strong></td>
<td>What can be dangerous?</td>
<td><em>Melons</em> and <em>pumpkins</em> can be dangerous.</td>
</tr>
<tr>
<td><strong>vehicles</strong></td>
<td>What did the group want price reports for?</td>
<td>The group wanted price reports for <em>taxis</em> and <em>rockets</em>.</td>
</tr>
<tr>
<td><strong>vegetables</strong></td>
<td>What did Andrew forget on the train?</td>
<td>Andrew forgot the <em>spinach</em> and the <em>mushrooms</em> on the train.</td>
</tr>
<tr>
<td><strong>clothing</strong></td>
<td>What was confiscated by the customs agent?</td>
<td>The <em>raincoat</em> and the <em>mittens</em> were confiscated by the customs agent.</td>
</tr>
<tr>
<td><strong>fruit</strong></td>
<td>What were traditional courtship gifts in</td>
<td><em>Coconuts</em> and <em>strawberries</em> were traditional courtship gifts in</td>
</tr>
</tbody>
</table>
Typicality and word order

Albania?

Albania.

clothing What are white?

Undershirts and handkerchiefs are white.

vegetables What did Karen buy because they were on sale?

Karen bought some sauerkraut and broccoli because they were on sale.

tools What ended up in the cabinet?

Some sandpaper and screwdrivers ended up in the cabinet.

II. EXPERIMENT 2: ad-hoc related/unrelated items in conjunctions

<table>
<thead>
<tr>
<th>Category</th>
<th>Ad Hoc Context</th>
<th>Control Context</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>bad diet foods</td>
<td>The doctor told Jane she could stand to lose a few pounds, so at the buffet what did she eye enviously?</td>
<td>The restaurant was really crowded and Jane did not want to be pushy, so at the buffet what did she eye enviously?</td>
<td>At the buffet Jane eyed the chocolate and celery enviously.</td>
</tr>
<tr>
<td>bad diet foods</td>
<td>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?</td>
<td>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?</td>
<td>Zach ate the sweets and bread quickly.</td>
</tr>
<tr>
<td>bad diet foods</td>
<td>Losing weight was not easy for Diana, Saving money was not easy for Diana,</td>
<td>Yesterday Diana resisted the</td>
<td></td>
</tr>
</tbody>
</table>
but sometimes she could resist eating things. What did she resist yesterday?

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.

The jewelry and candy were wrapped up and ready to go.

Scott forgot the card and tie when he left work.

Greg's family was really into sketching and always had lots of pictures around. What was Greg's favorite?

The CD and the keychain were Greg's favorite.

Since it wasn't a holiday month, the cake and art looked good.
Typicality and word order

present

store had a display for birthday presents. store had a display without a theme. in the window.

What looked good in the window? What looked good in the window?

camping

Camping really wasn't Nancy's idea of a good time. Nancy decided to give away the sleeping bag and opener that had been in the basement for years.

equipment
good time. What did she decide to give away? 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

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.

equipment
gone camping he'd forgotten what? gone to New Jersey he'd forgotten what? forgotten the matches and the swatter.

Sam remembered that the last time he'd gone to New Jersey he'd forgotten what? forgotten the matches and the swatter.

camping

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 the tent and pots on the ground.

equipment
to unload the van carefully, so she wasn't surprised to see what on the ground? 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.

camping

The campsite was swarming with people. The people were searching for knives and fuel.

equipment

The store was swarming with people. The people were searching for knives and fuel.

clothes
to

On his way home from school Mark got Mark was glad to have on his
wear in snow  stuck in a huge snow storm. He was glad to have on his what?
clothes to  Karen worried about her first snowy wear in snow  Montana winter. She wondered if she would need some new what?
clothes to  Brandon was packing for his skiing wear in snow  vacation in Colorado. What went into the bag immediately?
clothes to  After it stopped snowing, Marcia was wear in snow  outside in the cold shoveling the driveway. What was she wearing?
personalities  Patricia had strong opinions about what of non-friends  kinds of people she wanted as friends—
definitely not what kind of people?
personalities of non-friends

Max was glad to leave his old school. 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

There were some people he couldn't make friends with—the ones who were what?

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 *unfriendly* and *sarcastic* right away.

personalities of non-friends

Anyone 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 *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. *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

Alicia always found the company Alicia was trying to guess what would Usually there was lots of
activities  picnics fun. Usually there was lots of  be on the news. Usually there was lots  frisbee and reading.
what?
what?

picnic  After Dan and his pals had been at the  After Dan and his pals had been at the  Dan suggested doing some
activities  picnic for several hours what did he  symphony for several hours what did he  barbecuing and some
suggest doing?  suggest doing?  somersaulting.
suggest doing?
suggest doing?

take out  In the middle of the night, Laura's house  A local charity dropped by Laura's  Laura ran out with her pets
during fire  burst into flames. What did she run out  house for donations. What did she run  and clothes.
with?  out with?
with?

take out  When Franklin got home, his house was  As soon as Franklin got home, he  Franklin ran in and grabbed
during fire  on fire. What did he run in and grab?  decided to go back out again. What did  photos and blankets.
he run in and grab?

take out  When the smoke detector went off, what  When the alarm clock went off, what  Knut ran around yelling,
during fire  did Knut run around yelling?  did Knut run around yelling?  "Where is my family and my
camera?"
during fire  Carmen was terrified of a fire in her  Carmen was worried about moving to a  Carmen imagined dire scenes
home. She imagined dire scenes in  new house. She imagined dire scenes in  in which she was carrying her
<table>
<thead>
<tr>
<th>Activity</th>
<th>Distance</th>
<th>Consideration</th>
<th>Mode of Travel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which she was carrying</td>
<td>Typicality and word order</td>
<td>children and dishes.</td>
<td></td>
</tr>
<tr>
<td>Long-distance transport</td>
<td>Dave needed to get from Chicago to San Francisco for a conference. He contemplated going by what?</td>
<td>Dave thought of the short holiday that was coming up. He wanted to go on a trip. He contemplated going by what?</td>
<td></td>
</tr>
<tr>
<td>Long-distance transport</td>
<td>Andy and his friends discussed how to get from Seattle to Florida. Travel by what had come up?</td>
<td>Andy and his friends hated a good many things. Travel by what had come up? had come up.</td>
<td></td>
</tr>
<tr>
<td>Long-distance transport</td>
<td>Shannon wondered how she would get across the country next summer. What did her brother suggest?</td>
<td>Shannon, who was only 5, was trying to think of something that moved for a class project. What did her brother suggest?</td>
<td></td>
</tr>
<tr>
<td>Long-distance transport</td>
<td>For a wedding, Vicki wanted to go to Atlanta from her home in Connecticut. How did she think she might go?</td>
<td>For no particular reason, Vicki wanted to travel around and see some new places. How did she think she might go?</td>
<td></td>
</tr>
<tr>
<td>Weekend entertainment</td>
<td>Some friends were coming for the weekend, so Nadia was thinking of</td>
<td>She was teaching a class on retirement, so Nadia was thinking of some with movies and yard work.</td>
<td></td>
</tr>
</tbody>
</table>
Typicality and word order

exciting things to do. So far she'd come up with what? activities. So far she'd come up with what?

weekend 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?

entertainment weekend 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?

entertainment weekend 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?

III. EXPERIMENT 3: taxonomically related/unrelated items in conjunctions and non-conjunctions

<table>
<thead>
<tr>
<th>Category</th>
<th>Question</th>
<th>Conjunction Sentence</th>
<th>Non-Conjunction Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>vehicles</td>
<td>What is it that became</td>
<td>Somehow trolleys and rowboats became</td>
<td>Somehow trolleys became affiliated with</td>
</tr>
</tbody>
</table>
affiliated? affiliated. rowboats.

vegetables What bordered the yard quite elegantly? Eggplants and garlic bordered the yard quite elegantly.

tools When the truck went over a bump, what bounced out? When the truck went over a bump, the sandpaper and the screwdriver bounced out.

tools On the right, what was clustered? The hacksaws and the hatchets were clustered on the right.

fruit What was combined to make the secret sauce? The secret sauce was made by combining melons and pumpkins.

vegetables In the recipe, what did Andy confuse? In the recipe Andy confused the spinach with the mushrooms.

vegetables What was consolidated into bins? Leftover lettuce and pickles were consolidated into bins.

weapons What was the correlation that the spy noticed? The spy noticed the correlation between daggers and poison.
<table>
<thead>
<tr>
<th>Category</th>
<th>Sentence 1</th>
<th>Sentence 2</th>
<th>Sentence 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>birds</td>
<td>What was it that dated?</td>
<td>The <em>eagle</em> and the <em>chicken</em> dated.</td>
<td>The <em>eagle</em> dated the <em>chicken</em>.</td>
</tr>
<tr>
<td>weapons</td>
<td>What is it that it is important to distinguish <em>hand grenades</em> It is important to distinguish <em>hand grenades</em> distinguish?</td>
<td>and <em>razor blades</em>.</td>
<td>from <em>razor blades</em>.</td>
</tr>
<tr>
<td>clothing</td>
<td>What was it that fluttered down to the ground mysteriously?</td>
<td>Mysteriously, a <em>dress</em> and a <em>hat</em> fluttered down to the ground.</td>
<td>Mysteriously, a <em>dress</em> fluttered down to the ground with a <em>hat</em>.</td>
</tr>
<tr>
<td>tools</td>
<td>What was it that Sam fused for his art project?</td>
<td>Sam fused some <em>sawhorses</em> and <em>scissors</em> for his art project.</td>
<td>Sam fused some <em>sawhorses</em> to <em>scissors</em> for his art project.</td>
</tr>
<tr>
<td>birds</td>
<td>In the zoo's plan, what was to be herded together?</td>
<td>In the zoo’s plan, <em>pigeons</em> and <em>penguins</em> were to be herded together.</td>
<td>In the zoo’s plan, <em>pigeons</em> were to be herded with <em>penguins</em>.</td>
</tr>
<tr>
<td>tools</td>
<td>What was it that Linda finally got to intersect?</td>
<td>Linda finally got the <em>T-square</em> and the <em>plumb-line</em> to intersect.</td>
<td>Linda finally got the <em>T-square</em> to intersect with the <em>plumb-line</em>.</td>
</tr>
<tr>
<td>furniture</td>
<td>In the back room, what was it that was jumbled?</td>
<td><em>Bureaus</em> and <em>mirrors</em> were jumbled in the back room.</td>
<td><em>Bureaus</em> were jumbled with <em>mirrors</em> in the back room.</td>
</tr>
<tr>
<td>clothing</td>
<td>During the wash cycle, what was it that got knotted?</td>
<td>During the wash cycle, the <em>shirts</em> and the <em>belts</em> got knotted.</td>
<td>During the wash cycle, the <em>shirts</em> got knotted with the <em>belts</em>.</td>
</tr>
<tr>
<td>Category</td>
<td>Question</td>
<td>Sentence 1</td>
<td>Sentence 2</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>clothing</td>
<td>What was it that Roberto lined up neatly?</td>
<td>Roberto lined up the <em>nightgowns</em> and the <em>sandals</em> neatly.</td>
<td>Roberto lined up the <em>nightgowns</em> with the <em>sandals</em> neatly.</td>
</tr>
<tr>
<td>clothing</td>
<td>Historically, the manufacture of what are linked?</td>
<td>Historically, the manufacture of <em>undershirts</em> and <em>handkerchiefs</em> are linked.</td>
<td>Historically, the manufacture of <em>undershirts</em> is linked with <em>handkerchiefs</em>.</td>
</tr>
<tr>
<td>clothing</td>
<td>During the move, what was it that Greg lumped on the table?</td>
<td>During the move, Greg lumped <em>skirts</em> and <em>gloves</em> on the table.</td>
<td>During the move, Greg lumped <em>skirts</em> with <em>gloves</em> on the table.</td>
</tr>
<tr>
<td>vehicles</td>
<td>What was massed at the landing zone?</td>
<td>The <em>taxis</em> and the <em>rockets</em> were massed at the landing zone.</td>
<td>The <em>taxis</em> were massed with the <em>rockets</em> at the landing zone.</td>
</tr>
<tr>
<td>clothing</td>
<td>What was usually packaged together?</td>
<td><em>Sweaters</em> and <em>slippers</em> were usually packaged together.</td>
<td><em>Sweaters</em> were usually packaged with <em>slippers</em>.</td>
</tr>
<tr>
<td>tools</td>
<td>After the earthquake, what was it that was scrambled?</td>
<td>After the earthquake, the <em>hammers</em> and the <em>paintbrushes</em> were scrambled.</td>
<td>After the earthquake, the <em>hammers</em> were scrambled with the <em>paintbrushes</em>.</td>
</tr>
<tr>
<td>tools</td>
<td>What was sent by overnight mail?</td>
<td><em>A drill</em> and <em>an axe</em> were sent by overnight mail.</td>
<td><em>A drill</em> was sent with <em>an axe</em> by overnight mail.</td>
</tr>
<tr>
<td>clothing</td>
<td>What did Zoe have to separate?</td>
<td>Zoe had to separate the <em>raincoats</em> and the <em>mittens</em>.</td>
<td>Zoe had to separate the <em>raincoats</em> from the <em>mittens</em>.</td>
</tr>
<tr>
<td>Typicality and word order</td>
<td>Page 64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>clothing</strong></td>
<td><strong>During the reorganization, what was it that was shuffled up?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During the reorganization, <strong>parkas</strong> and <strong>aprons</strong> were shuffled up.</td>
<td>During the reorganization, <strong>parkas</strong> were shuffled up with <strong>aprons</strong>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>clothing</strong></td>
<td><strong>Later in the week, what would Emily have to split up?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Later in the week, Emily would have to split up the <strong>socks</strong> and the <strong>ties</strong>.</td>
<td>Later in the week, Emily would have to split up the <strong>socks</strong> from the <strong>ties</strong>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>clothing</strong></td>
<td><strong>What is it that supports the retail business?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sweatshirts</strong> and <strong>earmuffs</strong> support the retail business.</td>
<td>Along with <strong>sweatshirts</strong>, <strong>earmuffs</strong> support the retail business.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>clothing</strong></td>
<td><strong>What swayed in the breeze?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some <strong>pants</strong> and some <strong>scarves</strong> swayed in the breeze.</td>
<td>Some <strong>pants</strong> swayed with some <strong>scarves</strong> in the breeze.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>fruit</strong></td>
<td><strong>What did Audrey swirl in her bowl?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audrey swirled the <strong>strawberry</strong> and the <strong>coconut</strong> in her bowl.</td>
<td>Audrey swirled the <strong>strawberry</strong> with the <strong>coconut</strong> in her bowl.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>vehicles</strong></td>
<td><strong>In the garage, what was it that was touching?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the garage the <strong>van</strong> and the <strong>raft</strong> were touching.</td>
<td>In the garage the <strong>van</strong> was touching the <strong>raft</strong>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>birds</strong></td>
<td><strong>What was it that took a walk around the lake?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A <strong>robin</strong> and a <strong>turkey</strong> took a walk around the lake.</td>
<td>A <strong>robin</strong> took a walk around the lake with a <strong>turkey</strong>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>birds</strong></td>
<td><strong>In the next match, what is it?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The next match will feature a <strong>sparrow</strong> and...</td>
<td>The next match will feature a <strong>sparrow</strong>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
that will wrestle?  

_ an ostrich wrestling. _

wrestling _ an ostrich. _
Author Note

This work was supported by NIH grants MH41704, HD21011, and MH66089, NSF grants SBR98-73450 and BCS0214270, and NIH training grant T32 MH1819990. Correspondence may be addressed to Kristine Onishi, Department of Psychology, Stewart Biology Building, 1205 Dr. Penfield Avenue, McGill University, Montreal, H3A 1B1 CANADA.

Footnotes

1. Thanks to Scott Watter for suggesting this idea.
Table 1

*Sample Materials and Procedure in Experiments 1-3*

STUDY PHASE, Block 1: 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)

TEST PHASE, Block 1: 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)

STUDY and TEST PHASES for remaining blocks
Table 2

Normative properties of noun phrases in each experiment

<table>
<thead>
<tr>
<th>Noun Phrase Type</th>
<th>Typicality rating</th>
<th>Typicality Difference Within Pairs (Low - High)</th>
<th>Frequency</th>
<th>Frequency Difference Within Pairs (High - Low)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Typicality</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High typical</td>
<td>1.64 (1.0 - 2.4)</td>
<td>2.7 (1.2 - 3.9)</td>
<td>12 (0 - 63)</td>
<td>0 (-37 - 41)</td>
</tr>
<tr>
<td>Less typical</td>
<td>4.35 (3.3 - 5.9)</td>
<td></td>
<td>12 (0 - 71)</td>
<td></td>
</tr>
<tr>
<td>Experiment 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High typical</td>
<td>6.24 (5.3 - 7.0)*</td>
<td>3.4 (1.1 - 5.9)*</td>
<td>78 (0 - 620)</td>
<td>6 (-476 - 584)</td>
</tr>
<tr>
<td>Low typical</td>
<td>2.83 (1.0 - 4.9)*</td>
<td></td>
<td>72 (0 - 486)</td>
<td></td>
</tr>
<tr>
<td>Experiment 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High typical</td>
<td>1.75 (1.0 - 2.4)</td>
<td>42.6 (1.2 - 3.6)</td>
<td>9 (0 - 63)</td>
<td>-2 (-37 - 17)</td>
</tr>
<tr>
<td>Low typical</td>
<td>4.39 (3.6 - 5.9)</td>
<td></td>
<td>11 (0 - 71)</td>
<td></td>
</tr>
</tbody>
</table>

*In Experiment 2, the typicality scale ran from 1 (atypical) to 7 (typical), whereas in the other experiments, the scale (from Rosch’s, 1975, ratings) was reversed. For ease of comparison, the difference scores in Experiment 2 are calculated as High - Low.
Figure Captions

1. Switch proportions for typical-first and typical-last conjunctions of same- and different-category nouns, Experiment 1.

2. 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 (indicated as 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. The left panel shows the connections between concepts in the domain of fruit and their lexical counterparts. Note that the more typical fruit, apples, is more strongly connected to its superordinate (as indicated by the thicker line). The right panel shows the spreading of activation (indicated as 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.