When Pseudowords Become Words - Effects of Learning on Orthographic Similarity Priming

Georg Dorffner (GEORG@AI.UNIVIE.AC.AT)
Dept. of Medical Cybernetics and Artificial Intelligence
University of Vienna
Freyung 6-2, A-1010 Vienna, Austria
and Austrian Research Institute for Artificial Intelligence

Catherine L. Harris (CHARRIS@BU.EDU)
Psychology Department, Boston University
Boston, MA 02215

Abstract

This paper investigates empirical predictions of a connectionist model of word learning. The model predicts that, although the mapping between word form and meaning is arbitrary (thus rendering words as being symbols in the semiotic sense), novel pseudowords will be able to prime the concepts corresponding to word forms that are orthographically similar. If, however, pseudowords acquire meaning through an arbitrary mapping, this priming should be reduced. Two experiments support this hypothesis. Pseudowords, derived from and thus orthographically similar to English words, primed a categorization task involving those similar words. After a subsequent learning phase, in which subjects are asked to learn meanings for the pseudowords, this priming disappears. This interplay between iconic and symbolic use of words is proposed to emerge from connectionist learning procedures.

Introduction

This work is concerned with how humans respond to novel words and how they learn meanings for them. According to semiotics (Sebeok, 1994), words of any human language are mostly symbolic in their nature of being signs. This refers to the observation that the form of a word usually does not reflect its meaning in any way. Language users implicitly know that phonological and orthographic similarity between words need not imply semantic similarity. [1] Even a single differing letter (or phoneme) can mean an arbitrary difference in meaning. This abstraction from stimulus similarities may be at the heart of symbolic behavior.

Such abstraction from similarity runs against our usual way of treating stimuli. When facing a novel situation (e.g. attempting to get to downtown Boston by subway) one is trained to rely on similarities to other situations (e.g. taking the subway to downtown Vienna), but also to attend to slight differences (e.g. that one must buy tokens instead of tickets in Boston) and modify one's behavior accordingly. When facing new words, this strategy is of no help. On perceiving a new word token (e.g. the written string 'tiser'), one can either decide to view it as being *identical* to a known word (say, 'tiger', the 's' being just a typo) or conclude that one does not know anything about the word's meaning. It is not possible to take the representation of the closest word's meaning (the concept *tiger*) and modify it according to the perceived differences.

Nevertheless, we argue, humans are not perfect in pure arbitrary processing of symbols, but tend to associate meanings the "natural" way, based on similarities to other stimuli. Thus, when seeing the novel word 'tiser' the meaning of 'tiger' is activated to some extent, such that further processing (e.g. aspects concerning animals) are primed. This, however, does not apply to the same extent to words that exist in one's language and are equally similar. Responses to the word 'timer', although not more different to 'tiger' than 'tiser', will not activate the concept *tiger* or prime the concept *animal*.

This observation suggests the following hypothesis: nonwords are treated according to their similarities to existing words, while known words are relatively immune to the influence of words in their similarity neighborhood. This hypothesis was motivated by work with a connectionist model for word learning (Dorffner, 1996; Dorffner et al., 1996). This model consists of two components that categorize sensory stimuli and one component which learns links between the two other components (see figure 1). The model can identifies visually presented objects and visually (or acoustically) presented word tokens by categorizing them into classes and then builds a link to establish the word-meaning mapping.

The links between words and their meaning implement the arbitrary nature of the observed word-meaning mappings in humans. However, an arbitrary links of this type go against similarity-sensitive activation mechanisms which are the hallmark of connectionist networks. If not explicitly trained otherwise, a neural network, such as a multilayer perceptron (Rumelhart et al., 1996) will treat each novel stimulus according to its similarities to previously trained stimuli. Therefore, to implement arbitrary word-meaning mappings, alternative connectionist mechanisms are needed.

In Dorffner's model, word-meaning mappings are implemented via winner-take-all strategies and special learning algorithms. These mechanisms can overwrite or suppress the usual similarity-sensitive process. If these mechanisms are not fully employed (for instance, at the beginning of learning or with noisy inputs) similarity sensitivity can enter the scene again.

This duality between connectionist mechanisms in the

[1] To be more precise, one should speak of *morphemes* instead of *words*, since different inflected or derived word forms do in fact show meaningful overlap. Since in this paper we are mainly concerned with words consisting of just one morpheme, we will continue using 'word' synonymously to 'morpheme'

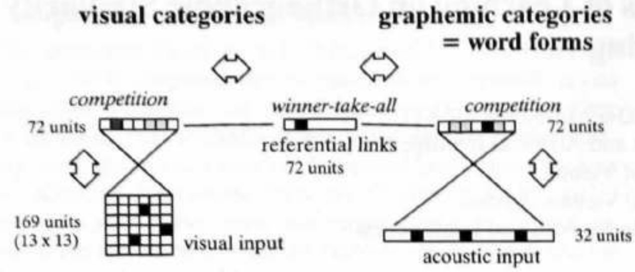

Figure 1: A model for word learning, consisting of two inputs, two categorization layers and a layer linking two category representations

word learning model intuitively seems to reflect the above-mentioned duality of humans treating environmental stimuli, on one hand, and treating language input, on the other. The fact that in connectionist networks, similarity-sensitivity is the more "natural", underlying mechanism, which has to be suppressed when it comes to language, raises the hypothesis that human cognitive processes might be similar. With respect to language, this would mean that humans tend to apply similarity-sensitive associations to novel language inputs, unless or until they have explicitly learned not to. To test this hypothesis we have designed a psycholinguistic experiment probing subjects' responses to novel (pseudo-) words and changes in those responses when those words acquire meaning.

Experiment 1

Method

Subjects. 19 undergraduate students and volunteers from Boston University participated in the experiment. The students participated in return for course credit. All subjects were native English speakers or learned English before the age of 11.

Apparatus. Stimuli were presented and responses recorded using the PSYSCOPE software (Cohen et al., 1993) on Macintosh computers at Boston University and the University of Vienna.

Procedure. The experiment consisted of three phases. Phase 1 and 3 were identical and consisted of a priming and categorization task. One out of a list of 12 possible pseudowords[2] was used as the prime and presented visually for 1500 ms. The target was an English word, which subjects were asked to categorize as quickly as possible. The six possible categories were *Animal, Body part, Clothing, Furniture, Fruit,* and *Vehicle.*

The priming phase consisted of 122 prime–target pairs. For half of the pairs, the prime originated in a word from the same category as the target (e.g. arple (from 'apple') – banana; called a *related pair*). For the other half or the pairs, the category of the word the prime originated in was unrelated to the target category (e.g. traim (from 'train') – bed; called an *unrelated pair*). Prime–target pairs were presented in random order.

Before the priming task in phase 1 subjects were asked to familiarize themselves with the six categories. An example of a word for each category (not occurring in the list of target words) was given. Subjects were further instructed to attend to the pseudoword primes at each trial, letting them "sink in" and perhaps imagine a possible meaning.

Subjects were told that reaction times were important and that they should respond as quickly as possible. Reaction times between the onset of each target and the uttering of the category name by the subject were recorded. The following trials were excluded from analysis:

- for each category, the first trial that contained a target word from that category; this was done to account for the problem that initially subjects were not totally familiar with each category.
- responses that were incorrectly recorded
- incorrect responses (although they occurred very rarely)
- statistical outliers

Phase 2 of the experiment consisted of a learning phase, in which subjects should learn some rough meaning of the new words. To mimic a "natural" way of learning word meanings, the new words were presented in the context of utterances in colloquial English, presented visually on the screen as sentences in quotation marks. The learning phase consisted of 84 trials. Each utterance or short dialog presented at one trial contained a placeholder (_______) which substituted for one of the new words. The subjects' task was to guess the correct word that should be filled in instead of the placeholder. For that, the list of the twelve possible words was always visible on the right part of the screen, in alphabetical order. After picking a word, subjects were asked to say the word out aloud. Subsequently, the correct word appeared on the screen, as well as the same sentence, now with the word instead of the placeholder, in the lower left part of the screen. Subjects were asked to read the sentence with the word in it one more time. When they were done, the next sentence would appear.

Meanings for the new words were created by assigning the words to concepts in one of three categories:

1. concepts that directly correspond to English words, such as *mistake*, or *liar*; this is called the *synonym condition*.
2. concepts that are generally known to English native speakers, but no direct English words exist, such as *stuff one has to scrape off the wall or flat surfaces to make it smooth*; this is called the *known concept condition*.
3. concepts that exist in other languages (Rheingold, 1988), but are usually not in the conceptual repertoire of English native speakers, such as *things one is likely to forget when making up a budget*; this is called the *unknown concept condition*.

For each concept, a total of seven sentences (or short dialogues) was created to lead to 84 different stimuli for the learning phase. Each of the sentences was presented exactly once. Sentences were presented in random order in seven blocks, each block containing one sentence for each of the twelve concepts. Words were randomly assigned to meanings

[2]Two of those 12 pseudowords had to be excluded from the analysis, since they showed unexpected interference with the categorization task, mainly due to strong sound symbolisms.

to create a mapping which was held constant for each subject. No more than two subjects were confronted with the same mapping.

For phase 2, subjects were told that they would learn the meanings of the new pseudowords. Subjects were told that the sentences they were about to see were utterances in colloquial English, supposedly somebody said to somebody else. Subjects were told that in the beginning they would have to guess, but that they should let their intuitions guide their answers. Subjects were told that they should not try to turn this into an elaborate guessing game, with taking notes or evaluating hypotheses, but respond by what intuitively came to their mind. They were further told that there could be an English word that corresponds to the meaning conveyed by the sentence, but that there need not be one – thus they should not try hard to guess which English word the new word substituted for.

Before phase 3 (the second priming phase, identical to phase 1), subjects were asked to go over the list of all new words and report what they thought they would mean. All responses (e.g. "has something to do with pregnancy") were written down by the experimenter, including a mark for no responses ("I have no idea what that word means").

Phase 3 was identical to phase 1. Subjects were again asked to attend to the pseudoword and then categorize the English word.

The entire experiment lasted between 45 and 60 minutes.

Stimulus materials. For the priming phases, words were chosen from the list used in Irwin and Lupker (1983), plus a few others, with the restriction that words should be no longer than six characters. Pseudowords were constructed the following way. Twelve English words from the same six categories (two per category) were taken, and one letter other than the initial letter was altered such that the most likely pronunciation of the pseudoword would not differ by more than one phoneme from the original English word. With one exception, all pseudowords were five characters long. Examples of the pseudowords appear in the appendix.

Most sentences (or short dialogues) used to set the context for the meaning did not implicitly define all properties of the concept to be learned, but the whole of the seven sentences per concept did. Care was taken to avoid using cue words (such as 'feeling' for concept "What one feels for someone one once loved but no loner does") more than once (or in rare cases, twice). Placeholders did not always appear on the same position, but were randomly placed at the beginning, in the middle, and at the end of sentences. The division of meanings into three categories was done for further investigation into how quickly subjects learn the meaning for new words. For the questions relevant in this paper, they were of no further interest.

Results

For each of the 19 subjects, mean response times were calculated for each of the 10 pseudowords in each of the four conditions, *before learning / related pairs*, *before learning / unrelated pairs*, *after learning / related pairs*, and *after learning / unrelated pairs*. Data from one subject was excluded since there were missing trials for one of the pseudowords. Thus, 18 times 10 times 4 cases were input for the anova.

Condition		RT (unrel.–rel.)	Difference
before	related	962	
before	unrelated	994	32
after	related	965	
after	unrelated	963	-2

Table 1: Mean response time RT (in ms) in experiment 1

Condition		RT (unrel.–rel.)	Difference
before	related, learned	947	
	unrelated, learned	993	46
	related, not-learned	957	
	unrelated, not-learned	997	40
after	related, learned	968	
	unrelated, learned	971	3
	related, not-learned	978	
	unrelated, not-learned	953	-25

Table 2: Mean response time (in ms) for the eight conditions including *word learned / not learned* in experiment 1.

Table 1 lists the resulting mean response times for the four conditions

Before learning, categorization times were 32 ms faster in the related condition, compared to the unrelated condition, a statistically reliable difference, $F=6.9$, $p < 0.02$. After the learning phase, RTs in the unrelated and related conditions differed by only 2 ms. The difference in amount of priming before and after learning was statistically significant, $F=6.9$, $p < 0.02$.

We then investigated whether the lack of priming in the post-learning task varied depending on whether the subject was able to retrieve a word meaning after the learning phase, as measured by responses to the question "Can you tell me what this word means". For each subject, pseudowords were divided into two groups: pseudowords for which the subject had retrieved some meaning (irrespective of whether this meaning was the intended one), called the *word learned* condition; and words for which the subject gave no response, called the *word not learned* condition. Two subjects had to be excluded, since they had learned a meaning for all ten pseudowords. Thus the anova received 17 times 2 times 4 inputs.

The resulting mean response times for the eight conditions are shown in table 2.

For the learned words, there is no priming after learning. For the non-learned words, the facilitation tends to turn into an inhibition (difference of -25 ms) after the learning-phase. Furthermore, response times for related learned words are higher after learning (968 ms) than before learning (947 ms), suggesting a kind of inhibition due to learning. However, these differences did not reach statistical significance.

Therefore, to further test the hypothesis that the decrease in priming for related stimuli is due to the learning of meanings (and not due to some other reason, such as habituation with the stimuli), we performed a second, control, experiment.

Condition		RT (unrel.–rel.)	Difference
before	related	1031	
before	unrelated	1086	55
after	related	974	
after	unrelated	1014	40

Table 3: Mean response time RT (in ms) in experiment 2

Experiment 2

Method

Subjects 14 undergraduate psychology students from Boston University participated for course credit. All of them were native English speakers.

Procedure As in experiment 1, experiment 2 consisted of three phases. The first and the third phase were identical to the respective phases in experiment 1. Instead of presenting semantic contexts for the pseudowords, the second phase consisted of a letter completion task involving all pseudowords. The aim was to mimic as much as possible the middle phase of experiment 1 without giving subjects a chance to deduce any meaning for the pseudowords.

Each trial in this task consisted of the presentation of a string containing one to three letters of one of the pseudowords in their correct positions and placeholders '_' standing in for the missing letters (e.g. 'a___e' for 'arple'). This string was presented in the upper half of the screen. As in the corresponding task of experiment 1, the list of all possible pseudowords was visible on the right half of the screen. Subjects were asked to pick the pseudoword from the list which would be a correct completion for the string, by saying out loud the word. There was always a unique solution. Whenever they picked a word, the correct word was displayed on the screen. Subjects then pressed a button to continue with the next string. For each pseudoword, seven different strings existed, leading to a total of 84 stimuli. Response time was recorded, although not further investigated.

The way this phase was designed, each pseudoword was the correct solution the same number of times as in the corresponding phase of experiment 1, thus had to be pronounced the same number of times (on average), and was presented on the screen the same number of times. Subjects, were not told that this is a control experiment, but instead that this experiment tested responses to novel words.

Results

Response times were evaluated for the first and second priming task (condition *before / after letter completion* instead of *before / after learning*) the same way as in experiment 1. Table 3 depicts the results.

A priming effect in the related pairs condition is again clearly visible (difference in RT = 55 ms). This facilitation, however, is still present in the second priming task (condition after letter completion; difference in RT = 40 ms). While the overall difference due to facilitation is statistically significant ($F=5.81$, $p < 0.05$), the difference between the two priming tasks is not.

Discussion

The results from experiments 1 and 2 supported the predictions generated from Dorffner's (1996) word learning model. The experimental subjects showed similarity priming to the novel word stimuli. After meanings were acquired for the new words, this priming was eliminated.

Psycholinguists have long noted that the initial stages of processing in non-words is similar to words. Reading a non-word appears to cause partial activation of the words in its orthographic neighborhood, just as happens when real words are read. The extent to which real words are influenced by their orthographic and phonological neighbors has also been shown to vary as a function of their frequency (Seidenberg and McClelland, 1989). For example, pronunciation of the low frequency word 'sweat' is slow because its orthographic neighbors are pronounced differently ('meat,' 'treat'). But irregularly spelled high frequency words like 'have' and 'done' are not slowed by their many regularly spelled neighbors. If we accept the common assumption that frequency indexes learning, then these effects of frequency are consistent with predictions of Dorffner's (1996) model.

Previous studies have demonstrated similarity priming of pseudowords in word identification tasks (Feustel et al., 1983; Rueckl, 1990). However, little focus has been laid on the possible influence of pseudowords on semantic tasks. Rueckl and Olds (1993) investigated how repetition priming involving pseudowords can benefit when subjects are asked to learn meanings for the pseudowords. In their experiments, pseudowords were paired with English words and subjects were told that they should learn these mappings in a vocabulary learning task. Each pair was presented up to three times. Results showed that pseudowords that acquired pseudomeanings this way showed repetition priming in a word identification task.

Rueckl and Olds (1993) also gave a connectionist explanation for their findings. They noted that during learning associative mappings can support each other if their outputs are similar, but can inhibit each other when outputs are dissimilar. Since word-meaning mappings are arbitrary, outputs of these mappings are dissimilar, in general, and therefore inhibition occurs leading to effects of repetition priming (since repeated stimuli reach asymptote only slowly). This inhibition is suggested by our results distinguishing between pseudowords which had apparently acquired meaning and those that had not (table 2), although statistical significance could not be shown. Since we did not study repetition priming, we cannot say whether a prime like 'arple' would prime the categorization of 'apple' after repeated presentation. Instead, we could demonstrate that similarity priming can also extend to semantic tasks when a pseudoword does not have a meaning attached.

Whittlesea and Cantwell (1987) investigated the influence of assigning meanings to pseudowords in word perception (i.e., the facilitation of recognizing letters in known words). Their experimental setting with respect to learning meanings is similar to ours, in that arbitrary meanings are assigned by definitions and sample sentences, and in a control experiment experience with the pseudowords was gained without meaning assignment. Subjects identified pseudowords more accurately when associated with meanings. Semantic priming was not investigated.

Dagenbach et al. (1990) investigated how much learning is necessary to enter new lexical entries into semantic memory, such as to produce automatic priming. They showed that automatic episodic priming does not occur in a one-hour experiment, where subjects were given definitions for new words. We could show that within the time frame of our experiment (up to 60 mins), a lasting effect of learning meaning with respect to semantic priming is visible – something Dagenback et al. (1990) did no investigate.

Looking at the connectionist model briefly discussed above, a possible explanation of the observed phenomena is the following. Basically, in any model of word learning presupposing a separation between orthographic, phonological and semantic components (compare, for instance, Seidenberg and McClelland, 1989; Grossberg and Stone, 1986) there would be two pathways for activation to flow to explain the similarity priming of pseudowords. Either the pseudoword orthographically primes the corresponding English word (e.g. 'apple' for 'arple'), which then activates its meaning, leading to the priming of the categorization of *fruit*. Or the pseudoword directly partially activates the meaning of the corresponding English word, due to their similarities. The results of our experiments point in favor of the second possibility. If priming happened on the orthographical level, it would not be reduced when the pseudoword acquires meaning.

Our model offers an even more intricate explanation. The basic component for both sides of the model (semantic and orthographic) is a collection of categorization layers, operating with a variant of competitive learning and adaptive resonance theory (Grossberg, 1987). Heavy focus is put on inhibition by competition. Dependent on context, each stimulus leads to the activation of a small number of categories (compare, for instance the interactive activation effects reported by Rumelhart and McClelland, 1981). Therefore, given orthographical input, the units compete for the activation of a word form category. If the context is such that an English word is expected, 'arple' is likely to activate the 'apple' category, If, however, there is reason to believe that 'arple' is a separate word, a new category unit is recruited, creating a new word form category (a new lexical entry, so to speak). Competition in our model is done in a "soft" way, exploiting interactive activation and inhibition that is adaptive during learning. A category will activate more than the winning unit, especially in the beginning. In other words, category units corresponding to word forms that are orthographically similar (including 'apple') will be activated. Links between word forms and (for instance) visual categories are realized through a separate layer of units. Weights to a link unit grow via repeated and consistent presentations of word-meaning pairs. When the activation of that link unit exceeds a threshold, winner-take-all is triggered, erasing all similarities to other categories (a special learning rule further assures that only weights emanating from winning units can grow – see Dorffner, 1996). Before that, activation is spread in a regular associative manner, permitting similarities on the word-form level to influence associations on the meaning level (although much weaker).

In other words, through repeated pairings of word form and meaning, the model learns to implement an arbitrary link, reflecting the symbolic nature of a word. Before sufficient learning, however, word forms can preserve some iconic function by activating meanings based on similarities to other words. If a pseudoword is encountered it will first co-activate category units corresponding to similar word forms, thus leading to the observed priming. If that pseudoword becomes more familiar, competition among word forms (the new lexical entry 'arple' and the word 'apple') will increase, thus leading to decreased priming as observed in experiment 2 (table 3). If the pseudoword acquires meaning, there is competition on both the word form level and the meaning level, largely erasing the priming due to word form similarity.

Conclusion

We demonstrated that English speakers, when faced with pseudowords, tend to have associations based on similarities between the pseudowords and similar English words. This was shown in a priming experiment where pseudowords primed a categorization task involving the semantic categories of orthographically similar words. When subsequently subjects learned meanings for the pseudowords, the priming disappeared. This iconic–symbolic duality in language stimuli may be one of the properties of language which allow it to be used with maximum efficiency for both novel and routinized communicative tasks. We argue that connectionist models which capture both the iconic and symbolic aspects of language are to be preferred over those which emphasize only similarity-based processing.

Acknowledgements

Part of this research was performed when the fist author was on leave as a visiting scholar at the Psychology Department, Boston University. We thank the Center for Adaptive Systems at Boston University for providing office space during that time. The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Science and Transport.

References

Cohen J.D., MacWhinney B., Flatt M., Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers*, 25(2), 257–271.

Dagenbach D., Horst S., Carr T.H. (1990). Adding new information to semantic memory: how much learning is enough to produce automatic priming? *J. Exp. Psych.: Learning, Memory, and Cognition*, 16(4), 581–591.

Dorffner G. (1996). Categorization in early language acquisition - accounts from a connectionist model, *Language and Cognitive Processes*, to appear.

Dorffner G., Hentze M., Thurner G. (1996). A Connectionist Model of Categorization and Grounded Word Learning, in Koster C., Wijnen F. (eds.): Proceedings of the Groningen Assembly on Language Acquisition (GALA '95).

Feustel T.C., Shiffrin R.M., Salasoo A. (1983). Episodic and lexical contributions to the repetition effect in word identification, *J Exp. Psych.: General*, 112, 309–346.

Grossberg S. (1987). Competitive Learning: From Interactive Activation to Adaptive Resonance, *Cognitive Science*, 11(1), pp.23–64, 1987.

Grossberg S., Stone G. (1986). Neural dynamics of word recognition and recall: attentional priming, learning, and resonance, Psychological Review, 93(1), 46–74.

Irwin D.I., Lupker S.J. (1983). Semantic priming of pictures and words: a levels of processing approach, *J. of Verbal Learning and Verbal Behavior*, 22, 45–60.

Rheingold H. (1988). *They have a word for it: a light-hearted lexicon of untranslatable words & phrases*, Los Angeles: J.P. Tarcher.

Rueckl J.G. (1990). Similarity effects in word and pseudoword repetition priming, *J. Exp. Psych.: Learning, Memory, and Cognition*, 16(3), 374–391.

Rueckl J.G., Olds E.M. (1993). When pseudowords acquire meaning: effect of semantic associations on pseudoword repetition priming, *J. Exp. Psych.: Learning, Memory, and Cognition*, 19(3), 515–527.

McClelland J.L., Rumelhart D.E. (1981). An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings, *Psychological Review*, 88, 375–407.

Rumelhart D.E., Hinton G.E., Williams R.J. (1986). Learning Internal Representations by Error Propagation, in Rumelhart D.E. & McClelland J.L., Parallel Distributed Processing, Explorations in the Microstructure of Cognition, Vol 1: Foundations, MIT Press, Cambridge, MA.

Sebeok T.A. (1994). Signs, University of Toronto Press.

Seidenberg M.S., McClelland J.L. (1989). A distributed, developmental model of visual word recognition, *Psychological Review*, 96, 523–568.

Whittlesea B., Cantwell A. (1987). Enduring influence of the purpose of experiences: Encoding–retrieval interactions in word and pseudoword identification, *Memory and Cognition*, 15, 465-472.

Appendix

Examples of Pseudowords (listed with their corresponding English original and category) used in the priming phases of experiments 1 and 2:

arple (apple - fruit); chaim (chair - furniture); drest (dress - clothing); horfe (horse - animal)

Examples of Concepts used as meanings in phase 2 of experiment 1:

Synonyms: *Excited, Liar, Mistake*

Known concepts:

- *Unusual appetites during pregnancy*
- *Applying something cold to one's skin to cool off*

Unknown concepts:

- *Little flaws that create an elegant whole*
- *Items one is likely to forget when making up a budget*