The Relative Contributions of Taxonomic and Thematic Knowledge When Making Similarity Judgements

Niels Dickson, Christopher R. Cox

Louisiana State University

Abstract

Taxonomic knowledge is comprised of the facts we know about individual objects and what allows us to put objects in categories; thematic knowledge is the information we acquire about objects through its co-occurrence with other objects in different linguistic and environmental contexts. Can both kinds of knowledge be represented within a single system, as argued by the "hub-and-spokes model" of semantic storage in the brain, or do we require a "dual-hub" account in which each is represented by a separate system? Much research investigating the distinction has been devoted to discovering which knowledge store requires more effort to access. However, this has led to mixed results, with the amount of effort tracking task demands rather than knowledge type. The current study investigated if and how participants can jointly utilize thematic or taxonomic knowledge when making similarity judgements where both kinds of knowledge may be relevant. We utilized normed featural analysis data to determine the taxonomic similarity between two words and the Word2vec program to determine the co-occurrence rate in language between two words to indicate thematic similarity. We found a slight taxonomic preference for these similarity judgements but failed to predict behavior when dividing our trials into conditions.

Keywords: Semantic Cognition, Taxonomic, Thematic

The Relative Contribution of Taxonomic and Thematic Knowledge When Making Similarity Judgements

People develop an incredibly large repertoire of concepts over their lifetime. Semantic cognition refers to the cognitive mechanisms needed to actively manipulate these concepts in a meaningful way, giving rise to not only our basic understanding of language but also our ability to perform basic tasks that require knowledge about the objects in our environment. As it is typically understood, semantic knowledge is comprised of two kinds: taxonomic and thematic. Taxonomic knowledge refers to what we know about individual items, and how they can be arranged into categories based on similar perceptual features or functions. Thematic knowledge refers to what we know about how items and concepts associate with one another within environmental or linguistic contexts. To illustrate, "dog" and "wolf" are taxonomically related because dogs and wolves share many behaviors and physical attributes, while "dog" and "leash" are thematically associated because they occupy similar contexts. Conversely, dogs and wolves are not thematically related, nor are dogs and leashes taxonomically related. Research has produced competing ideas on the distinction between the two, and the current research will investigate how humans interact with the distinct kinds of knowledge. The distinction between thematic and taxonomic knowledge is currently at the heart of a major debate in cognitive neuroscience regarding whether semantic knowledge is represented in a single neural system or divided over multiple. The debate has focused around two models: the hub-and-spoke model which predicts a single system and the dual hub model which predicts two.

The dual hub account has been proposed in response to research that highlights key differences between the two forms of representation. Recently, Mirman, Landrigan, and Britt (2017) compiled relevant neurological, behavioral, and computational research to indicate the need for two distinct hubs that receive information for taxonomic and thematic representations separately. Mirman et al. (2017) make two arguments for this dissociation: topographic specialization and architectural specialization. The topographic argument hinges on the observation that taxonomic knowledge relies on the spatial resolution of the ventral visual stream to attend to the details of individual objects while thematic knowledge require the temporal resolution of the dorsal visual stream to track the co-occurrence of multiple objects. Because the ventral and dorsal streams are anatomically distant from one another, and long-range neural connections are metabolically costly and rare (Betzel & Basset, 2018; Horvát et al., 2016), the different kinds of knowledge might be handled in topologically distinct hubs that are close to where the relevant structure is encoded. The architectural argument, on the other hand, is that specialized taxonomic and thematic systems may be necessary to accommodate the different computational mechanisms required for extracting each kind of information from the environment. Mirman et al. (2017) argue that identifying and predicting are fundamentally different cognitive processes that are essential to taxonomic and thematic knowledge respectively. They cite evidence that the anterior temporal lobe plays an important role in identification and therefore taxonomic knowledge, and the temporoparietal cortex plays an important role in predicting and representing temporal and contextual information and therefore thematic knowledge.

Although it seems natural to ascribe taxonomic and thematic knowledge to separate systems, the dominant theory of semantic cognition in the literature today—the "hub-and-spokes model"—hypothesizes that there is a single semantic system (Rogers et al., 2004). This theory has its foundation in research with patients who suffer from semantic dementia. Semantic dementia is a condition that results in a general deterioration of semantic knowledge as a consequence of neurodegeneration that begins by consuming the anterior temporal lobes while initially sparing other parts of the brain (Hodges & Patterson, 2007). Research of this condition enabled Rogers et al. (2004) to develop a computational model of semantic representation that relies on a centralized hub to mediate between various spokes distributed across the brain that encode information in modality-specific cortices.

Proponents of the single-hub account maintain that one system can accommodate both taxonomic and thematic forms of semantic representation. Hoffman, McClelland, and Lambon Ralph (2018) recently attempted to reaffirm the single-hub model and the Controlled Semantic Cognition framework with an updated computational model capable of learning about abstract concepts (i.e., those lacking perceptual or functional features) and thematic associations in addition to the features associated with concrete objects. In so doing, they demonstrate the plausibility of a single-hub approach considering the dual hub arguments outlined above.

With this fundamental division in theories of semantic cognition, an understanding of the distinction between taxonomic and thematic representation is important to uncover some of the underpinnings of semantic cognition. One way of studying the differences between the systems that is popular among researchers investigating this topic and is relevant to the current proposed study is measuring the amount of effort devoted to making semantic judgements. Recent research advancing the controlled semantic cognition framework has suggested that semantic cognition relies on executive control process that mediate between various areas in the brain that are responsible for semantic representation (Lambon Ralph, Jeffries, Patterson, & Rogers, 2017). This control manipulates activation to give rise to behaviors that rely on contextual information (Lambon Ralph et al., 2017). The executive control characteristics related to taxonomic and thematic representation play a vital role in understanding how we interact with our semantic representation.

In a recent study, Thompson et al. (2017) attempted to work within the controlled semantic cognition framework to suggest that the retrieval of thematic associations is heavily linked to semantic control. The researchers worked with semantic aphasia (SA) patients to perform semantic tasks of identity matching and thematic matching. As described above, semantic dementia results in a generalized semantic deficit that affects both taxonomic and thematic knowledge. Semantic aphasia, on the other hand, is associated with deficits in semantic control not the actual knowledge. The patients with SA were asked to choose which of three words best matched a picture. On thematic trials, the strength of the association between image and correct word was manipulated. On taxonomic trials, whether the correct word was an identity label or a super-ordinate category label for the picture was manipulated. In this study, Thompson et al. (2017) used psycholinguistic databases to match the words on imageability and familiarity in their list while supplementing it with participant ratings. The researchers used latent semantic analyses to obtain association of the words through co-occurrence measurements in the language.

Based on measurements of response time and accuracy, Thompson et al. (2017) found that the thematic association task was more difficult than the identification task for the participants, and the patients with SA showed greater signs of impairment in the thematic tasks than the identity tasks when compared to the control group. This result indicates that thematic judgements require more executive control as patients with SA performed worse on these tasks.

Thompson et al. (2017) expanded upon these results to test if they could induce a similar pattern of results in healthy participants by dividing attention and exceeding the person's ability to perform tasks relying on cognitive control. Researchers gave healthy participants the same simple task that they administered in the first experiment, but the healthy participants were given a time constraint to respond and an additional counting task during the semantic tasks that varied in difficulty (easy or difficult). This allowed the researchers to investigate important features of thematic judgments in neurotypical participants because the counting task divided the participants' attention which caused significantly greater disruption for the thematic task than the identity task (Thompson et al., 2017). In particular, the response times to the low association thematic task condition were significantly longer than any other condition which mirrored the pattern of results for the patients with SA. The results indicate that recollection of thematic associations is more reliant on cognitive control than identity or category membership taxonomic tasks.

While identity and category membership tasks do rely on taxonomic knowledge, they are not representative of the full range of tasks that rely on taxonomic knowledge. Category inference requires a participant to attempt to generate a category label that contains two or more exemplars. In contrast to the category membership task, where the category label is provided and an exemplar is evaluated with respect to whether it belongs to that category or not, category inference is expected to be a more challenging task. Category inference is arguably more like what participants do when evaluating if two exemplars are thematically associated: they may complete the task by inferring a theme or context that contains the two items. When comparing category and thematic inference tasks, results in the literature are more variable and there is not a clear advantage for taxonomic over thematic knowledge as seen in Thompson et al. (2017).

For example, Savic, Savic, and Kovic (2017) used event-related potential data and found that category inference taxonomic trials were associated with a larger N400 component than thematic trials. N400 components are often used as a marker of semantic incongruity. One would expect large N400 responses when a person encounters an unexpected combination of meanings, so Savic et al. (2017) used this component as another measure of elevated semantic processing in their study. On each trial, participants saw a picture and then determined if the following word matched the picture. Mismatch trials would involve either a thematic mismatch, a taxonomic mismatch, or an unrelated word. Taxonomically and thematically related word pairs used for the stimuli were identified based on ratings by groups of students who did not also participate in the cognitive neuroscience experiment. Raters evaluated pairs of words on a 7-point scale based on the perceived taxonomic and thematic similarity between the words. The researchers found that the largest N400 components occurred on trials with weak thematic association and effectively no taxonomic relationship. They interpret the ERP evidence to indicate that thematic processing is easier than taxonomic processing in the initial processing phase.

In a recent study, Geller, Landrigan, and Mirman (2019) assessed the cognitive effort associated with taxonomic and thematic similarity judgments using pupillometry, exploiting the fact that pupil dilation correlates with cognitive effort. In their study, the participants were presented with pairs of words and asked to decide if they were related or unrelated. As in Savic et al. (2017), pairs of related words have either a taxonomic or thematic relationship; thematically related pairs were chosen to have a very weak taxonomic relationship and vice versa. This manipulation was possible thanks to a prior study in which the taxonomic and thematic relatedness of each pair of words was rated on a 7-point scale by an independent sample, similar to Savic et al (2017). Surprisingly, Geller et al. (2019) reported that taxonomic judgments required more cognitive effort than thematic judgments as seen in longer reaction times and a steeper pupil dilation slope indicating increased pupil activity throughout the duration of taxonomic judgments. The researchers interpret their results in the context of evidence in the literature showing opposite effects by suggesting that neither taxonomic nor thematic judgments are fundamentally easier. Rather, they conclude that the amount of cognitive effort required is task dependent.

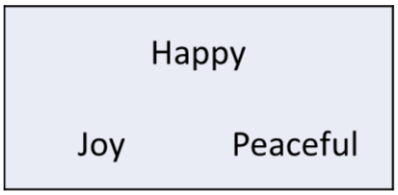
With the contradictory results produced by various studies attempting to determine the effort required to perform taxonomic and thematic judgements, it is reasonable to conclude that the amount of effort required to make these judgements depends on the specific task, and therefore does not reveal much on the nature of the distinction between taxonomic and thematic systems of representation. Instead, the current study, inspired by the methods of Lawson, Chang, and Willis (2017), aimed to determine which system people rely on more when making similarity judgements. Lawson et al. (2017) conducted a free-sorting study, where participants were given a stack of cards with either words or pictures printed on them, and participants were instructed to sort them into categories in whatever way they felt most natural. In so doing, they aimed to uncover whether semantic cognition was driven primarily by taxonomic or thematic structure.

Lawson et al. (2017) noticed that prior free-sorting studies drew conclusions about the nature of our stored semantic knowledge based on stimulus sets that were constructed to contain primarily taxonomic or thematic structure. This limits the scope of what prior work can say about the natural order of semantic knowledge in the brain. Instead, Lawson et al. (2017) administered a free-sorting task of different sets of concrete objects designed to have both kinds of structure present. In several experiments, the researchers found that the participants made significantly more thematic groupings from the free-sorting task than taxonomic. They determined the taxonomic or thematic characteristic by having independent participants rate how taxonomically similar the groups were and how thematically similar the groups were after instructions on the distinction between taxonomic similarity and thematic similarity. They do, however, stress the difficulty in determining the relation between objects, and the distinction between taxonomic and thematic associations is not always clear. Despite this difficulty, Lawson et al. (2017) concluded that because of the dominant preference for thematic sorting, our semantic knowledge prefers thematic representation over taxonomic representation.

In the current study, we used a combination of semantic feature norms and word co-occurrence data to estimate the taxonomic and thematic similarity among a set of words used in prior experiments. Previous studies reviewed here have used independent raters to determine taxonomic and thematic relations, but we hypothesized that normed data will be more representative of the actual distinction. To acquire these data tracking the theoretical distinction between taxonomic and thematic systems better than the ratings based on the human conception of the distinction, we used Google's Word2vec to obtain the word co-occurrence analysis of two words, and we used previous normed data to obtain featural analysis data. These data allowed us to construct a feature space that can determine the similarity of two words based on shared features. In the main experiment of this study (trial illustrated in Figure 1), participants performed a simple similarity judgement task, labeled the triplet task, where they decided which of the two items presented is more similar to a target item (in the form which is more similar to A: B or C). Using the data from the Word2vec program and the featural space, we attempted to determine if the participants are relying on predominately taxonomic or thematic storage to make their judgements or if they are relying on a combination of the two systems. We also investigated if there are individual differences that arise from these judgements as individuals may rely on a specific judgement strategy to determine the similarity. These questions of the preference for taxonomic or thematic judgements are important to our understanding of the representation of these two systems. Specifically, we predicted to see a combination of the two systems to make the judgments which would suggest that the two systems are highly interactive.

Figure 1

Example of Similarity Judgement Trial



Note. The Target word is “Happy.” Participants will select either “Joy” or “Peaceful” depending on which concept they believe to be more like “Happy.”

Method

Participants

Participants were recruited from the Louisiana State University student population using the SONA research participation system. 116 participants completed the experiment. Eight of the participants completed it on the computers in Dr. Cox’s lab space in the basement of Audubon Hall at LSU, while the remaining 108 participants signed up remotely and received course credit through SONA. All participants completed the study on the survey platform Qualtrics.

Of the participants in the study, 78% reported being female, 76% reported being white, 94% reported being non-Hispanic, and they had a mean age of 19.42 (SD = 1.37).

Materials and Design

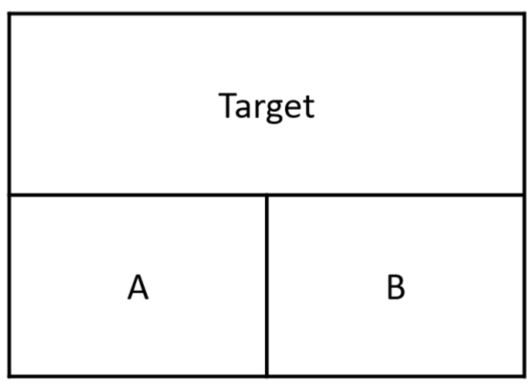
Using the 112 words with normed featural analysis data, we also obtained the word co-occurrence data for these 112 words using the word2vec program applied to the Google News corpus. We constructed our triplets with these 112 words and measured the taxonomic or thematic similarity between each member of this triplet by the cosine similarity between the featural analysis vectors and the word2vec vectors respectively. Cosine Similarity is a measure of similarity useful to compare vectors in a space and is popular in natural language processing work. We then had the taxonomic and thematic distance of reference word A from the target word and the distances of reference word B from the target word. For the purposes of the triplet task, the *relative* distances are a more important driver of behavior. When participants are making the similarity judgements, they are deciding a certain reference word is more similar to the target word relative to the second, unselected reference word. To obtain the relative distances as desired, we added the taxonomic distance of reference word A to the target word and the taxonomic distance of reference word B to the target word. Then, we divided each distance by the previous sum to obtain the relative taxonomic distance of the particular triplet. For example, consider a situation where there is a triplet where both reference words are taxonomically similar to the target word. The taxonomic distance from reference word A to the target word is 0.1 as found by the cosine similarity between the two vectors in the featural analysis. The taxonomic distance from reference word B to the target word is 0.3. The relative taxonomic distances for this triplet would then be the following: relative tax distance = 0.1/ (0.1 + 0.3) = 0.25. In this example, the relative taxonomic distances indicate that we would expect someone relying on taxonomic knowledge to select A for this triplet. We then calculated the relative thematic distances in the same way. This relative distance basically plots a point somewhere between option A and option B depending on what the data suggest where option A is represented by 0 and option B is represented by 1. So, in the previous example, 0.25 is indicating option A is more taxonomically similar to the cue, but if the relative distance were 0.90, we would conclude the opposite. The relative distance allows us to assess not only which is the predicted response but also the strength of the data suggesting one option over the other.

After having computed the appropriate taxonomic and thematic similarity data for the triplets, we then defined five conditions to sort the triplets and find the triplets that would be most interesting to the current research. We conceptualized these conditions based on the fact that a small relative distance would push participants to decide that the two words are similar, large relative distances would push participants to decide that the two words are dissimilar, and a relative distance of 0.5 would give the participants relatively little information. The conditions, illustrated in Figure 2, were as follows: 1. taxonomic and thematic knowledge pointing to the same answer, 2. taxonomic and thematic knowledge pointing to different answers, 3. taxonomic knowledge pointing to an answer while thematic knowledge does not give much information, 4. thematic knowledge pointing to an answer while taxonomic knowledge does not give much information, and 5. both knowledge stores are not giving much information.

Figure 2

Schematic Illustrating the Breakdown of Conditions

Condition	Decision based on ...	
	Taxonomic similarity	Thematic similarity
Same	A	A
Different	A	B
Taxonomic	A	—
Thematic	—	A
No Info	—	—



Note. The left indicates the responses to the triplet task that the taxonomic and thematic data predicts for each condition. We were not interested in the actual position of the option word, so “Same” condition could also be “B – B”, “Different” condition could also be “B – A”, etc.

We separated the triplets into the conditions by the percentage of the data that was most representative of that condition. For condition 1, we found the points in the relative taxonomic and thematic distances that marked the bottom and top 2.5% of the data. This percentage demarcation allowed us to find the triplets where the data pointed strongly in the same direction. The bottom 2.5% of the relative distances indicates the group of triplets with the highest target to option 1 similarity as low relative distances indicate that the data predicts option 1 being more similar to the target word than option 2. The top 2.5% of the relative distances indicates the group of triplets with the highest target to option 2 similarity as high relative distances indicate that the data predicts option 2 being more similar to the target word than option 1. This procedure found the triplets where both the taxonomic and thematic data most strongly suggested either option 1 or option 2. Using a similar percentage demarcation, we found the triplets from our list that was most representative of the condition.

Procedures

Feature norming task

For the featural analysis data, we used previous normed data from McRae et al. (2005) and Dilkina and Lambon Ralph (2012). In these studies, participants saw a word on the screen, and then they checked off all the features from a list of features provided that apply to the certain object the word is describing.

Associative norming with Google Word2Vec

Google Word2Vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) is the name of an algorithm for text analysis developed by google researchers. Informally, it often refers to pre-computed model solutions obtained by applying the Word2Vec algorithm to large text corpora and shared online of research purposes. We utilized Word2Vec solutions based on the text in

Google News, a 100-billion-word corpus, which is successful at capturing some important aspect of semantic structure for a very large set of English words (Pereira, Gershman, Ritter, & Botvinick, 2016). With the understanding that Google News does not represent all of language use, we chose to use this corpus because of its availability and its volume as we are able to encounter more instances of our stimuli words in a larger corpus to gather more accurate information. In the solution, each of several words is represented by a 300-element vector---an “embedding” of the word in a metric space---where the distance between vectors corresponds to the contextual similarity between the two words in the Google News corpus.

The utility of text-derived word embeddings is that, in contrast to the feature norms, they have no direct insight into the features of individual items. The embeddings are derived exclusively from text co-occurrence structure. There is a large literature interested in how much of semantic knowledge can be extracted from linguistic co-occurrence statistics. The important caveat from this literature is that quite a lot of structure, including structure that appears rather taxonomic, can be extracted from these statistics. This is why Mirman et al. (2017) warn that such text-based norms do not properly isolate thematic knowledge.

This is an interesting discussion– if taxonomic structure can be acquired through associative mechanisms, does that make it “thematic” or “taxonomic”? Should we sort based on content or by means of acquisition? With these questions in mind, we have made a practical decision that a text-based associative database such as that provided by the Word2Vec model is a reasonable proxy for thematic knowledge. It is also consistent with the latent semantic analysis approach adopted by Thompson et al. (2017).

Similarity judgement task

For the experimental task, participants were presented with a series of trials containing three word "triplets" arranged in a triangular formation: a target word presented at the top, center of the screen, and two reference words presented under the target word to the left and the right. Option A and option B randomly appeared on the screen either on the right or the left to ensure there was not a bias simply from the position on the screen. Participants were given instructions to simply choose which of the two words presented is more similar to the target word and encouraged to use whatever they know about and whatever experiences they have had with the referenced concepts to inform that decision. We phrased the instructions with the intent to prime the participants to consider both their taxonomic and thematic knowledge while completing the experiment. We also carefully constructed the instructions to deter the participants from basing their decisions off orthographic information. Orthographic information can drive behavior in many psycholinguistic studies, so we included a couple aspects to the instructions to ensure the participants were not comparing the structure or sounds of the words themselves. First, we avoided verbiage that might suggest that the task involves comparing words. We wanted to prime the participants to access the concepts that the words refer to, so we included the following sentence in our instructions: "The task is to pick the word that goes best with the target word based on what you know about the things that the words refer to." Second, we included an example that illustrated an inappropriate decision for our experiment task that is based on orthographic similarity. The example triplet was 'cat,' 'cap,' and 'dog' with 'cat' as the target word and 'cap' and 'dog' as the two reference words. We then instructed the following: "'Cap' and 'cat' share many letters and sounds, but they refer to things that have little to do with one another. We would expect one to choose 'dog' as being more similar to 'cat' than 'cap'."

After responding to all the triplets, the participants also explained a two of their decisions in a follow-up section. We designed the experiment to show every participant the same two triplets with a reminder of which option they chose. We wanted to ask these follow-up questions to gauge if the participants were consistently making associations based on taxonomic information or thematic information. The follow-up questions allowed us to understand some of the reasoning behind the decisions that the data output would not allow. The instructions, the examples, and the follow-up questions can be seen in Appendix B.

Removing an outlier, the participants completed the experiment between 4.5 minutes and 77.5 minutes with a mean of 14.5 minutes. The 1st quartile of the duration in minutes is 9.71, and the 3rd quartile of the duration in minutes is 16.08.

Results

In our initial data analysis, we were interested in what proportion of the responses aligned with what our taxonomic and thematic metrics predicted. If the taxonomic distance (as determined by the feature vectors of the norming task) of option one from the triplet target word is less than the taxonomic distance of option two from the target word, then decisions relying solely on taxonomic data would produce option one because the smaller distance indicates more similarity. If the second option has a smaller taxonomic distance, then the taxonomic data predict the participant to choose the second option. Using this framework, we calculated the proportion of responses that aligned with what the taxonomic data predicts for each condition. Using the same logic, we found what proportion of the responses aligned with what we would expect if the participant relied on thematic information. Here, we found the thematic distances based on the word co-occurrence vectors gathered from applying the word2vec algorithm.

Through this analysis, we found that the taxonomic and thematic distances that we collected were not strongly predictive of human behavior in this similarity judgement task. These data are summarized below in Table 1 where the "Tax Response" and "Them Response" columns indicate the percentage of responses that aligned with the taxonomic data or thematic data respectively. For example, if the participant chose option A when the taxonomic distance of A to the cue is 0.1 and the taxonomic distance of option B to the cue is 0.5, then this would be a response in line with the taxonomic data. The last column, "Lev Response," refers to the percentage of responses that aligned with Levenshtein Distance data. Levenshtein distance is a way of measuring the similarity of two strings of characters. It is commonly used to measure the similarity of two words by finding the minimum number of edits required to change the one to the other. We used this measurement to compare the taxonomic and thematic data results to the percentage of responses that aligned with simply orthographic decisions of similarity based on the words themselves. Looking at all responses from the 116 participants of the 150 triplets, we found that when the taxonomic data suggested the participant should choose a reference word, they chose that option 61.08% of the time. We also found that when the thematic data suggested the participant should choose a reference word, they chose that option 54.95% of the time. For each of the 150 triplets (30 per condition), we computed the proportion of participants that chose each option. An option might be the taxonomic choice, the thematic choice, align with both sources of similarity, or neither source of similarity. In Table 1, we report the proportion of triplets that align with the taxonomic response or the thematic response, respectively. Note that the proportions in a row can sum to more than 1 because sometimes an option is aligned with both sources of similarity. Statistical significance is determined as a 1-sample t-test against 0.5, a value that would indicate indifference between the taxonomic and thematic options. Tests within conditions have 29 degrees of freedom. Tests over all conditions have 149 degrees of freedom.

Table 1

The percentage of the Triplet Responses Aligning with the Normed Data

Condition	Tax Response (%)	Them Response (%)	Lev Response (%)
Same	55.10	54.81	57.75*
Different	65.64**	51.51	56.92
Tax	65.04**	49.46	54.53
Them	60.31*	45.64	39.80*
No	59.32	73.30**	46.24
Across All Conditions	61.08**	54.95**	51.05

Note. “Tax Response” refers to the percentage of the responses in the condition that align with the taxonomic data. If the relative taxonomic distance for a triplet is less than 0.5, then a response that is aligned with the data would be option 1. The same logic is used for the other two columns.

*p < .05

**p < .01

Relating these data back to our research question, one could look at the results of one of the condition relevant to our research question, the “different” condition where the taxonomic and thematic information are pointing in different directions, and conclude that humans tend to rely on taxonomic knowledge more than thematic knowledge. But because the other conditions produced results where the data is not predicting human behavior, we cannot make this conclusion. For example, the “same” condition, where both types of knowledge are prompting participants to choose the same option, was supposed to be the group of the easiest triplet decisions. We hypothesized that if participants could rely on both types of knowledge when making a decision for a triplet, then this condition would present the triplets with the most obvious decisions. But participants only selected the answer that the data predicts on a little over half of their responses and these proportions were not significantly greater than chance. We can also see that our data seem to be failing to predict behavior in the "them' condition where the thematic information should dominate the decisions, but participants relied more on taxonomic data in this condition. We do see a trend to rely on taxonomic associations more than thematic associations in our triplet task, but because we failed to see any sensible results from our condition grouping, more investigation is required.

We were also interested in seeing how often the participants agreed with one another. The initial data analysis told us how often the responses aligned with what the taxonomic and thematic data would predict, but it did not tell us if participants generally agreed with each other for each triplet. This question is particularly interesting for the current project because similar responses across participants would indicate that there is a knowledge representation shared among participants that our taxonomic and thematic data are not adequately capturing.

To find out how often participants agreed with one another, we found which response was the most common for each triplet, and then we found the proportion of responses that matched the most popular response. As seen in Table 2, the proportions of agreement for the 150 triplets ranged from 0.504 to 0.915. One can see that most of the time the participants had a fair level of agreement and responses were not simply random as the mean of the proportions of agreement was 0.704. But when dividing the proportion of agreement by the different conditions as seen in Figure 3, we found that the No Info condition had the highest level of agreement, 0.744, and the Same condition had the lowest level of agreement, 0.666. These results further demonstrate that our taxonomic and thematic data are not capturing essential elements of human representation because we expected that the Same condition would be easiest for the participants and have the highest level of agreement and that the No Info condition would be difficult for participants and would result in fairly random responses.

Table 2

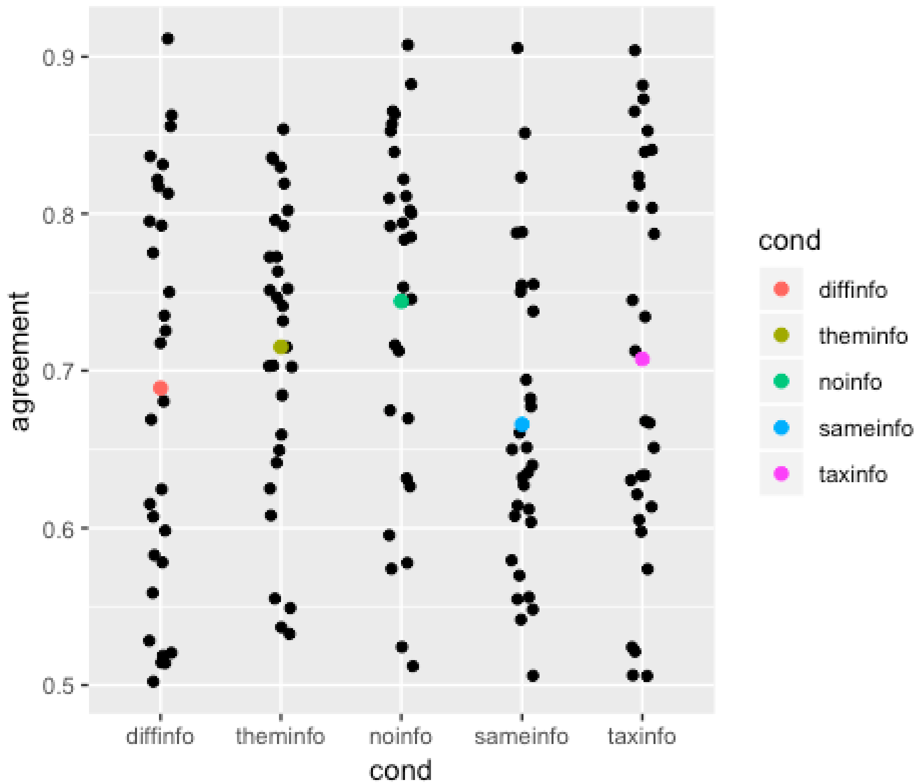
Percentage of Agreement for all 150 Triplets

Minimum (%)	Mean (%)	Median (%)	Maximum (%)
50.43	70.43	71.37	91.45

Note. Agreement was measured for each triplet by finding which response was most popular for the triplet, then finding the percentage of responses that was the popular response.

Figure 3

Graph plotting the agreement for each Triplet Grouped by Condition



Note. Each black dot represents the agreement score for one triplet, and the colored dots represent the mean for the condition.

Using these majority data, we also found the proportion of the triplets where the most popular response was consistent with the response that the taxonomic and thematic data would predict. We calculated the total proportion of taxonomically consistent responses of the 150 triplets which is summarized in Table 3. We found that 0.613 of the majority responses for each of the 150 triplets were consistent with the thematic data based on the majority decision. We also found that 0.680 of the majority responses for each of the 150 triplets were consistent with the taxonomic data based on the majority decision. Because this analysis compared each triplet's majority response against the normed data, this process does not warrant a one sample t test as performed for Table 1. Instead, we performed a z-test conventional for a binary measure, where we compared the proportion against chance.

Table 3

Percentage of the Triplets Where the Majority Response Aligned with the Normed Data

Condition	Tax Agreement (%)	Them Agreement (%)
Same	60.00	60.00
Different	80.00**	53.33
Tax	66.67	56.67
Them	66.67	40.00
No	66.67	96.67**
Across All Conditions	68.00**	61.33**

Note. We found the response that aligned with taxonomic data and the response that aligned with thematic data for each triplet. Then we determined if the majority response for each triplet was the same as the aligned responses.

*$p < .05$

**$p < .01$

Discussion

In this research project, we set out to answer how humans represent taxonomic and thematic knowledge. In reviewing the current literature, we found that many researchers were investigating the amount of effort required to access the two knowledge stores. These studies produced mixed results that suggested that accessing taxonomic knowledge may be more difficult for some tasks while accessing thematic knowledge may be more difficult for other tasks. We hypothesized that constructing trials that were either accessing taxonomic or thematic knowledge may be an imperfect system for understanding how knowledge is represented. These trials from previous studies are artificially constructed with data from participants who are rating the relationships of words after given the theoretical definitions of taxonomic and thematic knowledge. This process creates almost a circular argument of investigating a distinction between the two types of knowledge by imposing the distinction onto the participant raters. We wanted to apply two techniques to our procedure that would be novel for the current literature on the distinction between taxonomic and thematic knowledge and that would effectively investigate the distinction between the two. Firstly, we employed the triplet task format where participants chose which of two reference words was more similar to a target item (in the form which is more similar to A: B or C). Importantly, this procedure did not place any restrictions on the participants and enabled us to analyze how participants are freely accessing knowledge when making similarity judgements. Secondly, in order to analyze how participants access knowledge in the context of the taxonomic and thematic knowledge distinction, we divided our trials into conditions not by data from independent raters but by normed data that better reflect the cognitive processes involved in a theoretical distinction between taxonomic and thematic knowledge. In order to theoretically distinguish between taxonomic and thematic similarity without using data from independent raters, we used measurements of linguistic contextual co-occurrence, how often words appear together in language, to define thematic similarity, and we used normed featural analysis data to define taxonomic similarity. Because the motivations behind the normed data are a major component of this investigation, we will now summarize these procedures.

We used the word2vec algorithm to get measurements for thematic similarity as the word embeddings that this algorithm produces after being trained on a multi-billion-word corpus perform well on a variety of natural language processing tasks (Levy & Goldberg, 2014). The foundation of this work is the assumption that words that appear in similar contexts have similar meanings, so the word2vec program hoped to construct a vector space of word embeddings that predicts nearby words (Mikolov et al., 2013). The algorithm produces a 300-dimension vector space where words that share similar contexts are represented by vectors that are close together in this vector space. We hypothesized that this linguistic co-occurrence metric would be a theoretical approach to thematic knowledge because thematic knowledge is knowledge that we acquire about a certain object through the context that we encounter the object. Even though we realized that this linguistic co-occurrence metric might also capture more than this limited definition of thematic knowledge, we expected that a linguistic context metric would mimic how humans build a knowledge store based on context.

We used the feature norming work from Dilkina and Lambon Ralph (2012) and McRae et al. (2005) as an estimate of taxonomic similarity. The feature norming task consisted of participants checking off all the features from a list that apply to a certain word. From the results of this norming task, we constructed feature vectors so that we can compare the featural similarity of two words by how close their two vectors are to one another in the vector space. By relying on the feature norming task, we are not claiming that humans represent knowledge as a list of features. But researchers do hypothesize that these data help in our understanding of representation because the participants must exploit the knowledge representations they have developed when making decisions on the presence or absence of certain features (McRae et al., 2005). In the current project, the feature norming approach was taken because it draws attention entirely to the perceptual qualities, behaviors, and functions of individual items. Associations with other items are irrelevant to the feature norming objective, and so the resulting feature vectors isolate structure thought to be relevant specifically for taxonomic knowledge.

One difficulty in investigating the distinction between taxonomic and thematic knowledge with our featural analysis and co-occurrence data is that these two metrics capture much of the same information. Theoretically, we were hoping to account for only facts about an individual object through its features (which would reflect taxonomic knowledge) but can we really say that the features applied to certain concepts are completely devoid of how the concept is perceived in particular contexts? Similarly, we hoped to target information that was strictly acquired through contextual co-occurrence with the linguistic co-occurrence data (which would reflect thematic knowledge) but couldn't we say that some things may appear in similar contexts because they have similar features? This difficulty in simply conceiving the theoretical distinction between taxonomic and thematic knowledge suggests that these two types of knowledge should be treated as a continuum as opposed to separate bins where some knowledge is stored and acquired strictly taxonomically and other knowledge is stored and acquired strictly thematically.

With this foundation of how we applied word co-occurrence and featural analysis data, we must now attempt to understand the implications of the experimental results not predicting human behavior. We found that participants often agreed with one another on which option was most similar to the target word with an average of 70.43% agreement across all 150 triplets. With this research project, we wanted to investigate if one type of knowledge would dominate the decisions on the triplet task when both types of knowledge may be relevant to the decision. To illustrate with a previous example, one may be asked to decide which of the two concepts, "wolf" or "leash," is more similar to the concept "dog." To answer this question, the person may employ taxonomic knowledge and decide "wolf" is more similar, or the person may employ thematic knowledge and decide "leash" is more similar. We attempted to construct triplets of this nature in the "different information" condition and theorized that a preference of one knowledge system indicates that the two knowledge systems cannot be represented in a single system. We found that there was a slight preference for participants to employ taxonomic knowledge in this "different" condition which indicated the prior sentiment. But we constructed our other conditions to represent all the ways that these two knowledge systems might interact. We made predictions about the responses we might see for each of the conditions. One of our predictions was that the "same" condition would be the easiest condition for the participants under the assumption that the participants could employ both types of knowledge to arrive at the answer. However, the proportion of responses that aligned with the taxonomic data and the thematic data was not significantly greater than chance. Further, we constructed the "thematic" condition to reflect the triplets where the participants cannot rely on taxonomic data to make a decision (i.e. the taxonomic similarity between the target and option 1 is similar to the taxonomic similarity between the target and option 2) and the thematic data point strongly to one of the options. We predicted that the participants would rely on thematic knowledge for this condition, but we observed that the proportion of responses that aligned with thematic data was not significantly different from chance. Because we were unable to see these predicted results in our other conditions, we cannot make a conclusion about the nature of representation based on the "different" condition.

We must now discuss why the data did not predict human behavior for our experimental task and what we can learn from this result. A potential source of the difficulty in predicting human behavior in this triplet task comes from the strong link between the word and the concept. As explained earlier, we meticulously wrote the instructions to avoid any suggestion that the task involves comparing words. Hoping to prime the participants to only be comparing the concepts that the words refer to, we even included an example that illustrated that judging that the concepts “cat” and “cap” are related because of the similar letters and sounds would be inappropriate and incorrect for the current task. Despite these precautions, in the responses for one of the follow-up questions, 48 participants of the 116 participants cited some variant of orthographic information to explain why they made their decision. In the follow-up question, they were prompted with the triplet (dress: kite or drum) as they saw it in the experiment task, and they were asked why they decided “drum” was more similar to “dress” than kite. 48 participants explained that they chose “drum” because both the words start with “dr,” both the words start with “d,” or some other explanation related to the structure of the words. This issue does raise questions about motivating participants to carefully read instructions and thoughtfully complete experiments, but the issue also points to a potential shortcoming of the task itself. When presenting participants with three words on the screen and asking them to make similarity judgements, even conscientious participants who try to strictly compare the concepts might fall into relying on orthographic similarity for the less obvious triplets. We did consider word frequency when constructing the triplets to ensure that the participants did not choose an option simply because it was a higher frequency word, but a more robust orthographic matching could eliminate the temptation to fixate on the word structures in this task. Despite the issues that this follow-up question raised, the proportion of responses that agreed with the Levenshtein data (a metric for the orthographic similarity between two words) was not significantly different from chance which suggests that although many participants relied on orthographic information for the follow-up question triplet, the participants did not seem to reliably rely on orthographic information.

Upon further investigation of our procedure, we found an error in the coding of our experiment that incorrectly arranged the triplets in Qualtrics. The words that were supposed to be the target words and the words that were supposed to be the first option were flipped when the triplets were entered into Qualtrics. This error accounts for the unpredictable results observed in the conditions as the triplets of each condition no longer reflect the data used to construct that condition. Despite this error, our results across conditions indicate that participants relied on both taxonomic knowledge and thematic knowledge as the proportion of responses that aligned with taxonomic data and the proportion of responses that aligned with thematic data were both significantly greater than chance. This result offers promising evidence that humans can utilize both taxonomic and thematic knowledge fluently when making similarity decisions which supports a single-hub knowledge representation hypothesis where a single-hub may manipulate these two types of information in appropriate situations. Because our conditions no longer reflect our initial construction, further analysis of the interaction between the two types of knowledge is limited. But we suggest that future investigation build off the research on the differences between the two types of knowledge to investigate how humans actually use these types of knowledge and how these two systems interact.

References

Betzel, R. F., Basset, D. S. (2018). Specificity and robustness of long-distance connections in weighted, interareal connectomes. *PNAS, 115(21).*

Chen, Y., Shimotake, A., Matsumoto, R., Kunieda, T., Kikuchi, T., Miyamoto, S., … Lambon Ralph, M. A. (2016). The ‘when’ and ‘where’ of semantic coding in the anterior temporal lobe: Temporal representational similarity analysis of electrocorticogram data. *Cortex, 79*, 1-13.

Dilkina, K., Lambon Ralph, M. A. (2012). Conceptual structure within and between modalities. *Frontiers in Human Neuroscience, 6.*

Geller, J., Landrigan, J., Mirman, D. (2019). A pupillometric examination of cognitive control in taxonomic and thematic semantic memory. *Journal of Cognition*, 2, 1-10.

Hodges, J. R., Patterson, K. (2007). Semantic dementia: A unique clinicopathological syndrome. *Lancet Neurology, 6,* 1004-1014.

Hodgson, C., Lambon Ralph, M. A. (2008). Mimicking aphasic semantic errors in normal speech production: Evidence from a novel experimental paradigm. *Brain and Language*, 104, 89-101.

Hoffman, P., McClelland, J. L., Lambon Ralph, M. A. (2018). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychological Review*, 125, 293-328.

Horvát, S., Gamanut, R., Ercsey-Ravasz, M., Magrou, L., Gamanut, B., Van Essen, D. C., … Kennedy, H. (2016). Spatial embedding wiring cost constrain the functional layout of the cortical network of rodents and primates. *PLoS Biology, 14.*

Jogan, M., Stocker, A. A. (2014). A new two-alternative forced choice method for the unbiased characterization of perceptual bias and discriminability. *Journal of Vision*, *14*.

Lambon Ralph, M. A., Jeffries, E., Patterson, K., Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, 18, 42-55.

Lawson, R., Chang, F., Willis, A. J. (2017). Free classification of large sets of everyday objects is more thematic than taxonomic. *Acta Psychologica*, 172, 26-40.

Levy, O., & Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. *Advances in Neural Information Processing Systems*, 27, 8–13.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J. (2013). Distributed representations of words and phrase and their compositionality. *In Proceedings of NIPS*.

Mirman, D., Landrigan, J., Britt, A. E. (2017). Taxonomic and thematic systems. *Psychological Bulletin*, *143*, 499-520.

McRae, K., Cree, G. S., Seidenberg, M. S., McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*, 547-559.

Pereira, F., Gershman, S., Ritter, S., Botvinick, M. (2016). A comparative evolution of off-the-shelf distributed semantic representations for modelling behavioural data. *Cognnitive Neuropsychology*, *33*, 175-190.

Ratcliff, R., McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks.

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. Psychological Science, 9(5), 347–356.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review, 111*, 205-235.

Savic, O., Savic, A. M., Kovic, V. (2017). Comparing the temporal dynamics of thematic and taxonomic processing using event-related potentials. *PLOS ONE*, 12.

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Brecher, A., Faseyitan, O. K., Dell, G. S., … Coslett, H. B. (2011). Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 8520-8524.

Thompson, H., Davey, J., Hoffman, P., Hallam, G., Kosinski, R., Howkins, S., … Jefferies, E. (2017). Semantic control deficits impair understanding of thematic relationships more than object identity. *Neuropsychologia, 104*, 113-125.

Appendix A

List of all the triplets used in the experiment where the majority decision is highlighted and the "Proportion of Agreement" column added to indicate what proportion of the responses agreed with the majority decision.

Target Word	Option 1	Option 2	Proportion of Agreement
camel	motorcycle	cow	0.82905983
whistle	giraffe	cannon	0.86324786
bear	dress	skirt	0.52991453
bee	shell	owl	0.72649573
kangaroo	tractor	dog	0.79487179
bee	anchor	owl	0.79487179
kangaroo	car	ostrich	0.82051282
zebra	cymbals	suitcase	0.51282051
fly	car	duck	0.73504274
bat	shell	owl	0.77777778
mouse	hammer	dog	0.85470085
web	dress	skirt	0.52136752
snail	tractor	monkey	0.66666667
snake	cymbals	drum	0.51282051
crab	rhinoceros	lion	0.58119658
mouse	church	dog	0.81196581
gorilla	wheel	monkey	0.83760684
rooster	harp	piano	0.55555556
lobster	anchor	screw	0.75213675
monkey	guitar	piano	0.52136752
seahorse	anchor	owl	0.5982906
raccoon	slide	tractor	0.60683761
mouse	anchor	dog	0.91452991
whistle	rooster	ladder	0.68376068
mouse	tractor	monkey	0.82051282
gun	pig	cow	0.5042735
ladybug	tractor	monkey	0.62393162
crab	car	duck	0.71794872
owl	flute	violin	0.61538462
giraffe	cymbals	bicycle	0.58119658

kangaroo	shell	fence	0.79487179
raccoon	truck	lion	0.75213675
kangaroo	ski	cat	0.52991453
camel	swing	lion	0.76068376
basket	gorilla	lion	0.75213675
gorilla	cymbals	clock	0.7008547
ant	sheep	cow	0.83760684
kangaroo	lemon	horse	0.8034188
zebra	lemon	elephant	0.68376068
cymbals	dog	bus	0.74358974
fly	giraffe	penguin	0.76923077
bell	whale	trumpet	0.60683761
shell	rooster	duck	0.83760684
camel	stroller	suitcase	0.73504274
rhinoceros	violin	lion	0.64102564
kangaroo	cymbals	bicycle	0.71794872
giraffe	kite	vase	0.82051282
goat	basket	lion	0.54700855
raccoon	pumpkin	squirrel	0.79487179
whistle	camel	trumpet	0.62393162
bear	bicycle	car	0.7008547
kangaroo	trumpet	violin	0.64957265
deer	harp	guitar	0.55555556
bell	raccoon	trumpet	0.82905983
penguin	violin	guitar	0.76923077
whistle	bee	crown	0.65811966
cymbals	harp	motorcycle	0.7008547
bell	fly	slide	0.85470085
whale	flute	trumpet	0.74358974
zebra	clock	lion	0.53846154
clock	wheel	apple	0.79487179
bell	cow	tomato	0.90598291
shell	clock	apple	0.78632479
frog	whale	cherry	0.8034188
frog	watermelon	piano	0.58119658
mouse	raccoon	wheel	0.86324786
lobster	pig	helicopter	0.71794872
frog	whale	screw	0.67521368
fence	accordion	cow	0.85470085

mouse	ant	windmill	0.78632479
cannon	trumpet	chicken	0.51282051
seahorse	church	hammer	0.63247863
frog	butterfly	tractor	0.75213675
bear	caterpillar	web	0.81196581
rhinoceros	screw	cherry	0.82051282
cigar	fence	motorcycle	0.66666667
slide	car	ostrich	0.74358974
web	dress	watermelon	0.57264957
goat	bat	web	0.85470085
snowman	watermelon	helicopter	0.8034188
church	crown	dog	0.88034188
ski	cannon	fence	0.70940171
rabbit	ant	lion	0.83760684
rooster	penguin	helicopter	0.5982906
frog	mouse	fence	0.52136752
anchor	iron	hammer	0.86324786
caterpillar	tractor	watermelon	0.81196581
giraffe	kangaroo	cannon	0.62393162
deer	seahorse	shell	0.79487179
fly	zebra	wheel	0.74358974
mouse	cat	chicken	0.82051282
rabbit	zebra	motorcycle	0.55555556
lobster	seahorse	pliers	0.55555556
cow	tomato	scissors	0.85470085
mouse	gorilla	clock	0.75213675
strawberry	pumpkin	sled	0.68376068
mouse	ski	lemon	0.63247863
church	hammer	elephant	0.64957265
rabbit	pig	harp	0.53846154
tortoise	seahorse	accordion	0.78632479
dress	kite	drum	0.58119658
camel	mouse	seahorse	0.69230769
camel	tiger	owl	0.61538462
lobster	spider	flute	0.63247863
gun	snowman	truck	0.61538462
bat	lobster	harp	0.90598291
goat	seahorse	drum	0.60683761
ladybug	lobster	elephant	0.5042735

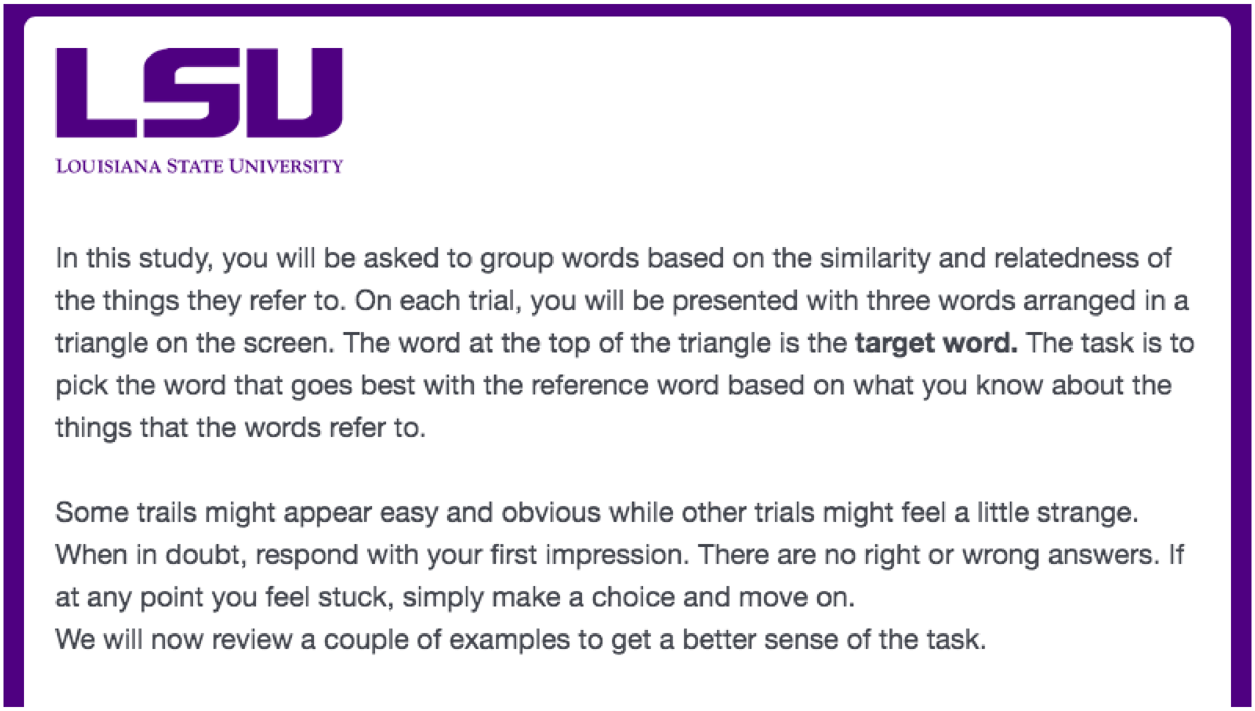
elephant	horse	apple	0.73504274
frog	caterpillar	drum	0.67521368
violin	guitar	sled	0.57264957
raccoon	apple	scissors	0.75213675
camel	lobster	pig	0.54700855
nut	swing	owl	0.62393162
zebra	stroller	cherry	0.64102564
tortoise	pig	cake	0.75213675
nut	pumpkin	cow	0.64957265
mouse	seahorse	fence	0.78632479
goat	tortoise	dress	0.65811966
goat	strawberry	screwdriver	0.60683761
zebra	windmill	monkey	0.63247863
rabbit	web	bicycle	0.78632479
stroller	elephant	lion	0.83760684
raccoon	dress	dog	0.82051282
anchor	elephant	lion	0.62393162
brush	car	hammer	0.63247863
truck	monkey	squirrel	0.5042735
cake	wheel	bicycle	0.86324786
kangaroo	brush	hammer	0.90598291
ant	clock	fence	0.64957265
basket	web	bicycle	0.70940171
rabbit	clock	fence	0.82051282
camel	anchor	wheel	0.57264957
seahorse	web	bicycle	0.5042735
butterfly	fence	monkey	0.87179487
deer	anchor	cow	0.52136752
snail	rooster	chicken	0.66666667
butterfly	windmill	monkey	0.8034188
tractor	owl	helicopter	0.88034188
ladybug	wheel	chain	0.8034188
ladybug	wheel	ladder	0.52136752
giraffe	clock	fence	0.74358974
crown	rooster	chicken	0.61538462
ant	anchor	cannon	0.83760684
ant	lemon	lion	0.66666667
gorilla	train	cow	0.60683761
fish	wheel	bicycle	0.73504274

ant	truck	lion	0.5982906
bee	web	bicycle	0.63247863
shell	elephant	lion	0.85470085

Appendix B

Figure 4

Image of the instructions that the participants saw in Qualtrics.



LSU
LOUISIANA STATE UNIVERSITY

In this study, you will be asked to group words based on the similarity and relatedness of the things they refer to. On each trial, you will be presented with three words arranged in a triangle on the screen. The word at the top of the triangle is the **target word.** The task is to pick the word that goes best with the reference word based on what you know about the things that the words refer to.

Some trails might appear easy and obvious while other trials might feel a little strange. When in doubt, respond with your first impression. There are no right or wrong answers. If at any point you feel stuck, simply make a choice and move on.
We will now review a couple of examples to get a better sense of the task.

Figure 5

Image of the first example presented to the participants to illustrate that one can use different types of information to decide two concepts are similar.

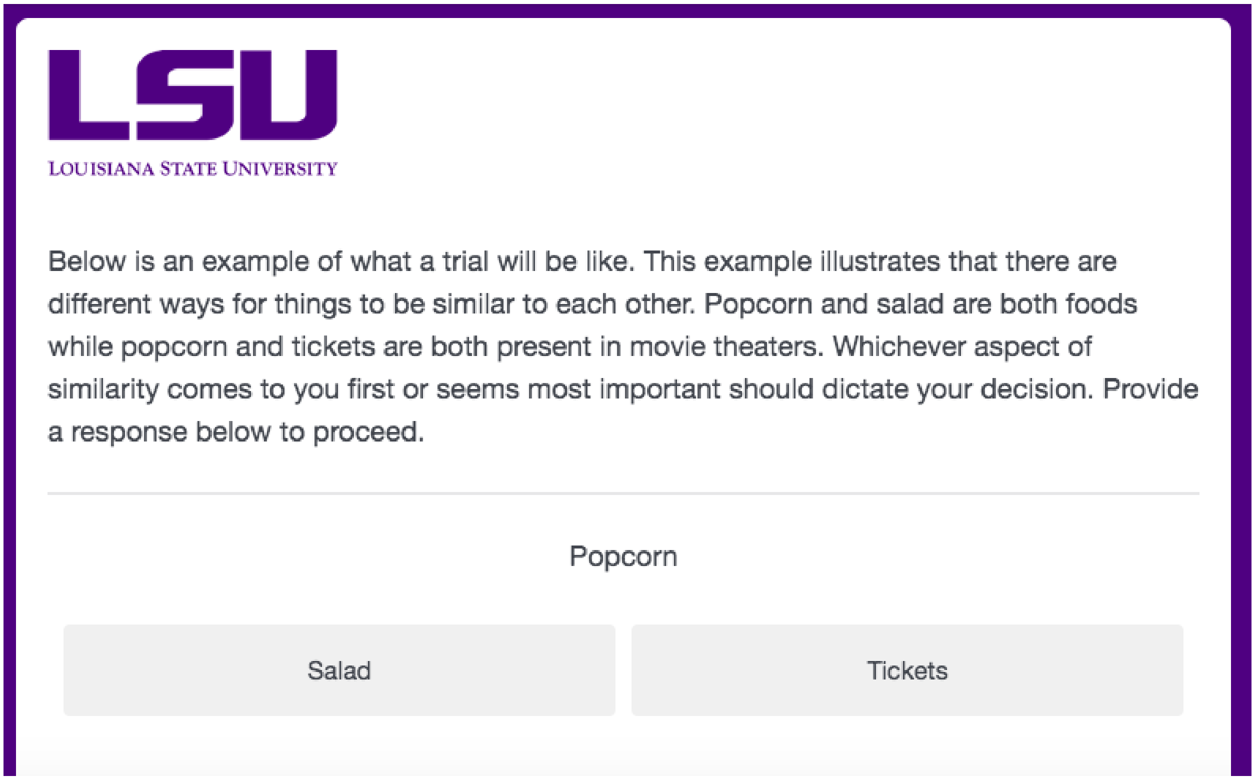


Figure 6

Image of the second example presented to the participants to illustrate that using orthographic information from the words themselves to make the similarity judgment would be inappropriate for the research task.

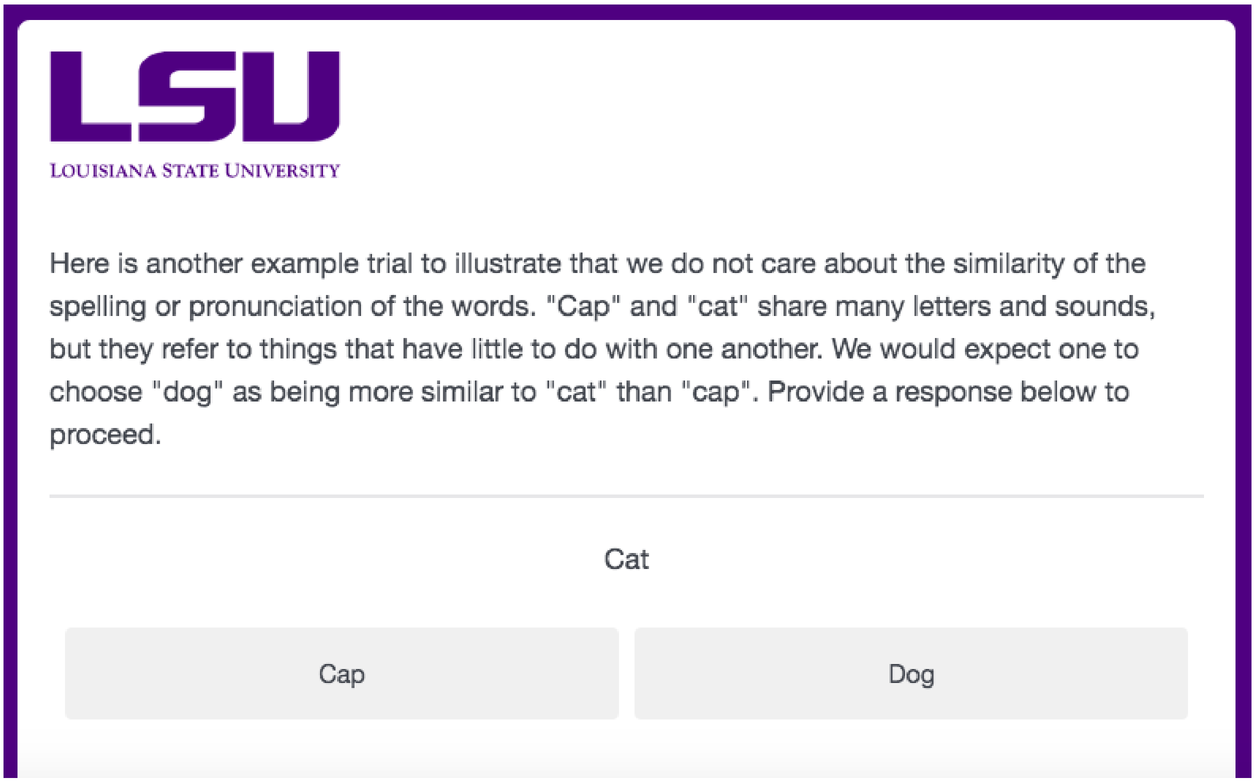


Figure 7

Image of the first follow up question after the participants responded to the 150 triplets.

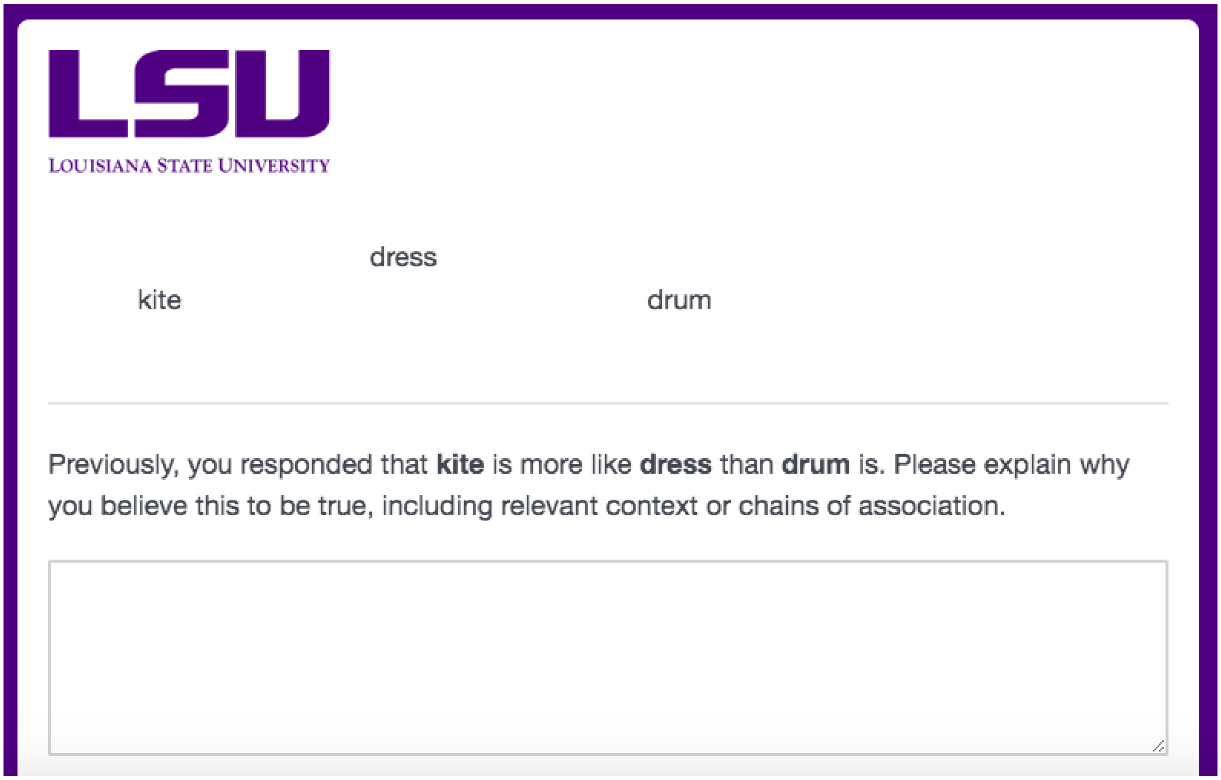


Figure 8

Image of the second follow up question after the participants responded to the 150 triplets.