4-22-2009

Jam vs. Jelly: How Children Reason Analogically with Semantically Similar Words

Sheela Ramesh
Carnegie Mellon University

Running head: CHILDREN'S REASONING WITH SYNONYMS

Jam vs. Jelly: How Children Reason Analogically with Semantically Similar Words

Sheela Ramesh

Advisor: Anna Fisher_____

Carnegie Mellon University

Abstract

Children's ability to name an object by multiple semantically similar labels (i.e. synonyms) has been documented by the age of three; however, the ability to reason using semantically similar labels remains sparsely investigated. The present study expands our understanding of children's ability to utilize semantically similar labels in two relational reasoning tasks – a semantic completion task (Experiment 1) and a semantic substitution task (Experiment 2). In Experiment 1, participants were presented a base pair of related words (e.g., castle:rock), then given a partially completed target word-pair (castle: ?) to complete with a label that made the target word-pair relationally identical to the base word-pair (e.g., stone). Additional response options included a label that was thematically related (king) and an unrelated label (milk). In Experiment 2, the same semantic relationships were explored by presenting the labels in stories with prompt questions. Results indicated that four-year-olds exhibited difficulty with semantically similar labels in the semantic completion but not in the semantic substitution task. Theoretical implications of these results are discussed.

Jam vs. Jelly: How Children Reason Analogically with Synonyms and Semantically Similar Words

Many objects in the world can be referred to by more than one label, a phenomenon called polyonomy. For example, a pet could be accurately referred to as *Fluffy*, *kitty*, *cat*, and *animal*. It is well-documented that in the beginning stages of language acquisition, children struggle with this phenomenon; however by three years of age children are able to accept and produce multiple labels in reference to the same object, both in the form of taxonomically-related labels (such as *cat-animal*) and semantically similar labels at the same level of taxonomic hierarchy (i.e. synonyms, such as *kitty-cat*) (Banigan & Mervis, 1988; Blewitt, 1994; Deák & Maratsos, 1998; Haryu & Imai, 1999; Johnson, Scott, & Mervis, 1997; Liitschwager & Markman, 1994; Mervis et al., 1994). Learning to refer to an object by more than one label may signify development of understanding that labels denote categories rather than individual objects. However, mature understanding of labels as category markers requires that one is not only capable of using multiple labels in reference to the same object, but moreover is willing to rely on multiple related labels to perform a variety of reasoning tasks, such as categorization, inductive reasoning, and analogical reasoning.

Research investigating development of the ability to use hierarchically-related labels in reasoning tasks indicates that this ability does not mature until 7- to 8-years of age (Gelman & O'Reilly, 1988; Johnson, et. al., 1997). However, research into children's ability to use semantically similar labels at the same hierarchical level – synonymous or nearly synonymous labels – in reasoning tasks has not been conducted extensively, and the existing research has produced mixed results. In particular, Gelman and Markman (1986) observed that four-year-old children can perform inductive reasoning tasks at above chance levels with identical labels (e.g.,

generalizing a property from one rabbit to another rabbit, rather than from a squirrel to a rabbit) as well as semantically similar labels (e.g., generalizing a property from a bunny to a rabbit, rather than from a squirrel to a rabbit).

However, while Gelman and Markman's study provides valuable insight into children's reasoning with semantically similar labels, several factors warrant further investigation of this phenomenon. First, some of the stimuli used in this study included taxonomically-related labels (e.g. *rose-flower* and *cobra-snake*) rather than synonyms, and the ability to reason with taxonomic labels matures later developmentally (Gelman & O'Reilly, 1988). Second, some of the semantically similar labels used in this study are likely to co-occur in the speech of children and their caregivers as compound noun-phrases (e.g., *bunny-rabbit* and *puppy-dog*) according to the CHILDES database (MacWhinney, 2000). For example, the word "bunny" occurred in CHILDES 803 times, the word "rabbit" occurred 579 times, and these words co-occurred 103 times. At the same time, other semantically similar labels used in the Gelman and Markman (1986) study (e.g., *rock-stone*) *never* co-occurred in the CHILDES database. It is likely that children associate highly co-occurring words with each other regardless of semantic relationships. Therefore, it is possible that effects of semantic similarity of labels on inductive reasoning in four-year-old children were amplified by co-occurrence frequency of some of the label pairs used in Gelman and Markman's research. In support of this hypothesis, Matlen and Fisher (2008) found that four-year-old children successfully relied on semantically similar labels in a property induction task only if these labels were likely to co-occur in child-directed speech; when provided with non-co-occurring semantically similar labels, children's performance was not significantly different from chance. Therefore, the extent to which young children can rely on semantically similar labels in the course of reasoning tasks remains unclear.

The goals of the research reported below were three-fold. The first goal was to explore to what extent young children are capable of using semantically similar labels as a basis for reasoning. Children's ability to reason with semantically similar labels had until now been investigated only within a property induction task (Gelman & Markman, 1986; Matlen & Fisher, 2008), so the research at hand utilized relational reasoning tasks to explore the conditions under which children can reason with such labels relying on semantic similarity. Unlike property induction tasks, tasks involving relational reasoning have typically been utilized in research aimed at assessing children's analogical thinking (see Goswami 1991 for review). These tasks tend to follow the format of A:B::C:?. For example, Goswami and Brown (1990) assessed 4- and 5-year-olds' ability to perform analogical reasoning tasks with familiar relations by presenting them with a base word-pair (e.g., Spider-Web) and an incomplete target word-pair (e.g., Bee-?). Children could complete the target word-pair with a relational choice (e.g., Hive), or with a word that did not preserve the relation specified in the base word-pair: a thematic lure (e.g., Honey). Goswami and Brown found that by four years of age children were competent at completing these analogies based on the relational choice, even in the presence of a thematic lure (cf Rattermann and Gentner, 1998).

The second goal of the present research was to examine children's understanding of linguistic labels as markers of category membership. It has been suggested that understanding of labels as category markers develops as early as two years of age (Gelman & Coley, 1991; Welder & Graham, 2001). Therefore, labels influence children's performance on a variety of cognitive tasks because children realize that objects referenced by the same label (or by semantically similar labels) refer to objects of the same kind, and objects of the same kind are likely to share many common properties.

However, it has been demonstrated that under many conditions auditory input, including labels, attenuates processing of corresponding visual input in infants and young children (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). As a result of attenuated visual processing in the presence of auditory information, entities that share the same label are perceived as looking more similar than the same entities presented without a label (Sloutsky & Fisher, 2004; Sloutsky, Lo, & Fisher, 2001). Therefore, it has been argued that labels influence children's performance on cognitive tasks by contributing to the overall perceived similarity automatically computed over visual and auditory features of an object. According to this theoretical perspective, understanding that labels refer to specific kinds has a protracted developmental course, with understanding of labels as category markers continuing to mature beyond preschool years.

The third goal of this research deals with children's ability to reason in applied versus abstract contexts. In many domains of learning and especially in problem-solving, the objective is generally twofold: to understand abstract principles and to comprehend their utility in specific applied situations. In problem-solving, a common strategy involves using examples of a problem in which the solution is known to solve a new problem that contains the same structure, by determining which principle to apply and how to apply it to the problem. The tendency of students to apply this strategy, analogical transfer, instead of extracting the appropriate abstract principles from each new problem, leads to the question of how children assess abstract and applied problem-solving tasks differently. Prior research emphasizes exemplar-based models (see Reeves & Weisberg, 1994, for review), in which content and exemplar-specific details are favored over abstract, content-free information for reasoning. According to this theoretical perspective, abstract information is understood as it is associated with and utilized in specific

exemplars (Medin & Ross, 1989). Therefore, presentation of reasoning tasks in the applied, familiar context of narratives may result in increased demonstration of children's reasoning abilities. Results of a study by Holyoak et al (1984) provide additional evidence for this perspective. They presented four-year-olds with analogy tasks and found that the children were more successful in reasoning analogically when the analogies were also presented as story analogs to be mapped onto transfer problems.

To achieve the goals outlined above, the present study utilized two novel reasoning tasks: a semantic completion task (Experiment 1) and a semantic substitution task (Experiment 2). In the semantic completion task children were presented with a base word-pair relation (e.g. Castle:Rock), and a partially completed target word-pair relation (Castle: ?). Children could complete the target word-pair with a label that preserved the relationship specified in the base word-pair (i.e., a label semantically similar to the second term in the base word-pair, such as "Stone") or with a word that did not preserve the relation specified in the base word-pair: a word thematically related to the first term in the target word-pair (e.g., King; in the Thematic Lure condition) or an unrelated word (e.g., Milk; in the Unrelated Lure condition). Thus, this task followed an A:B::A:B' format (where the B- and B'-terms were semantically similar, or synonymous labels). Children's ability to correctly solve relational reasoning tasks using semantically similar labels was compared to their ability to correctly solve relational reasoning tasks using identical labels.

If children have acquired mature understanding that identical as well as semantically similar labels refer to objects of the same kind (Gelman & Markman, 1986; Gleman & Coley, 1991; Jaswal, 2004), then children should have little difficulty in completing relational reasoning tasks using semantically similar labels. Furthermore, children's performance on reasoning tasks

with semantically similar labels should be similar to their performance on reasoning tasks with identical labels (Gelman & Markman, 1986). However, if understanding that labels are category markers has a protracted developmental course and is not yet complete by four years of age (Matlen & Fisher, 2008; Sloutsky & Fisher, 2004; Sloutsky, Lo, & Fisher, 2001) then children may have difficulty in completing relational reasoning tasks using semantically similar labels, though they should succeed in completing relational reasoning tasks using identical labels. Experiment 1 tests these possibilities.

In order to address the third goal, children's ability to apply relational reasoning in a concrete context, Experiment 2 utilized a semantic substitution task, in which participants were presented with a series of stories involving the same semantic labels and relationships as in Experiment 1. Whereas the A:B::A:B' task requires purely conceptual abstract reasoning, the story task also requires children to extract the underlying concept necessary to complete the task using semantically similar labels. If the ability to reason with semantically similar labels is solidly embedded in four-year-olds, children should be highly successful in choosing the semantically similar label response for both experimental tasks. If, however, children understand the context of a narrative as providing valuable problem-solving cues for analogical transfer (Reeves & Weisberg, 1994), they should be more likely to reason relationally in this task than in the abstract task. Additionally, if children comprehend information better when it is provided in a narrative context (Black & Bower, 1980) and necessarily attend to the relations between items in the text when encoding (Einstein, McDaniel, Owen, & Cote, 1990; McDaniel, Einstein, Dunay, & Cobb, 1986), it follows that they should be better able to apply the information onto the given reasoning task and should be less likely to be distracted by thematic lures. In fact, prior research suggests that preschool-aged children benefit in analogy tasks when the task is

presented in the form of a story (Holyoak, Junn, & Billman, 1984). The following experiments address the above possibilities through an abstract relational reasoning task (Experiment 1) and an applied relational reasoning task (Experiment 2).

Experiment 1

Method

Participants

Participants were 33 male and 28 female four-year-old children (M = 4.46 years, SD = .28 years) recruited from preschools and child care centers in the Pittsburgh area.

Design

For the purpose of brevity, semantically similar labels will be referred to as "synonyms" henceforth. The study had a two (Lure type: Thematic vs. Unrelated) by two (Label type: Identical vs. Synonymous) between-subjects design. Children were randomly assigned to experimental conditions. The experiment was administered on a laptop computer using SuperLab Pro software. The order of trials was randomized for each participant.

Materials

Materials consisted of 12 picture sets presented on a computer screen, accompanied by 12 label sets provided by the experimenter (see Table 1 for the list of labels used in the experiment). Each picture set consisted of a series of four pictures: the first picture contained four doors in two rows of two. One by one, the first three doors disappeared to reveal objects hidden behind them (see Figure 1 for a schematic depiction of the task). As the doors disappeared to reveal hidden objects, the experimenter labeled each object. The objects behind the first and second doors had a clear relationship, and the object behind the third door was identical in picture and

label to the object behind the first door. The fourth door revealed no hidden object, but instead the experimenter provided two response options for the participant to guess the final object.

In the Synonymous Label condition, the relational choice was communicated by a label that was semantically similar to the B-term of the base word-pair. For example, in the trial where the base word-pair consisted of the words "Castle:Rock", the relational choice consisted of the word "Stone". To avoid the potential confound of co-occurrence frequency influencing children's performance, only synonyms that never co-occurred in child-directed speech according to the CHILDES database (MacWhinney, 2000) were chosen for this study. Thus, common synonym pairs used in prior research, such as *puppy-dog* and *bunny-rabbit,* were not utilized.

In order to confirm that four-year-old children were familiar with the synonym labels used in the task, and furthermore that they accepted the multiple synonymous labels as referents to the same objects, a familiarity check was performed with a separate group of four-year-olds (N = 9, M = 4.22 years) using a picture-naming task analogous to the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997). During the task, participants (none of whom participated in the experiment proper) were presented a series of pictures of familiar objects on a computer screen. Pictures were presented four at a time, with one Target picture for the children to identify (counterbalanced in location on the screen) and three distracter pictures. Children were asked to select the Target picture according to the label spoken by the experimenter ("Point to the rock," Point to the frog," etc.). All the synonym pairs used in the experiment proper were assessed, for a total of 24 trials, during which each Target picture was referenced by both synonymous labels, on nonadjacent trials. For example, the trials for "Frog" and "Toad" included the same Target picture. The order of trials was counterbalanced across participants. Item-analyses of the

picture-naming data indicate that for all items (the B-terms and B' Synonym terms in Table 1), children pointed to the appropriate Target picture (M = .98%, all individual item $M > 0.88$, above chance, all one-sample ts > 5.75, ps < .001). This indicates that the synonym labels used in Experiment 1 were familiar to four-year-old children. Importantly, four-year-old children were willing to accept the synonymous labels chosen in Experiment 1 for the same objects.

To ensure that the children recognized and understood the relationship between the A and B terms, another calibration experiment was performed with the same group of four-year-olds who participated in the prior familiarity check (N = 9, M = 4.22 years). The familiarity check and the present calibration were performed in the same session, with the order of tasks counterbalanced. During the calibration, participants (none of whom participated in the experiment proper) were presented a series of triads depicting the A term, the B term, and an unrelated lure. Children were provided verbal labels for all three pictures and asked to select the two objects that "go together." As in the experiment proper, the labels serving as B-term were counterbalanced across participants, with half receiving, for example, "rock" and half receiving "stone." Additionally, to confirm that the children not only recognized that the A and B terms are related, but furthermore understood the nature of the relationship, the children were asked to explain why the two pictures they had chosen "go together," and their explanations were recorded. Item analyses of the calibration data indicated that all of the B-terms presented in Table 1 were judged to be identifiably related to the A-term of the target word-pair by four-year-old children: on each trail, children selected the B term over the unrelated lure and correctly specified the nature of the A-B term relationship (M = 95%, all individual item $M > 0.88$, above chance, all one-sample ts > 5.0, ps < .01).

In the Identical Label condition, the relational choice was communicated by a label identical to the B-term of the base word-pair. For example, in the trial where the base word-pair consisted of the words "Castle:Rock", the relational choice consisted of the word "Rock" (the B-terms and B'-terms were counterbalanced across participants; see Table 1 for the explanation).

In the Thematic Lure condition, the lure choice was thematically related to the A-term of the target word-pair. For example, in the trial where the base word-pair consisted of the words "Castle:Rock", thematic lure consisted of the word "King". All thematic lures were confirmed to have strong thematic relationship to the A-term of the target word-pair in a calibration experiment with a separate group of four-year-old children ($N = 12$, $M = 4.77$ years). During the calibration, participants (none of whom participated in the experiment proper) were presented a series of triads depicting two thematically related objects and an unrelated lure (the A-term, thematically-related B'-term and unrelated B'-term from Table 1). Children were provided verbal labels for all three pictures and asked to select the two objects that "go together." Item analyses of the calibration data indicated that all of the thematic lures presented in Table 1 were judged to be strongly thematically related to the A-term of the target word-pair by four-year-old children: on each trail, children selected the thematically related lure over the unrelated lure ($M =$ 94%, all individual item $M > 0.91$, above chance, all one-sample ts > 7.0, ps $< .001$). In the Unrelated Lure condition, the lure was unrelated to any of the other labels in the set.

The outcome measured was each participant's proportion of relational responses (i.e. choosing identical or synonymous labels over thematic and unrelated lures) across the 11 experimental trials after the initial practice trial confirmed that the participant understood the task.

Procedure

All participants were tested individually in a quiet room in their schools. Participants were presented with pictures of four doors in two rows of two on a computer screen. The experimenter explained that there were objects hiding behind all of the doors, and that after showing the objects behind the first three doors, the participant would have to guess what was hiding behind the last door. Out of the 12 trials, one was a practice trial and 11 were experimental trials. The practice trial for all participants consisted of the base word-pair "Bread:Jam" (see Table 1) and was always presented first. The order of the rest of the trials was randomized for each participant. The goal of the practice trial was to illustrate the task and explain to participants that the A-term and the B-term are related to each other. For instance, when bread and jam were revealed during the practice trial, participants were told, "bread and jam go together, because you put jam on bread". Participants were then asked to guess what object was hiding behind the fourth door and presented with two response options; participants were asked to choose the option that "goes with the bread the same way that jam goes with the bread". Upon completing the practice trial children were provided with corrective feedback. No feedback was provided after the experimental trials. At the conclusion of the practice trial children were told that they would keep playing the game, and that to solve the task they needed to think how the objects behind the first two doors go together. At the conclusion of the experiment all participants were thanked by the experimenters and awarded a small prize for their participation (i.e., a sticker, a slinky, a puzzle, etc.).

Results and Discussion

Proportions of relational choices in each experimental condition are presented in Figure 2. A 2-way Analysis of Variance was performed on the proportions of relational choices to determine the effect of labeling condition (Synonymous vs. Identical) and lure type (Thematic

vs. Unrelated). The results indicated a main effect of labeling condition (F (1, 61) > 33, p < .001). No other effects or interactions were significant (all F's < 2.5, p's > .12).

Follow-up analyses were performed to compare children's responses in Identical and Synonymous Label conditions. In the Identical Label/Unrelated Lure condition children averaged 94% of relational choices, whereas in the Synonymous Label/Unrelated Lure condition they averaged 73% of relational choices; while performance in both of these conditions was above chance level (chance = 50%) (both one-sample ts > 4.2, ps < .005), the proportion of relational responses was higher when children could rely on identical labels compared to synonymous labels (independent-samples t(29) > 3.2, one-tailed p < .001). Similarly, the rate of relational responding was higher in the Identical Label/Thematic Lure condition than in the Synonymous Label/Thematic Lure condition (94% and 58% of relational responses respectively, independent-samples t(28) > 4.76, one-tailed p < .001). However, the rate of relational responding in the Synonymous Label/Thematic Lure condition was not different from chance (one-sample t(16) < 1.2, p > .25).

To further understand the source of differences among conditions, analysis of the individual patterns of responses was conducted. Participants were judged to be relational responders if they selected the relational choice on at least 8 out of 11 trials (binomial p < .05). In the Identical Label/Thematic Lure condition all 13 participants exhibited relational pattern of responding, whereas only five out of 17 participants (or 29%) in the Synonymous Label/Thematic Lure condition exhibited this pattern of responding (this difference was significant, Fisher exact p < .001). Similarly, in the Identical Label/Unrelated Lure condition 13 out of 14 participants (or 93%) exhibited relational pattern of responding, whereas 9 out of 17

participants (or 53%) in the Synonymous Label/Unrelated Lure condition exhibited this pattern (this difference was significant, Fisher exact $p < .001$).

This experiment sought to examine young children's ability to rely on semantically similar labels in the context of an abstract relational reasoning task. When presented with the task of utilizing identical labels to perform relational reasoning, four-year-old children performed at ceiling levels in the presence of both unrelated and thematic lures. However, when the same task was presented with semantically similar labels, children's performance decreased significantly in both lure conditions; furthermore, when semantically similar labels were pitted against thematic lures, performance decreased to chance level.

Results from Experiment 1 provide evidence suggesting that four-year-old children can reason relationally with semantically similar labels to moderate success, as long as they are not provided with a thematically alluring alternate choice. Ceiling level performance on the Identical label conditions could be a result of a different, better-developed process such as automatic generalization, than the deliberate reasoning process used by children during the Synonym label conditions, if the task is performed successfully (Sloutsky & Fisher, 2005).

In sum, thus far we have shown that young children exhibit difficulty in reasoning relationally with semantically similar labels in an abstract reasoning task, particularly in the presence of thematic lures. In Experiment 2, we explore the extent to which the application of relational reasoning in the context of a story influences success in reasoning with semantically similar labels, both with and without the presence of thematic lures.

Experiment 2

Method

Participants

Participants were 21 male and 20 female four-year-old children (M = 4.47 years, SD = .30 years) recruited from preschools and child care centers in the Pittsburgh area.

Design

The study used a between-subjects design with the same two lure conditions used in Experiment 1. Because the participants in both identical label conditions had performed at ceiling for Experiment 1, those conditions were omitted from Experiment 2. Additionally, a control condition was created to ensure that all participants understood the task and did not have a bias toward either answer choice. Children were randomly assigned to one of the three experimental conditions. The experiment was administered on a laptop computer using SuperLab Pro software. The order of trials was randomized for each participant.

Materials

The materials in Experiment 2 consisted of twelve short narratives that were presented to participants along with pictures on the computer screen. The twelve trials explored the same twelve synonym pairs and A-B relationships examined in Experiment 1 and displayed in Table 1. Table 2 provides the stories and labels in all trials and all conditions. Each story explicitly stated the A-B relationship and was illustrated by a picture containing the A- and B-term objects, identical to the pictures used in the reasoning task in the previous experiment, but without doors (see Figure 3). Following the presentation of each story, the participant was asked a question by the experimenter, to make a prediction about an event following the story, based on the information given. The child was given the choice between the relational response (the Synonym B' term) or a distracter (either a Thematic Lure to the story or an unrelated choice).

In the Synonym/Unrelated condition, children were asked to choose between the relational response and an unrelated choice to answer the question. In the Synonym/Thematic

condition, children were asked to choose between the relational response and a lure that was thematically related to the story. Because the thematic lures from Experiment 1's task were no longer appropriate thematically, new lures were chosen. To ensure that children had no inherent bias toward the relational response (the Synonym B' term) or the Thematic lure, a control condition of the task was created. In the control condition, children were not presented with the entire narrative and A-B-term relationship; rather, they received only the prompt question and the choice between relational response and thematic response. Additionally, the control condition differed from the two experimental conditions in that children did not receive pictures, as the pictures portrayed the A and B terms, which were avoided in the control condition prompts.

Procedure

All participants were tested individually in a quiet room in their schools. The experimenter explained that they were going to play a story game where the experimenter would tell the child a story and show some pictures (except during the control condition), and then it would be the child's turn to answer a question about the story. Out of the 12 trials, one was a practice trial and 11 were experimental trials. The practice trial for all participants consisted of the base word-pair "Bread:Jam" and was always presented first. The order of the rest of the trials was randomized for each participant. The goal of the practice trial was to illustrate the task and confirm that participants understood how to complete the task. For instance, when bread and jam were revealed during the practice trial, participants were told, "Katie loves to bake bread every day, and she loves to make sandwiches by putting jam on her bread. One day, Katie bakes the perfect loaf of bread and opens the refrigerator. What do you think she is looking for?" Participants were then asked to choose between the two answers: jelly and either butter (in the

Synonym/Thematic condition) or carrots (in the Synonym/Unrelated condition). In the control condition, participants were only told "One day Katie bakes the perfect loaf of bread and opens the refrigerator. What do you think she was looking for?" and are asked to choose between "jelly" (the synonym during experimental conditions) and "butter" (the thematic lure during experimental conditions). Upon completing the practice trial children were provided with corrective feedback. No feedback was provided after the experimental trials. At the conclusion of the experiment all participants were thanked by the experimenters and awarded a small prize for their participation (i.e., a sticker, a slinky, a puzzle, etc.).

Results and Discussion

Proportions of relational choices in each experimental condition were analyzed. The practice trial data were excluded from analyses because of the corrective feedback provided. In the Synonym/Thematic condition ($N = 14$), children averaged 86% relational choices, compared to the Synonym/Unrelated condition ($N = 14$), for which children averaged 94% relational choices. Performance in both of these conditions was significantly above chance (chance = 50%) (both one-sample ts > 19.4, ps < .01) and significantly different from each other: a two-sample t-test on the proportions of relational choices determined a significant effect of lure type (Thematic vs. Unrelated), $t(13) = 2.7$, $p = .01$.

To further understand the source of differences among conditions, analysis of the individual patterns of responses was conducted. Participants were judged to be relational responders if they selected the relational choice on at least 8 out of 11 trials (binomial $p < .05$). In both experimental conditions, 100% of the participants exhibited this relational pattern of responding.

For the control condition, all trials, including the Bread-Jam practice trial, were included in analyses because no feedback was given and because the purpose of this condition was to establish both answer choices in all trials as equally appealing to children when no relational context was provided. A one-sample t-test on proportion of relational responses in the control condition ($N = 13$, $M = .57$) indicated that performance in this condition (57% correct) was not statistically different from chance, $t(12) = 1.73$, $p = .11 > .05$. Individual item-analyses of the control condition trials also indicated that for all trials, performance was not statistically different from chance (for all means, $.46 < M < .65$, all one-sample ts < 1.2, all ps $> .05$). Thus, children did not have any inherent bias toward either the relational or thematic responses when they were not provided with the relationship in the story context. However, a two-sample t-test comparing children's relational responses between the control condition (57%) and the same answer options in the Synonym/Thematic condition (86%) displays a significant difference in performance when children are presented with a story containing the appropriate information to answer the question relationally, $t(26) = 7.0$, $p < 0.01$. Therefore, it follows that the relational information provided in the two experimental conditions led to the high proportion of relational responses in those conditions.

The results of the applied story task in Experiment 2 were then compared against those of the abstract reasoning task in Experiment 1 to determine the effect of task type on proportion of relational responses (see Figure 4 for proportion of relational responses in Synonym conditions across both tasks). For the Synonym/Thematic Condition, performance on the abstract task in Experiment 1 was at 58%, not significantly above chance (one-sample $t(16) < 1.2$, $p > .25$), whereas performance on the applied task in Experiment 2 was at 86%, significantly above chance (one sample $t(13) = 19.4$, $p < .01$. A 2-sample t-test comparing performance on this

condition across tasks indicates that children chose the relational response at a significantly higher rate in the story task, $t(29) = 3.8$, $p < .01$. A comparison of patterns of individual response in this condition across both tasks was also run. For the Synonym/Thematic condition in Experiment 1's abstract task, 5 out of 17 participants, or 29%, proved to be relational responders (responding relationally on at least 8 out of 11 trials, binomial $p < .05$), whereas on the same condition in Experiment 2's applied task, all 14 participants, or 100%, were relational responders. This difference proved to be significant, Fisher exact $p < 0.001$.

Similarly, results were compared across tasks for the Synonym/Unrelated condition. In this condition, performance on the abstract task in Experiment 1 was at 73%, above chance (one-sample $t(16) = 4.3$, $p < 0.01$) but far from ceiling, whereas performance on the applied task in Experiment 2 was at 95%, also above chance (one-sample $t(13) = 11.5$, $p < 0.01$) and effectively at ceiling. A 2-sample t-test comparing performance on this condition across tasks indicates that again, children chose the relational response at a significantly higher rate in the story task, $t(29) = 3.4$, $p < .01$. A comparison of patterns of individual response in this condition across both tasks was also run. For the Synonym/Unrelated condition in Experiment 1's abstract task, 9 out of 17 participants, or 53%, proved to be relational responders, whereas on the same condition in Experiment 2's applied task, all 14 participants, or 100%, were relational responders. This difference was significant as well, Fisher exact $p < 0.001$.

Overall, data presented above indicate that four-year-old children were more successful on the story version of the relational reasoning task when they were not presented with thematically alluring choices, although it's clear that children this age can perform in this task well above chance level regardless of the nature of the lures. Moreover, children were much more successful in reasoning with synonyms when the task was applied in the context of a story

(Experiment 2) than when the task was purely in the abstract (Experiment 1). Specifically, children were able to consistently reason with synonyms in the presence of a thematic lure when doing so in the story task, where they were unable to do so in the abstract relational reasoning task.

This experiment sought to examine young children's ability to rely on semantically similar labels in the context of an applied relational reasoning task. When presented with a story and the task of utilizing semantically similar labels to perform relational reasoning, four-year-old children performed at ceiling levels in the presence an unrelated lure. Performance decreased when children were presented with a thematic lure, but was still considerably above chance level. In comparison with the same two conditions in Experiment 1, children performed significantly better on the story task in the presence of an unrelated lure, and notably were able to successfully reason relationally with synonyms in the presence of a thematic lure (which was at chance level in the abstract task of Experiment 1). In short, results of Experiment 2 provide additional evidence that young children can in fact reason with semantically similar labels, but that this ability is fragile and highly dependent on the context of the reasoning task.

General Discussion

The results from Experiment 1 indicate that children can reason with identical labels to high degrees of success regardless of lures, and can reason with semantically similar labels to moderate success when they are not presented with strong thematic lures. It could be argued that these results reported above stemmed from children being unfamiliar with some of the semantically similar labels used in the task. However, that possibility seems unlikely due to the results of the picture-naming task, which indicated that children were not only familiar with

these semantically similar labels, but furthermore that they were willing to except those multiple labels as referents to the same objects.

Another possibility is that children were unfamiliar with the relations used in the present research. This explanation seems implausible as well, because children in our sample were likely to have direct experience with many of the relations we used (i.e., cars go on roads, apples go in the tummy, etc.) and encounter other relations in books or at school. Moreover, results of the A-B-term calibration, during which children both identified and explained the relationships used in the study, seem to rule out that possibility.

A third possibility is that the findings in the Synonym label conditions are a result of children simply preferring to choose the thematically related option, rather than actually reasoning with the labels as intended. In the Synonym/Thematic label condition, both answer options were thematically related to the A-term. Therefore, if children were simply responding based on thematic preference, it follows that they should select between the two answer options at chance level, as found. In the Synonym/Unrelated condition, however, the Synonym choice was the only thematically appealing one. If children were choosing the answer choice purely thematically, it seems likely that performance would have been nearly at ceiling level, instead of at 73% as found. Instead, it is likely that children were in fact attempting to reason with the given labels, but that at four years, this ability is still fragile and in the process of developing to maturity.

Children's decreased ability to reason with semantically similar labels as opposed to identical labels in a relational reasoning task has implications for theories of learning early in development. For instance, some theories view early learning as a deliberative process of reasoning (Gelman & Medin, 1993). An alternative approach views learning early in

development as an automatic process driven primarily by the low-level mechanisms of perception, attention, and memory (Fisher & Sloutsky, 2005; Samuelson & Smith, 2000; Sloutsky & Fisher, 2004; Sloutsky & Fisher, 2005; Smith, 2000). According to the former view, when children are told that a "bunny" has a particular property and are then asked whether this property would be true of a "rabbit" or a "squirrel", children's answer is based on the understanding that bunnies and rabbits are the same kind of animal (Gelman & Markman, 1986). At the same time, it is often suggested that semantic knowledge is organized into networks of simple units, with connection strength among the units being a function of similarity, relatedness, or co-occurrence of various features (Anderson, 1983; Plaut & Booth, 2000; Rogers & McClelland, 2004; Thompson-Shill, Kurtz, & Gabrieli, 1998). Therefore, according to the latter view, children's choice of a synonymous label in the property induction tasks may stem from the fact that activation in the semantic network spreads from the word "bunny" to activate the word "rabbit", because these words frequently co-occur in child-directed speech (MacWhinney, 2000), whereas the word "squirrel" receives little to no activation from the word "bunny".

In the relational reasoning task presented above children may perform well when presented with identical labels because they can bypass the reasoning component of the task and rely on automatic matching of identical labels. However, performing the task with semantically similar labels that do not co-occur in child-directed speech, as opposed to identical labels, requires deliberative reasoning. Therefore, differences between the Identical Label and Synonymous Label conditions may arise because all children can rely consistently on the ontogenetically earlier appearing process of automatic generalization, rather than on the later developing process of reasoning (Sloutsky & Fisher, 2005).

The results from Experiment 2 suggest that four-year-old children exhibit even greater abilities to reason relationally with synonyms when they are asked to do so within the applied context of a story. These findings could arise for several different reasons. One possibility is that children were so successful at the relational reasoning task when doing so in the context of a story because the narrative provided a familiar context for reasoning. Children are inundated with stories from a very young age, and are highly familiar with narrative structure. It's possible that being presented with the task in such a familiar form simply made children more motivated to perform well. Similarly, the familiarity of the task structure may have decreased the level of difficulty in understanding how to solve the task. These possibilities seem unlikely, however, given that children fully demonstrated understanding of how to solve the task and motivation to do so in the Identical label conditions of the abstract relational reasoning task in Experiment 1.

Another possibility is that children performed better on the relational reasoning task in the story context because the A-B-term relationship was clearer. Narrative structures include tight causal chains, making salient the fact that each event in a story results in the following event (Lorch & Lorch, 1996). Thus, relationships between items and events in a story may be implicitly emphasized in a way that aids reasoning (Einstein, McDaniel, Owen, & Cote, 1990; McDaniel, Einstein, Dunay, & Cobb, 1986). Furthermore, the stories used in the relational reasoning task of Experiment 2 explicitly stated the relationship between A- and B-terms, whereas the abstract task in Experiment 1 demanded that children extract the A-B relationship in order to successfully reason with the terms. We know that young children are clearly familiar with all of the relationships used in the task, as evidenced by the results of the A-B calibration in which four-year-old children identified and explained all of the A-B-term relationships portrayed in the study. However, the extra act of extracting the relationships before using them to select

the synonym answer choice may have made the abstract task inherently more difficult than the story task, for which the relationships were directly provided. Further study is needed to draw conclusions as to the nature of this discrepancy.

A third possibility is that the real-life scenarios depicted in the stories allowed the children to test their answer choices, deciding if they made practical sense. In the abstract task, if a child was not fully utilizing an understanding that the semantically similar labels referred to common category members, there was no way for the child to check that his or her answer was logical. In the applied task, on the other hand, the context of the story allowed children to check their answers against the context of the story and potentially relate that to prior experience, confirming or disaffirming that the chosen answer was a logical conclusion to the story. For example, in the *Castle:Rock* trial, the abstract task essentially asked children to solve the analogy *Castle:Rock::Castle:B'*. If a child chose *King,* there would be no way to check whether this response made sense. In the same trial in the story task (see Table 2 for story details), the child could check the answer against the story. In this case, if the child was considering *Trees* and wanted to check, remembering that Bobby liked to collect *rocks* outside for his *castles* could dissuade the child from the answer *Trees*, which makes little sense considering Bobby's affinity for building castles with rocks, and instead prompt him to choose, correctly, *Stones*.

This possibility is consistent with exemplar-based theories of analogical transfer which suggest that content and exemplar-specific details are favored over abstract, content-free information for reasoning; that is, reasoning is better aided by referring new problems against examples, rather than applying abstract principles immediately onto new problems (Reeves & Weisberg, 1994). If extraction of abstract information from exemplars increases understanding of those abstract principles, at least insofar as their function in those specific exemplars (Medin

& Ross, 1989), the findings of the present research provide further evidence that young children can display greater reasoning skill in an applied context rather than in a purely theoretical context.

Regarding implications for theories of learning early in development, we can once again compare theories that view early learning as a deliberative process of reasoning (Gelman & Medin, 1993) against those that view learning early in development as an automatic process driven primarily by the low-level mechanisms of perception, attention, and memory (Fisher & Sloutsky, 2005; Samuelson & Smith, 2000; Sloutsky & Fisher, 2004; Sloutsky & Fisher, 2005; Smith, 2000). Evidence from Experiment 1 indicated that differences between the Identical Label and Synonymous Label conditions arose because children could consistently rely on low-level mechanisms (the latter view) earlier than on deliberate reasoning (the former view), Experiment 2 expands our understanding of this matter by providing evidence to suggest the possibility that children can in fact reason deliberately to high degrees of success with synonymous labels, when the task is provided in a rich, familiar context and certain areas of difficulty are avoided in the task presentation. It would be beneficial for future studies to explore these possibilities in greater detail.

While further research is needed to elucidate the developmental course of children's ability to utilize semantically similar labels in various forms of reasoning tasks, the present experiment contributes to the growing body of research suggesting that understanding of labels as markers of category membership has a protracted developmental course (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; Sloutsky & Fisher, 2004; Sloutsky, Lo, & Fisher, 2001), and that the development of reasoning skills stem from dependence on exemplars (Reeves & Weisburg, 1994; Medin & Ross, 1989) in addition to purely structure-mapping strategies

(Gentner, 1988) . Children's capacity to name objects by multiple labels at the same level of hierarchy demonstrates an initial step towards acquiring mature understanding of semantic similarity. However, as the present study shows, four-year-olds' conceptualization of semantically similar labels is still undergoing the process of development.

References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 2-32.

Banigan, R. L. & Mervis, C. B. (1988). Role of adult input in young children's category evolution: II. An experimental study. *Journal of Child Language, 15,* 493-504.

Black, J., & Bower, G. H. (1980). Story understanding as problem solving. *Poetics, 9,* 223-250.

Blewitt, P. (1994). Understanding categorical hierarchies: The earliest levels of skill. *Child Development, 65,* 1279-98.

Deák, G. O., & Maratsos, M. (1998). On having complex representations of things: Preschoolers use of multiple words for objects and people. *Developmental Psychology, 34,* 224-240.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-Third Edition: Manual*. Circle Pines, MN: American Guidance Services.

Einstein, G. O., McDaniel, M. A., Owen, P. D., & Cote, N. C. (1990). Encoding and recall of text; The importance of material appropriate processing. *Journal of Memory and Language, 29*, 566-581.

Fisher, A. V., & Sloutsky, V. M. (2005). When induction meets memory: Evidence for gradual transition from similarity-based to category-based induction. *Child Development, 76,* 583-97.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 65-95.

Gelman, S. A., & Coley, J. (1991). Language and categorization: The acquisition of natural kind terms. In S. A. Gelman, S. & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 146-196). New York: Cambridge University Press.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 189-209.

Gelman, S. A., & Medin, D. L. (1993). What's so essential about essentialism? A different perspective on the interaction of perception, language, and conceptual knowledge. *Cognitive Development, 8*, 157-67

Gelman, S. A., & O'Reilly, A. W. (1988). Children's inductive inferences within superordinate categories: the role of language and category structure. *Child Development, 59*, 876-87.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59,* 47-59.

Gentner, D., & Namy, L. L. (2006). Analogical processes in language learning. *Current Directions in Psychological Science, 15*, 297-301.

Goswami, U. (1991). Analogical reasoning: What develops? A review of research and theory. *Child Development, 62*, 1-22.

Goswami, U., & Brown, A. L. (1990). Melting chocolate and melting snowmen: Analogical reasoning and causal relations. *Cognition, 35*, 69-95.

Goswami, U. & Brown, A. L. (1990). Higher-order structure and relational reasoning: Contrasting analogical and thematic relations. *Cognition, 36*, 207-26.

Haryu, E. & Imai, M. (1999). Reorganizing the lexicon by learning a new word: Japanese children's interpretation of the meaning of a new word for a familiar artifact. *Child Development, 73*, 1378-91.

Holyoak, K. J., Junn, E. N., & Billman, D. O. (1984). Development of analogical problem-solving skill. *Child Development, 55*, 2042-2055.

Jaswal, V. K. (2004). Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category induction. *Child Development, 75,* 1871-85.

Johnson, K. E., Scott, P. & Mervis, C. B. (1997). Development of children's understanding of basic-subordinate inclusion relations. *Developmental Psychology, 33,* 745-763.

Lorch, R. F. Jr., & Lorch, E. P. (1996). Effect of organizational signals on free recall of expository text. *Journal of Educational Psychology, 88 (1),* 38-48.

Liittschwager, J. C., & Markman, E. M. (1994). Sixteen- and 24-month-old's use of mutual exclusivity as a default assumption in second-label learning. *Developmental Psychology, 30*, 955-968.

MacWhinney, B. (2000). The CHILDES Project: Tools for analyzing talk. Lawrence Erlbaum Associates.

Matlen, B. J., & Fisher, A. V. (2008). Development of synonym-based induction. Poster presented at the Cognitive Science Society, July 28th.

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language, 25,* 645-656.

Medin, D. LL., & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem solving, and induction. In R. J. Steinberg (Ed), *Advances in the psychology of human intelligence* (Vol. 5, pp. 189-223). Hillsdale, NJ: Erlbaum.

Mervis, C. B., Golinkoff, R. M. & Bertrand, J. (1994). Two-year-olds readily learn multiple labels for the same basic-level category. *Child Development, 65*, 1163-77.

Napolitano, A. C., & Sloutsky, V. M. (2004). Is a picture worth a thousand words? Part II: The flexible nature of modality dominance in young children. *Child Development*, 75 (6), 1850-1870.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review*, *107*, 786-823.

Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development, 13*, 453-478.

Reeves, L. M., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin, 115 (3),* 381-400.

Robinson, C.W., & Sloutsky, V.M. (2004). Auditory Dominance and its change in the course of development. *Child Development, 75 (5),* 1387-1401.

Rogers, T. T., & McClelland, J. L. (2004). Semantic Cognition: A parallel distributed processing approach. Cambridge, MA: MIT Press.

Samuelson, L. K., & Smith, L. B. (2000). Grounding development in cognitive processes. *Child Development, 71*, 98-106.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133 (2),* 166-188.

Sloutsky, V. M., Lo, Y., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels, similarity, and the development of inductive inference. *Child Development, 72,* 1695-1709.

Sloutsky, V. M., & Fisher, A. V. (2005). When induction meets memory: Evidence for gradual transition from similarity-based to category-based induction. *Child Development*,*76*, 583-597.

Smith, L. B. (2000). Learning how to learn words: An associative crane. In *Becoming a Word Learner*. New York: Oxford University Press.

Thompson-Schill, S. L., Kurtz, K. J., & Gabrieli, J. D. E. (1998). Effects of semantic and associative relatedness on automatic priming. *Journal of Memory and Language, 38*, 440-58.

Welder, A. N., & Graham, S. A. (2001). The influences of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development*, 72, 1653-1673.

Table 1

Labels provided during semantic completion task, Experiment 1

A-term	B-term[1]	B'-term: Synonym Choice	B'-term Thematic Lure	B'-term Unrelated Lure
Bread[2]	Jam	Jelly	Crumbs	Foot
Hand	Mitten	Glove	Foot	Ant
Castle	Stone	Rock	King	Milk
Cat	Couch	Sofa	Milk	Banana
Fly	Toad	Frog	Ant	Phone
Cheese	Rat	Mouse	Cracker	Dress
Apple	Belly	Tummy	Banana	Puppy
Duck	Lake	Pond	Feathers	TV
Vacuum	Carpet	Rug	Mop	Lion
Water	Ship	Boat	Fish	Cookie
Beach	Ocean	Sea	Sand	Chair
Car	Road	Street	Steering Wheel	Clock

[1] The labels that served as a B-term and a B'-term were counterbalanced across participants. For example, half of the participants were presented with the "Castle-Rock" base word-pair whereas the other half of the participants were presented with the "Castle:Stone" base word-pair. Consequently, the relational choice for half of the participants consisted of the word "Stone", and for the other half of the participants the relational choice consisted of the word "Rock".

[2] For all participants in all conditions, this was a practice trial. The order of the rest of the trials was counterbalanced across participants.

Table 2

Stories and labels provided during the semantic substitution task in Experiment 2.

Condition[3]	Story	Synonym[4]	Thematic lure	Unrelated lure
S/T and S/U	Katie loves to bake BREAD every day, and she loves to make sandwiches by putting JAM on her bread. One day, Katie bakes the perfect loaf of bread and opens the refrigerator. What do you think she is looking for?[5]	Jelly	Butter	Carrots
Control	One day, Katie bakes the perfect loaf of bread and opens the refrigerator. What do you think she is looking for?[5]	Jelly	Butter	
S/T and S/U	Bobby likes to collect ROCKS outside to make CASTLES. One day, his mom finds him in the yard and asks him what he's doing. What do you think he was playing with?	Stones	The trees	His dog
Control	One day, Bobby's mom finds him in the yard and asks what he's doing. What do you think he was playing with?	Stones	The trees	
S/T and S/U	Ashley's CAT often sleeps on the COUCH in Ashley's room. One day, Ashley decides to play with the cat, but isn't sure where she is. Where do you think she checks first?	Sofa	Litter box	Kitchen
Control	Ashley has a pet cat. One day, Ashley decides to play with the cat, but isn't sure where she is. Where do you think she checks first?	Sofa	Litter box	
S/T and S/U	Billy knew all about how RATS really love to eat CHEESE, so he thought he could catch one if he left some cheese out. One day, Billy placed a piece of cheese on the attic floor, but when he came back later, it had disappeared! What does he go looking for?	Mouse	Cat	A ball
Control	One day, Billy placed a piece of cheese on the attic floor, but when he came back, it had disappeared! What does he go looking for?	Mouse	Cat	
S/T and S/U	Susie likes her room to be very clean and VACUUMS her CARPET every single day. One day, her friend comes over and finds Susie in her room. What is she busy taking care of?	Rug	Bed	Homework
Control	One day, Susie's friend comes over and finds her in her room. What is Susie busy taking care of?	Rug	Bed	

[3] S/T and S/U refer to Synonym/Thematic and Synonym/Unrelated conditions, respectively. In the Synonym/Thematic Condition, only the Synonym and Thematic answer choices are presented. In the Synonym/Unrelated condition, only the Synonym and Unrelated answer choices were presented. In the Control Condition, only the Synonym and Thematic answer choices were presented.

[4] As in Experiment 1, the labels that served as the B term in the story and the Synonym answer choice were counterbalanced across participants. For example, half of the participants were told "Bobby likes to collect ROCKS to build CASTLES" and were offered the Synonym choice "STONES," whereas the other half of the participants were told "Bobby likes to collect STONES to build CASTLES" and were offered the Synonym choice "ROCKS." Consequently, the relational choice for half of the participants consisted of the word "STONE", and for the other half of the participants the relational choice consisted of the word "ROCK."

[5] For all participants in all conditions, this was a practice trial. The order of the rest of the trials was counterbalanced across participants.

S/T and S/U	Dottie loves cars and spends all her by the ROAD, looking at other CARS. One day, her dad can't find her in the house. Where do you think he looks for her?	Street	Backyard	Library
Control	One day, Dottie's dad can't find her in the house. Where do you think he looks for her?	Street	Backyard	
S/T and S/U	Jenny likes to eat an APPLE each day because she thinks it will keep her TUMMY healthy. Unfortunately, it doesn't seem to work and one day she gets a tummy ache! She goes to the doctor, and when the doctor asks Jenny asks what is wrong, where do you think she points?	Belly	Teeth	Foot
Control	Jenny goes to the doctor one day, and when the doctor asks Jenny asks what is wrong, where do you think she points?	Belly	Teeth	
S/T and S./U	Harry loves the water. Every day, he goes exploring on the WATER with his SHIP. One day, it gets cold out and his mother wants him to come inside, but she can't find him. When she finally spots him, what do you think he was playing with?	Boat	Surfboard	Bicycle
Control	One day, it gets cold out and Harry's mother wants him to come inside, but she can't find him. When she finally spots him, what do you think he was playing with?	Boat	Surfboard	
S/T and S/U	Molly hates the cold and wears GLOVES every day in the winter so that her HANDS stay warm. She won't go outside with bare hands. One winter day, her brother asks her to go play outside with him, so she goes to her room to find something. What do you think she was looking for?	Mittens	Scarf	Crayons
Control	One winter day, Molly's brother asks her to go play outside with him, so she goes to her room to find something. What do you think she was looking for?	Mittens	Scarf	
S/T and S/U	Daisy the DUCK lives on a LAKE and spends all of her time watching the other animals by the lake. One day, a hiker spots Daisy the Duck and tells his friend about it because his friend loves to look at ducks. When the friend goes hiking the next day, what is he looking for?	Pond	Flamingo	Field
Control	One day, a hiker goes hiking because his friend thought he should see something. What is he looking for?	Pond	Flamingo	
S/T and S/U	Ted was playing in a swamp one day, and noticed there were a lot of toads there. Whenever a FLY buzzed by, one of the TOADS would stick out its tongue and eat it! Ted realized that toads like to eat flies. The next time he saw a fly buzzing around, he figured out that it was about to get eaten. What did he think would eat the fly?	Frog	Spider	Puppy
Control	Ted was playing outside, and he saw a fly buzzing around. He figured out that it was probably about to get eaten. What did he think would eat the fly?	Frog	Spider	
S/T and S/U	Tommy liked to spend all of his summers at the BEACH because he loved to go surfing and swimming in the OCEAN. One summer, his friends came to visit him and see what he was up to. Where do you think he took them?	Sea	Swimming pool	Tennis Court
Control	One summer, Tommy's friends came to visit him while he was on vacation to see what he was up to. Where do you think he took them?	Sea	Swimming pool	

Figure Captions

Figure 1. Schematic depiction of the semantic completion task, Experiment 1: 1) All objects are hidden; 2) The first object (A-term of the base word-pair) is revealed; 3) The second object (B-term of the base word-pair) is revealed; 4) The third object (A-term of the target word-pair) is revealed; participants are asked to guess what object is hiding behind the last door.

Figure 2. Proportions of relational choices during the semantic completion task, Experiment 1, by experimental condition. Error bars represent standard errors of the mean. Dashed line represents chance level.

Figure 3. Picture displayed in the semantic substitution task, Experiment 2.

Figure 4. Proportions of relational choices, by experimental task and experimental condition. Error bars represent standard errors of the mean. Dashed line represents chance level.

Figure 1.

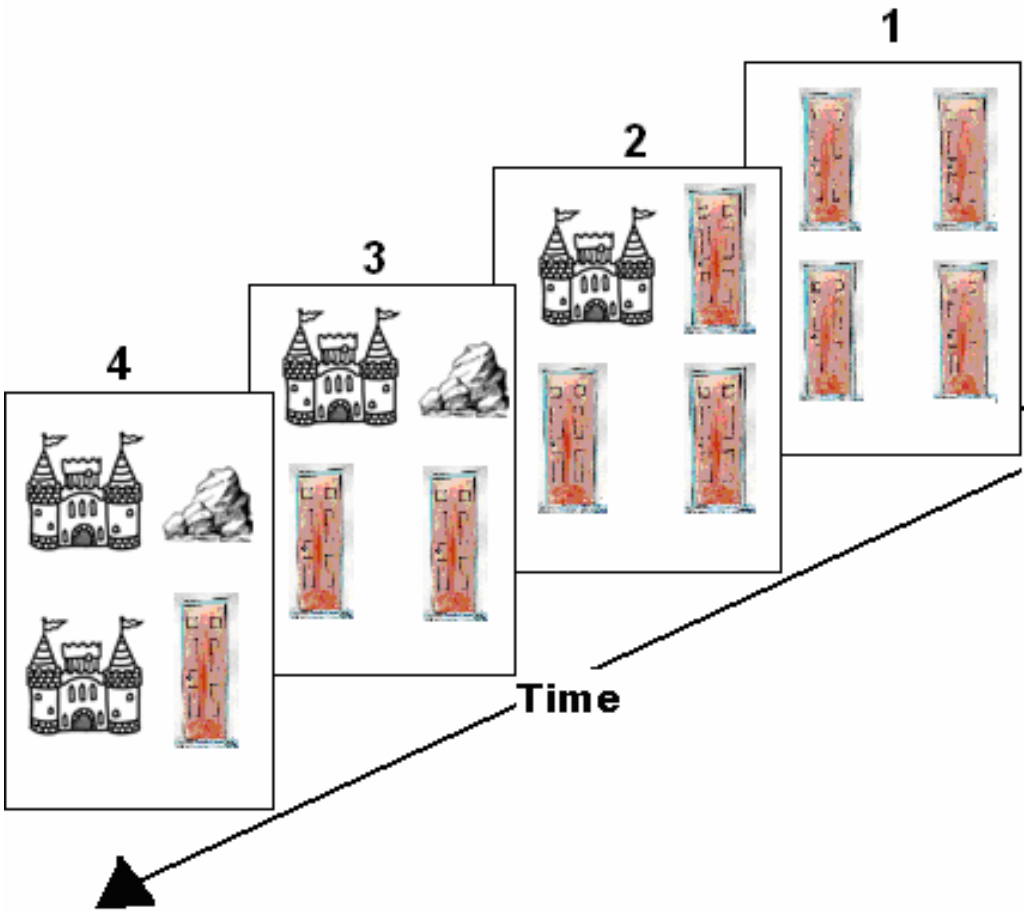


Figure 2.

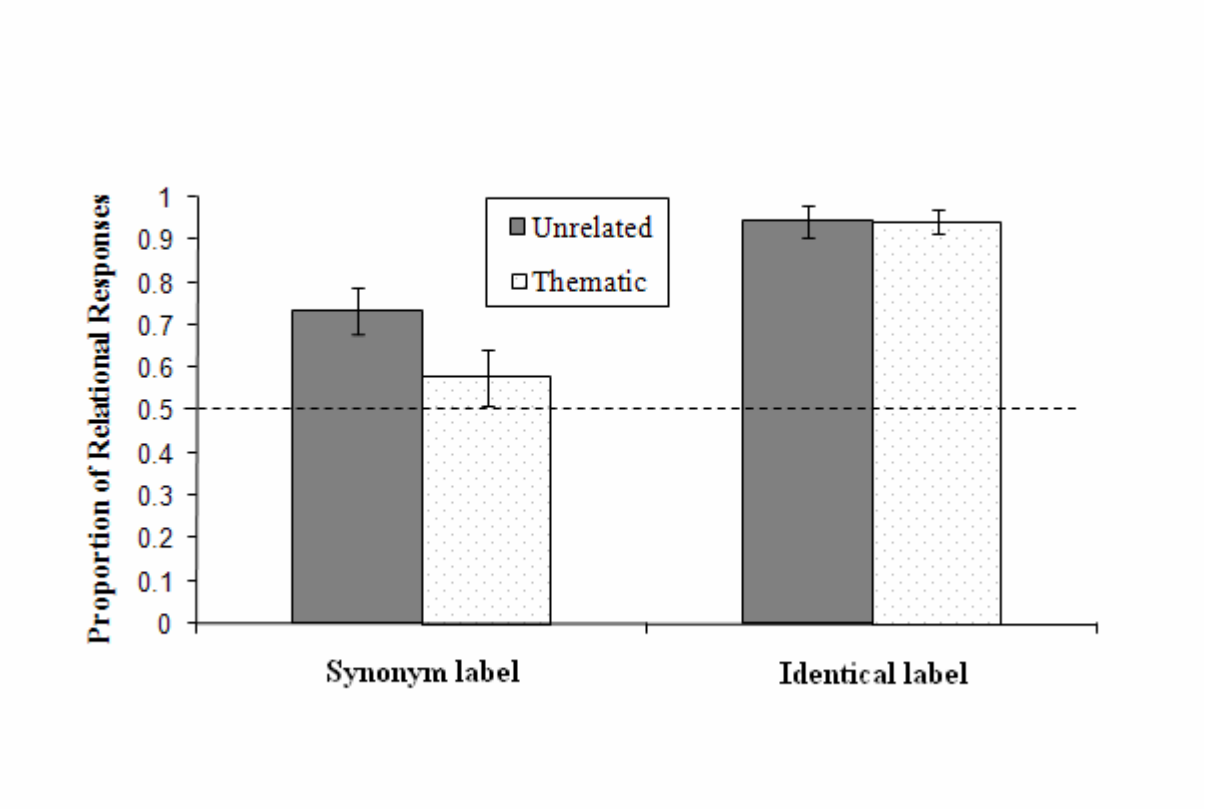
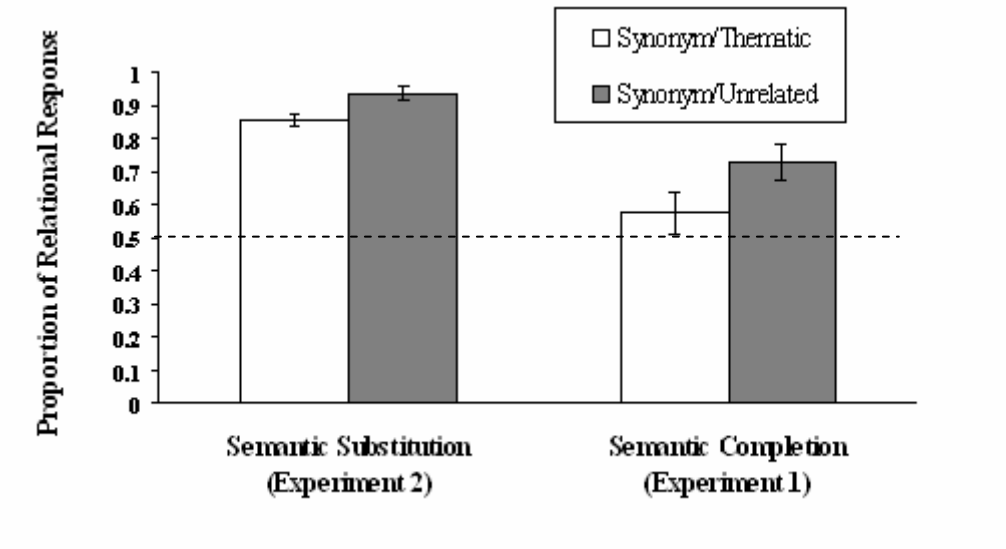

Proportion of Relational Responses
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Unrelated
Thematic
Synonym label
Identical label

Figure 3.

Figure 4.


Synonym/Thematic
Synonym/Unrelated
Proportion of Relational Response
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Semantic Substitution
(Experiment 2)
Semantic Completion
(Experiment 1)