

Cross-situational word learning in the right situations

Isabelle Dautriche and Emmanuel Chemla

Laboratoire de Sciences Cognitives et Psycholinguistique, DEC-ENS/EHESS/CNRS

Abstract

Upon hearing a novel word, language learners must identify its correct meaning from a diverse set of situationally relevant options. Such referential ambiguity could be reduced through *repetitive* exposure to the novel word across diverging learning situations, a learning mechanism referred to as cross-situational learning. Previous research has focused on the amount of information learners carry over from one learning instance to the next. The present paper investigates how *context* can modulate the learning strategy and its efficiency. Results from four cross-situational learning experiments with adults suggest that (1) Learners encode more than the specific hypotheses they form about the meaning of a word, providing evidence against the recent view referred to as “single hypothesis testing”. (2) Learning is faster when learning situations consistently contain members from a given group, regardless of whether this group is a semantically coherent group (e.g., animals) or induced through repetition (objects being presented together repetitively, just like a fork and a door may occur together repetitively in a kitchen). (3) Learners are subject to memory illusions, in a way that suggests that the learning situation itself appears to be encoded in memory during learning. Overall, our findings demonstrate that *realistic* contexts (such as the situation in which a given word has occurred, e.g., in the zoo or in the kitchen) help learners retrieve or discard potential referents for a word, because such contexts can be memorized and associated with a to-be-learned word.

Keywords: word learning, hypothesis-testing, language acquisition, memory, lexical representation

Author Note

Isabelle Dautriche, Laboratoire de Sciences Cognitives et Psycholinguistique, DEC-ENS/EHESS/CNRS; Emmanuel Chemla, Laboratoire de Sciences Cognitives et Psycholinguistique, DEC-ENS/EHESS/CNRS

Acknowledgments

The authors would like to thank Anne Christophe, Marieke van Heugten, Benjamin Spector, Judith Koehne and Lila R. Gleitman for stimulating and helpful contributions and discussions. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.313610 and was supported by ANR-10-IDEX-0001-02 and ANR-10-LABX-0087 and a PhD fellowship from the Direction Générale de l'Armement (DGA, France) supported by the PhD program FdV (Frontières du Vivant) to the first author.

Correspondence should be addressed to isabelle.dautriche@gmail.com

Introduction

Children observe their environment and learn the associations between word forms and their world referents. Yet the signal is noisy: a word is not uttered in the sole presence of its referent but in a complex visual environment where multiple word-to-meaning mappings are available (Quine 1964). One possible mechanism that may reduce the referential ambiguity is *cross-situational learning*, or the aggregation of information across several exposures to a given word (Akhtar & Montague, 1999; Pinker, 1989; Siskind, 1996).

Cross-situational learning has been studied experimentally with adults and infants (Smith & Yu, 2008; Smith, Smith & Blythe, 2011; Trueswell, Medina, Hafri & Gleitman, 2013; Vouloumanos & Werker, 2009; Yu & Smith, 2007). Typically, participants are asked to learn the meaning of several (up to 18) new words in situations simulating the ambiguity of the real world. For example, Yu and Smith (2007) exposed adults to a series of learning trials containing n words and a set of n possible referents. Each trial separately was thus underinformative, but towards the end of the study, participants selected the correct referent at greater-than-chance levels. Participants' success in these paradigms has been taken as evidence for an *accumulative account* of word learning (Smith & Yu, 2008; Smith et al., 2011; Vouloumanos & Werker, 2009; Yu & Smith, 2007). According to this view, each time a new word is uttered, children entertain a whole set of situationally plausible meanings and learning entails pruning the potential referential candidates as new instances of the word. The word-meaning mapping thus starts as a one-to-many association.

Such accumulative account of word learning has recently been challenged by an alternative *hypothesis-testing account*. (Medina, Snedeker, Trueswell & Gleitman, 2011; Trueswell et al., 2013). Unlike the accumulative account, the hypothesis-testing strategy does not require learners to remember multiple referents for a given word. Instead, based on a single exposure to a given word, a learner selects the most plausible interpretation of this word (a process referred to as *fast-mapping*). As new information becomes available in subsequent word usages, this hypothesis may be confirmed or falsified. In the case of falsification, the old referential candidate is promptly replaced by a new one. Thus, according to this view, word-meaning mapping involves a one-to-one association, which continues to be updated until it reaches a stable (adult) stage. Support for such an account comes from the observation of the sequence of hypotheses learners formulate during the course of word learning. In a modification of the original experiment of Yu and Smith (2007), Trueswell and colleagues (2013) presented adults with a series of learning trials containing *one* word and n candidate referents and asked subjects to select the word meaning at each trial. In line with previous work, participants learned the meaning of words over the course of the study. However, contrary to previous experiments in which analyses focused on participants' final performance, Trueswell and colleagues examined participants' trial-by-trial accuracy. Crucially, they found (a) that participants persisted in their choices (e.g., if they picked dog as the meaning for the word *blicket*, they would maintain this hypothesis as long as it is confirmed by the learning situation) and (b) participants picked a new meaning hypothesis *at chance* among the available candidates otherwise (we propose a refinement of this measure below). This was taken as evidence that participants had no memory for previously seen referents beyond the one they entertained as a possible meaning, as predicted by an hypothesis-testing account.

Work on cross-situational learning has typically focused on the nature of the word-meaning mappings during the learning process. On the one hand, a complete one-to-many word-meaning mapping (following the accumulative account) seems implausible given the memory cost this presupposes. On the other hand, one-to-one word-meaning mappings (following the hypothesis testing account) implies that a vast amount of potentially useful information is lost along the way.

In this study, we investigate one potential source of information left out by these two extreme views, the broader *context* of the learning situation, and examine its role in constraining word learning strategies. Although naturalistic word learning environments introduce a potentially more complicated set of referent candidates that are typically eliminated in lab-based settings, this richer context may in fact contain more structure and could, as a result, help learning. That is, the set of possible referents for a word in a real learning situation is not a pseudo-random set of unrelated objects; they co-occur in the real world and this could play an important role in cross-situational learning.

Our reasoning is best introduced with an example. In a zoo, people naturally talk about animals, whose name children may or may not know (“do you see the *blicket* there?”, “the *dax* seems hungry today!”). An accumulative word learner would encode the full one-to-many word-meaning mapping as constrained by the situations for each occurrence of a new word (a ‘*blicket*’ could mean lion, elephant or monkey, and so could ‘*dax*’ as this word has been heard in the same situation). By contrast, a hypothesis-testing learner would bind each word to one chosen referent (a ‘*blicket*’ could mean a lion while a ‘*dax*’ could mean a monkey). In both cases, however, subsequent learning could be constrained at a different level if the learner encodes that these words were encountered in a zoo. Hence, the information that a *zoo*-word refers to an animal may persist beyond the specific situation in which it was uttered and on top of the currently entertained hypotheses. In other words, learners may encode higher order properties of situations and use it to deduce meaning across situations (“I heard *blicket* in the zoo, it must be one of these animals...”).

We thus propose to investigate to what extent cross-situational learning relies on context to develop word-meaning mappings. To this end, we first replicate the results of previous word learning experiments using a paradigm similar to Trueswell et al. (2013) (Experiment 1) and introduce a novel measure that quantifies the amount of information stored and retrieved across trials in such a paradigm. Second, we investigate whether introducing more ecologically valid situations would further boost memory retrieval of previously encountered referents. Specifically, we manipulate higher order properties of a word-learning situation: the semantic relation among the possible referents (Experiment 2) and context consistency (Experiment 3) and test their effects on participants’ learning strategy using the measure developed in Experiment 1. And finally, we demonstrate that if context can improve word learning, this improvement is subject to memory illusions, in a way that suggests that the learning situation itself is memorized and associated to novel words during cross-situational learning (Experiment 4).

Experiment 1

We conducted a classical word-learning experiment using a paradigm similar to that used by Trueswell et al. (2013). Participants were exposed to a sequence of learning instances. In each instance, participants saw four images and a sentence featuring a to-be-learned word (e.g., “There is a *blicket* here”). At each learning instance participants were asked to select a plausible referent for the word (based on the current and past information they received). The correct word referent was present in all learning instances for that word.

Our goal was to develop a measure suitable to quantify the amount of information that participants store and retrieve from a previous learning instance. Our measures differed from the one used in Trueswell et al. (2013) in two ways. First, we did not base our measure on the actual accuracy of answers, but solely on their compatibility with previous learning instances. Second, we focused on learning instances of a word *W* where the referent selected in the previous learning instance for *W* is absent (and not on all cases in which this previous choice was incorrect, as Trueswell et al. did). According to the hypothesis-testing view, if participants remember only their conjecture for *W*, *these* are the cases in which they should randomly pick a novel referent among the current candidates since they cannot confirm their previous hypothesis. By contrast, if participants remember more than their single previous hypothesis for the word, their choice of a new referent should be informed by the set of referents that were present in previous learning instances.

Method

Participants. Fifty adults were recruited through Amazon Mechanical Turk (22 females, $M=34$ years, 48 native speakers of English, as per voluntary answers given on a questionnaire at the end of the experiment). The experiment lasted between 5 and 10 minutes and participants were paid \$0.85.

Stimuli & Design. Twelve phonotactically legal English non-words were selected from <http://elexicon.wustl.edu/> (*blicket*, *dax*, *smirk*, *zorg*, *leep*, *moop*, *tupa*, *krad*, *slique*, *vash*, *gaddle*, *clup*)¹, as well as 12 objects representing these non-words (*cat*, *dog*, *cow*, *rabbit*, *pants*, *hat*, *socks*, *shirt*, *pan*, *knife*, *bowl*, *glass*). For each of these 12 objects, five different photographs were selected. The one-to-one pairing between the 12 non-words and the 12 objects was fully randomized and differed for each participant.

The trial design follows the same constraints as that in Experiment 1 of Trueswell et al. (2013) with the exception that each learning instance contained 4 possible referents in our study, but 5 possible referents in theirs. As represented in Figure 1, each trial was a learning instance for a given word, e.g., *blicket*, consisting of 4 pictures aligned horizontally on a white background along with a written prompt “There is a *blicket* there”. The pictures were selected pseudo-randomly such that (1) the correct referent was always represented, (2) no incorrect referent occurred with a word more

¹ As one reviewer pointed out, three of these words are actually real words: *smirk*, *leep* and *slique* (although the latter two are spelled differently). However, because accuracy was not predicted by word type (non-word vs. real words: $z < 1$; $p > 0.4$), it is unlikely that *only* the small group of real (but infrequent) words induced the observed results.

than twice in the experiment, (3) each object appeared the same number of times (5 times as the correct referent and 15 times as a distractor), (4) all pictures occurred the same number of time in the experiment. There were 5 learning instances per word during the experiment, resulting in a total of 60 trials. The experiment consisted of 5 blocks each of which contained 12 trials, one for each to-be-learned word. The list of 12 words occurred in the same order in each of the blocks.

Procedure. Participants were tested online. They were instructed that they were to learn words by associating them with images displayed on the screen. Prior to test, participants were given a screenshot of a learning instance involving a word and a set of pictures that were not used at test). No information about the number of to-be-learned words or the number of learning instances was given. For each trial, participants were asked to click on the image they believed could represent the meaning of the word. Once they responded, the test continued with the next trial. We recorded participants' answers at each trial as well as their response times.

Data processing. Five participants were excluded from our analysis for obvious violations of the instructions (two always selected the left image, three had RT patterns indicating that they were 5 to 10 times faster in the last block than in the first and second block – including these participants in the analyses does, however, not impact the pattern of results). We also removed 5 responses out of 3000 for being implausibly fast (below 1 second) or slow (above 30 seconds, following Smith et al. (2011)). Participants who provided 50 or fewer responses out of 60 were discarded (but this criterion did not eliminate any participants in this first experiment).

Data analysis. Participants' responses were coded as 0 (incorrect) or 1 (correct) for each trial. Since we analyzed categorical responses we modeled them using logit models as recommended by Jaeger (2008). We ran mixed model analyses using R 2.15 and the lme4 package (Bates and Sarkar, 2007), plots have been realized using the ggplot2 package (Wickham, 2009). β estimates are given in log-odds (the space in which the logit models are fitted), with the odds of an event defined as the ratio of the number of occurrences where the event took place to the number of occurrences where the event did not take place. Significant positive β estimates indicate an increase in the log-odds, and hence an increase in the likelihood of occurrence of the dependent variable with the predictor considered (calculated using the inverse logit function (logit^{-1})). We computed two tests of significance: the Wald's Z statistic, testing whether the estimates are significantly different from 0, and the χ^2 over the change in likelihood between models with and without the considered predictor. Since the results did not change between the two tests, we report the Z statistic only.

The random effect structure chosen for each model is the maximal random effect structure justified by model comparison and supported by the data. We followed the procedure outlined in Baayen, Davidson and Bates (2008), starting with the full random effect structure and reducing the structure on a step-by-step basis until excluding a random term resulted in a significant decrease of the log-likelihood compared to the model including it. For the sake of clarity, the χ^2 comparisons between models are not reported.

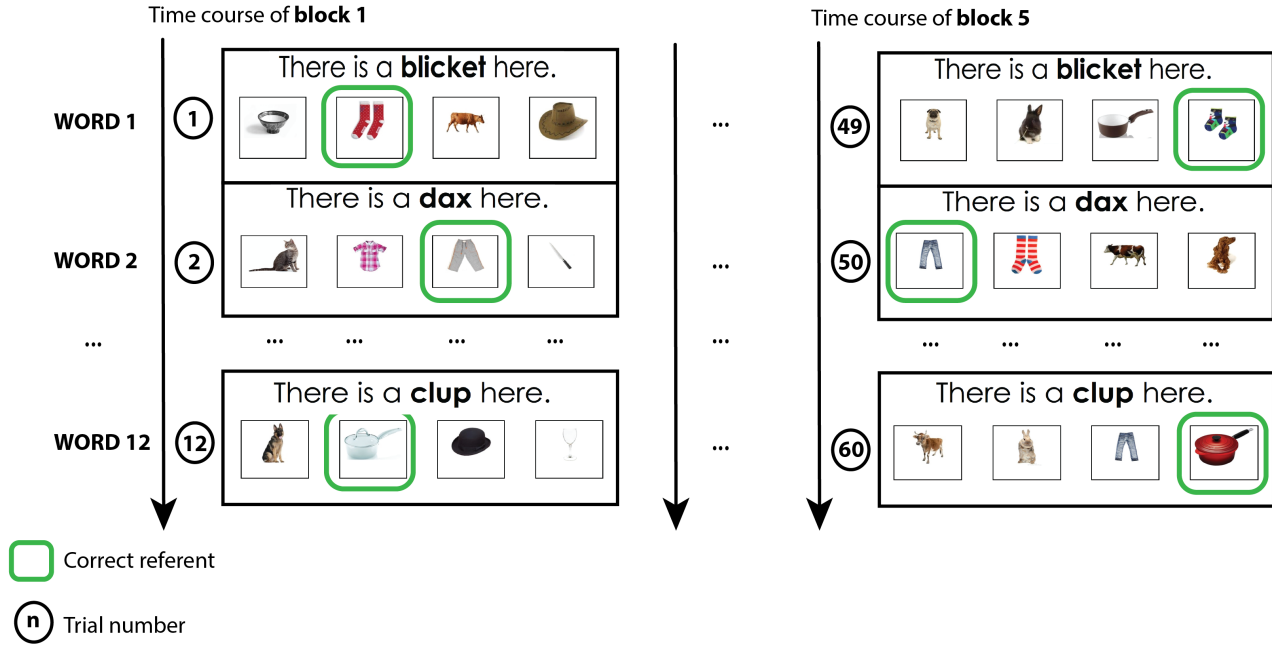


Figure 1: Experimental design. A learning trial of a to-be-learned word is a set of 4 candidate referents presented with the word in a simple declarative sentence. The 5 learning instances for each word are distributed in 5 blocks such that there is exactly one learning instance for a given word per block, hence 12 trials per block. As depicted, each block is an ordered list of 12 trials, such that there are exactly 11 intervening trials between two learning instances of the same word. This resulted in a total number of 60 trials. The word-referent pairings were randomly assigned for each participant.

Results & Discussion

We report three analyses looking at: (1) the learning curve, (2) accuracy as a function of the previous response, following Trueswell et al. (2013), and (3) a novel measure characterizing information retrieval from prior experience.

(1) Learning curve: a replication. Figure 2 presents participants' accuracy in each block. We modeled the accuracy with a mixed logit model using a predictor Block (1 to 5) with subjects and words as random effects on intercepts plus a random slope for the effect of Block with subjects. We found a significant effect of Block on accuracy ($\beta = 0.36$, $z = 10.25$, $p < 0.001$). The β coefficient indicates that for every new block, participants were 59% ($\text{logit}^{-1}(0.36)$) more likely to be accurate than in the previous block. We thus replicate previous findings showing that participants gradually learned word-meaning mappings across learning instances (Yu & Smith, 2008; Trueswell et al., 2013).

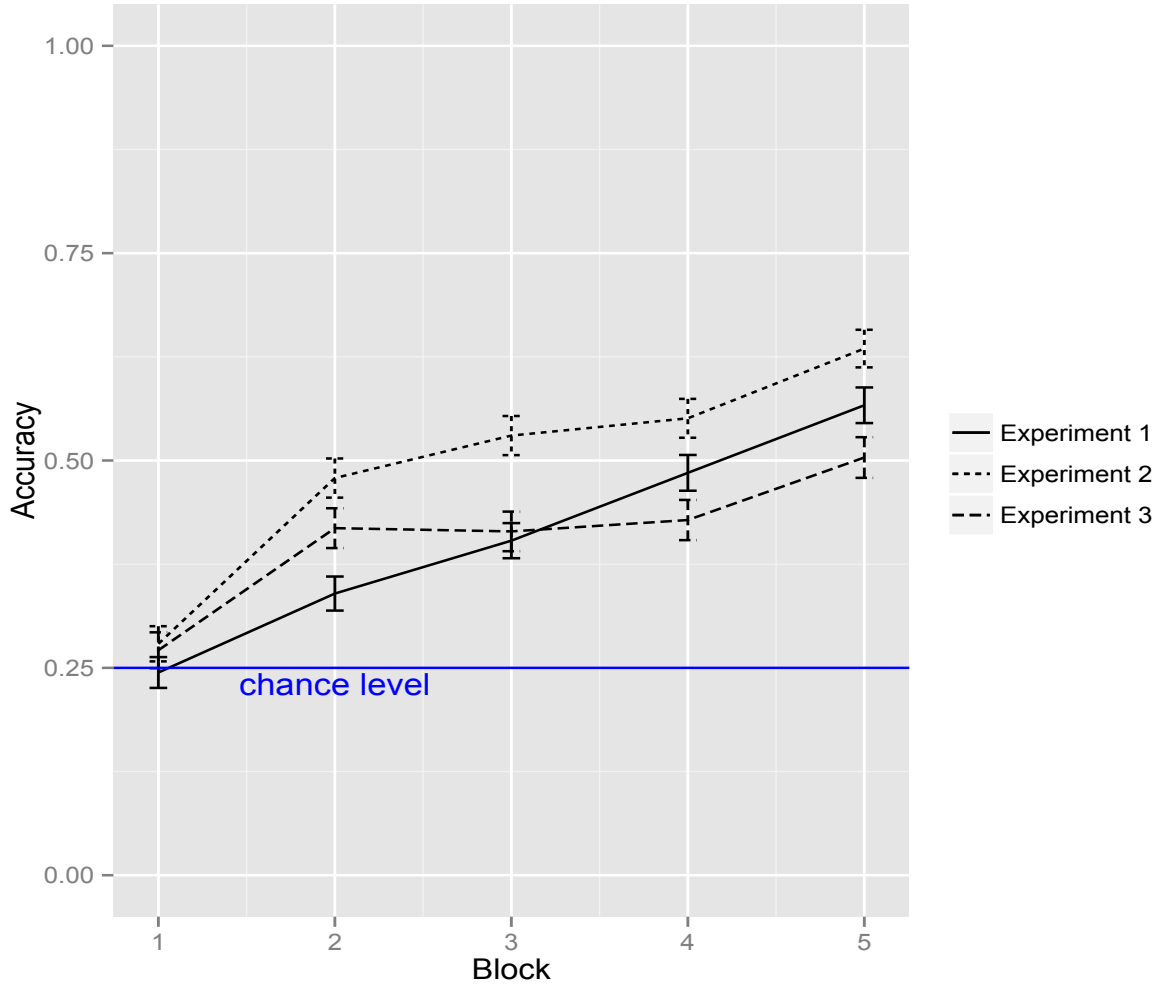


Figure 2: Learning curves. Average accuracy aggregated by subject for each block in Experiments 1-3. Error bars indicate standard error of the mean.

(2) Trial-by-trial analysis: Accuracy dependent responses. Using Trueswell et al.'s analysis on participants' responses, we compared the average proportion of correct responses in blocks 2-5 depending on whether the previous referent selection for that particular word was correct or incorrect (Figure 3). We modeled the proportion of correct responses using a predictor Previous Response Accuracy (Correct vs. Incorrect) with subjects and words as random effects on intercepts and a random slope for the effect of Previous Response Accuracy with subjects. We applied an offset corresponding to the logit of the chance level to the model (i.e. .25, the probability of being correct in a trial) to compare the intercept against chance level. We found a main effect of Previous Response Accuracy ($\beta = 1.40$, $z = 10.94$, $p < 0.001$) showing that participants were 80% ($\text{logit}^{-1}(1.40)$) more likely to be accurate when they were correct on the previous learning instance than when they were incorrect.

We then compared participants' average accuracy against chance level separately depending on whether their previous response was correct or incorrect. We found that (a) participants' accuracy was significantly above chance when they had been correct in the previous learning instance for that word (789 data points, $\beta = 3.13$, $z = 12.10$, $p <$

0.001), (b) accuracy also exceeded chance after being incorrect in the previous trial (1339 data points, $\beta = 0.33$, $z = 3.48$, $p < 0.001$).

While (a) aligns nicely with the results from Trueswell et al., (b) does not. Instead, Trueswell et al. found that after an incorrect response participants were at chance in the next learning instance.

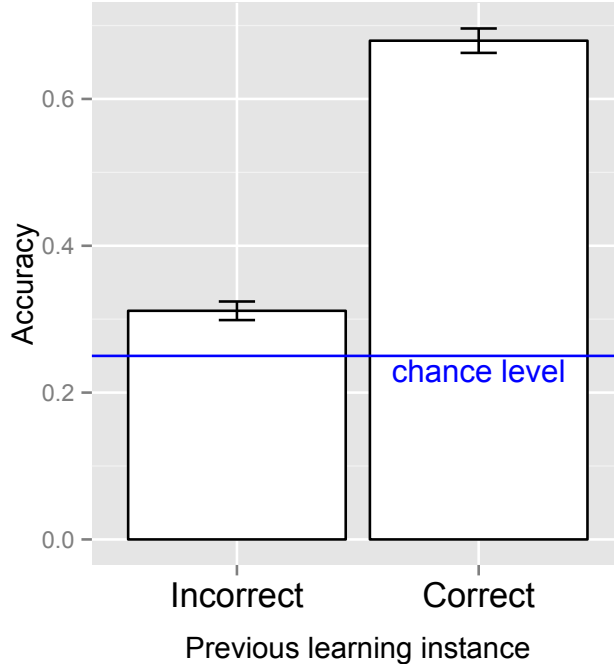


Figure 3: Accuracy dependent measure. Accuracy in blocks 2 to 5 for previously correct or incorrect words in Experiment 1. Error bars indicate standard error of the mean.

The apparent difference between our results and Trueswell et al.'s results could be explained when one takes into account that the current analysis collapses two situations for which the hypothesis-testing strategy predicts different behaviors: (I) if the participant's previous selection is present, participants should repeat their incorrect previous hypothesis and (II) if it is not present, participants should be at chance in selecting the correct referent. Hence, the outcome of this analysis is dependent on the proportion of instances of type (I) and (II).

Both Trueswell et al.'s first experiment and the present experiment are constrained in the same way: no object can be repeated more than twice as a distractor for a given word. However, in Trueswell et al., each trial displayed five possible referents (in contrast to the four referents displayed here), hence objects had to be repeated more often as distractors to account for the additional fifth picture on each trial. While both occurrences for a given distractor are not necessarily in two subsequent trials for a given word, there should be a higher proportion of instances of type (I) in Trueswell et al.'s study than in the present experiment (12% of type (I) instances on the total number of trials where the previous choice is incorrect). Since type (I) trials lead to incorrect responses, this difference could explain why the analysis reveals better results for the current experiment.

(3) New analysis: a measure of information retrieval. To distinguish learning strategies based on one-to-one and one-to-many word-meaning mappings, we need to quantify the amount of information stored and retrieved at each learning occasion during cross-situational learning. In the following, we propose such a measure.

We selected from block 2 all learning instances of type (II), i.e. learning instances for a word x in which the participant's choice for x from block 1 is not present. Figure 4 represents a measure of selecting a response that is informed by previously seen referents. Specifically, for each trial, we computed the set S of referents that were also present in the first block for this word. Figure 4 represents the proportion of responses that belong to S minus the expected proportion of falling in S by chance (cardinal of S divided by 4). We modeled the proportion of responses that belong to S with subjects and words as random effects on intercepts and applied an offset corresponding to chance to the model. Note that chance level of selecting a referent present in the previous learning instance is now trial-dependent (1, 2 or 3 images could be repeated from the previous trial), hence the offset applied to each trial was the logit of the corresponding chance level of selecting a previously seen referent (.25, .50 or .75).

The resulting measure significantly exceeds zero (336 data points, $\beta = 0.27$, $z = 2.28$, $p < 0.05$), i.e. participants were more likely than chance to select a previously seen referent. This result is not specific to block 2: while considering all learning instances where participants' previous choice was not present, the measure also significantly exceeded zero (1172 data points, $\beta = 0.27$, $z = 3.77$, $p < 0.001$). This analysis shows that in our paradigm, participants store more than a single hypothesis for the meaning of a word. Specifically, we show that participants resorted to previously encountered, but not chosen, referents in cases where their previous hypothesis is irrelevant.

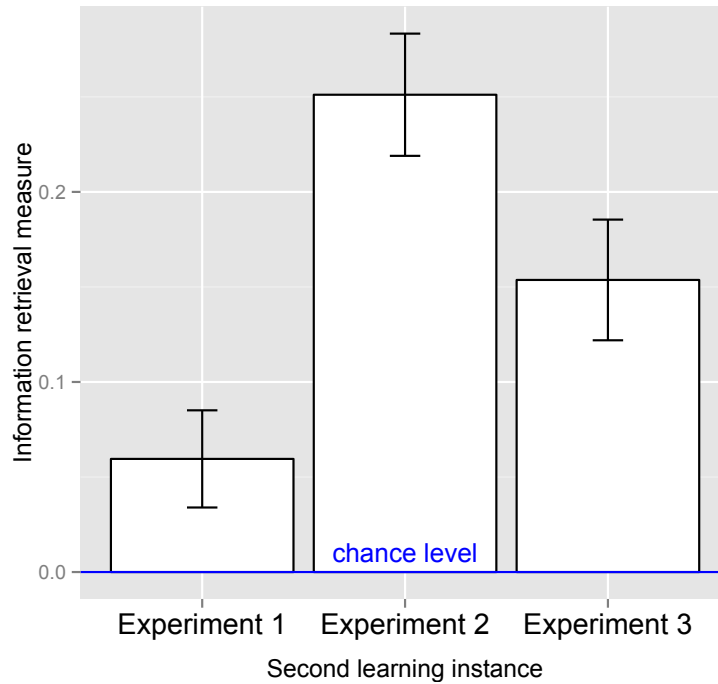


Figure 4: Information retrieval measure. Corrected tendency to select a previously seen referent: average for all second learning instances of 1 or 0 (depending on whether the answer

was in the previous learning instance) minus the chance of selecting a referent present in the previous learning instance.

Although Trueswell et al. did not employ this analysis, one would expect to find the same result in one of their experiments, their Experiment 3. In this experiment, participants were presented with only two objects on the screen at a time and, crucially, no single object was used twice as a distractor for a given word. Hence, the accurate answer (not selecting the distractor) corresponds to the only response that is fully coherent with previous learning instances (since the distractor was never previously presented with this word). The two measures are thus merged here. Yet, Trueswell et al. did not find improved accuracy following an incorrect selection. To explain the absence of evidence for accumulative learning in Trueswell et al.'s Experiment 3, one could think about reasons why 2-object and 4-object trials trigger different strategies. It is possible that participants' strategy depends on a tradeoff between the cost and the incentive to remember more than a single conjecture for a word in a given experimental situation. While memorizing two possible referents is easier than memorizing four possible referents, it is not clear that there is a real advantage of doing so to succeed in the task. Remembering only the object guessed means remembering 50% of the whole scene in the 2-object trial, hence an already quite high probability of success in the next trial (where chance is already at 50%). While the cost of remembering the objects may be higher in the 4-object trial, there would also be more incentive to do so given the higher ambiguity following the lower probability of success (chance is at 25%, so it may be worth investing resources into enhancing this probability). Albeit speculative, superficial aspects of the experimental situation could thus in principle alter participants' strategy. We leave the exploration of this issue for future research. Our current goal is to investigate the effect of context on prior experience retrieval, and we will do so with the novel, more restrictive measure we proposed.

4) Control analysis: participants' strategy in online vs. in lab experiments. So far, the discussion has not considered the possibility that there could be more fundamental differences between Trueswell et al.'s paradigm and ours. For instance, our participants were not present and monitored in the lab. It is thus possible that they completed the task in a different way (e.g., taking notes) and that their performance would therefore not reflect the natural learning ability. To assess this possibility, we analyzed participants' response times and we gathered more information about our population in a replication of Experiment 1.

1) In Experiment 1, participants took on average 5323ms (SE: 68) to associate a meaning to a word, making it unlikely they took notes. More objectively, a linear regression on the participants' accuracy in the final block using average RT throughout the experiment as a predictor did not reveal any effect of RT on accuracy ($z = -1.24$, $p > 0.2$). This suggests that there is no division within the population between participants who would have taken notes (thus being slow and accurate) and those who would not have taken notes (thus being relatively fast and inaccurate).

2) The same experiment was administered to 30 new participants recruited in exactly the same way from the same population. The crucial difference was the addition of a question at the end of the final questionnaire: "Did you take notes during the task?". Among the 28 participants who finished the task, none of them reported taking notes, suggesting that the new participants performed the task in the appropriate way. The results of this control experiment patterned with those of Experiment 1 on the three

analyses that were conducted.² This suggests that the methodology used in Experiment 1 corresponds to the type of cross-situational learning exercise we are interested in.

Summary. Our results provide evidence that participants store more than simple one-to-one word-meaning mappings. In the next experiments, we investigate whether external constraints on simultaneously presented referents for a word can alter prior information retrieval.

Experiment 2 – Encoding semantic relation

We adapted Experiment 1 to evaluate one of such contextual constraint: the semantic relation among the possible referents. We modified the first block such that all four pictures on each trial corresponded to one of the following natural categories: animals (dog, cat, rabbit, cow), dishes (pan, bowl, knife, glass), clothes (pants, socks, shirt, hat). For instance, if *blicket* referred to a dog, the three other distractor images it co-occurred with were *all* possible animal referents, mimicking a zoo-context. Furthermore, words belonging to a given category were presented on consecutive trials (allowing the learner to first learn words related to the zoo, and then words related to a bedroom and so on). By imposing these constraints on the situation of the first learning instance, we hope to reduce the overall memory cost for encoding the situation and thus improve cross-situational learning. As a consequence, we expect an increase in performance in the second learning instance for Experiment 2 compared to Experiment 1.

Method

Participants. Forty adults were recruited from Amazon Mechanical Turk (25 females, $M=40$ years, 37 native speakers of English). Two participants were excluded from our analysis because over 20% of their responses fell outside the 1-30 seconds response time window (See Experiment 1 – Analysis).

Stimuli, Design. The stimuli and the design were the same as in Experiment 1 except for new constraints on the first block of learning instances (see Figure 5 for a schematic description): (1) on all trials of the first block, each word was presented along with distractors from the target object category: *animals*, *clothes* or *dishes*, (2) the words from a given category were presented in consecutive trials.

² Regarding the learning curve, we modeled the accuracy with a predictor Block (1 to 5) and a predictor Experiment (Experiment 1, Control) with subjects and words as random effects on intercepts. There was no effect of the predictor Experiment ($z < 1$, $p > 0.4$) showing that the learning curves were similar.

Furthermore, accuracy was modeled after an incorrect response with a predictor Experiment (Experiment 1, Control) with subjects and words as random effects on intercepts and an offset of the chance level. There was no effect of the predictor Experiment ($z = -1.5$, $p > 0.1$) showing that control participants' accuracy after an incorrect response was not different from earlier participants' ($M_{\text{control}} = 0.32$; $SE_{\text{control}} = 0.02$; $M_{\text{exp1}} = 0.31$; $SE_{\text{exp1}} = 0.02$).

Finally, we modeled our measure of information retrieval with a predictor Experiment (Experiment 1, Control) with subjects and words as random effects on intercepts and an offset of the chance level and found no difference between the control and Experiment 1 ($z = 0.2$, $p > 0.8$; $M_{\text{control}} = 0.06$; $SE_{\text{control}} = 0.02$; $M_{\text{exp1}} = 0.06$; $SE_{\text{exp1}} = 0.01$).

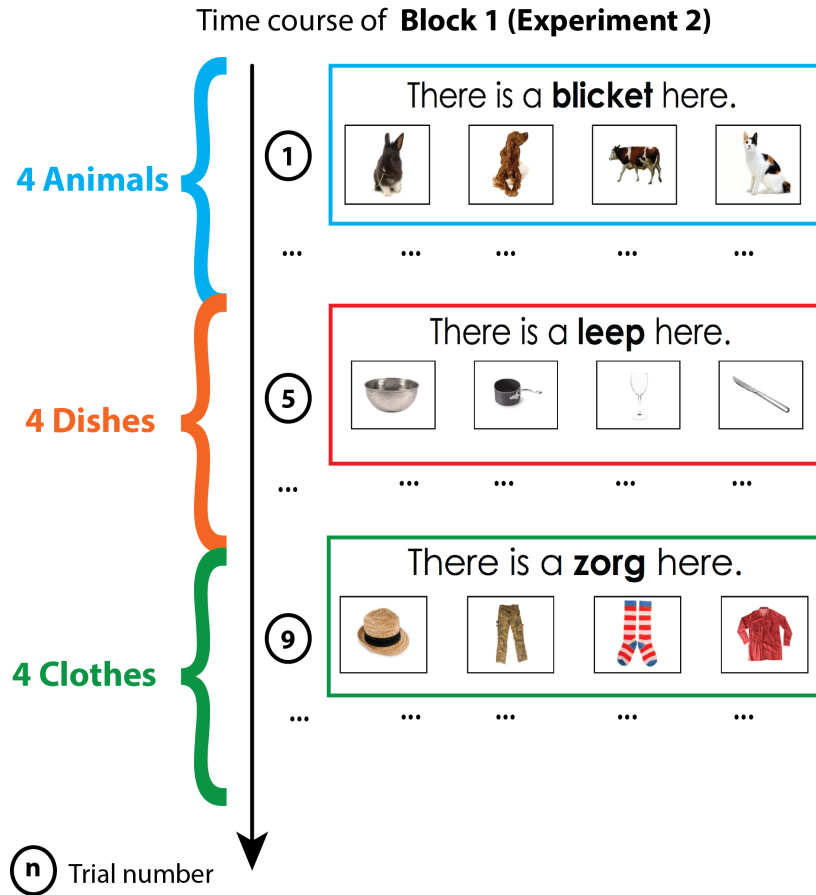


Figure 5: An example of the trial presentation in block 1 for Experiment 2. Adults saw 12 trials, one for each to-be-learned word, such that all objects in one trial were from the same natural category of the referent (animal, cloth, dish). All words referring to objects from the same natural category appeared in succession.

Procedure and analysis. The procedure and analysis are identical to those in Experiment 1.

Results

We replicated the two main results of Experiment 1. First, we modeled the accuracy with a mixed logit model using a predictor Block (1 to 5) with subjects and words as random effects on intercepts and a random slope for the effect of Block with subjects (model 1). Participants demonstrated a gradual learning of word-referent pairs across learning instances, as evidenced by a significant effect of Block on accuracy (Figure 2; $\beta = 0.39$, $z = 7.12$, $p < 0.001$). Participants were 60% ($\text{logit}^{-1}(0.39)$) more likely to be accurate than in the previous block. Second, we modeled the measure defined in Experiment 1 with subjects and words as random effects on intercepts (model 2). Participants stored more information during the first exposure of the word than expected by chance (Figure 4; 223 data points, $\beta = 1.20$, $z = 6.37$, $p < 0.001$).

We compared Experiment 1 and Experiment 2 along these two dimensions. First, we modeled participants' accuracy in block 1 and 2 for these two experiments similarly to model 1 but applied to the results of both experiments at once and with an additional predictor Experimental condition (Experiment 1, Experiment 2) and its interaction with

Block (1 vs. 2). We restricted the comparison to blocks 1 and 2 to ensure that distance or performance at or near ceiling would not mask the effect of block 1. As discussed above, we observed a significant effect of Block on accuracy. In addition, we also observed a significant interaction between Block and Experimental condition (Figure 2; $\beta = 0.43$, $z = 2.11$, $p = 0.03$). Second, we modeled our measure of information retrieval for Experiment 1 and 2 similarly to model 2 with a predictor Experimental condition (Experiment 1, Experiment 2). Our information retrieval measure shows that participants in Experiment 2 were significantly more likely than participants in Experiment 1 to resort to previously encountered, but not selected referents (Figure 4; $\beta = 0.90$, $z = 4.32$, $p < 0.001$). The probability of choosing a previously encountered referent increased by 71% ($\text{logit}^{-1}(0.90)$) in Experiment 2 compared to Experiment 1.

Discussion

The comparison between Experiment 1 and Experiment 2 shows that providing learners with an opportunity to rely on higher-order properties of situations allowed them to resort to previously encountered experience more efficiently than participants who were exposed to artificial, randomly assembled situations (Experiment 1).

As expected, richer contextual information boosted participants' use of a cross-situational learning strategy. There are three possible interpretations for this result. (1) **Context consistency and memory:** Participants used contextual information to inform their word learning strategy. We will come back to this issue in Experiment 4, but it is important to note that there are two possible explanations for such an effect. First, in a one-to-many mapping approach, temporary lexical entries may be easier to memorize if the multiple potential referents for a word are semantically coherent. Second, it could be that contextual information is stored as an independently accessible source of information: participants may memorize associations between a word *and* situations in which it was uttered, and these situations could directly inform word-meaning mappings in subsequent learning instances. (2) **A closest-match strategy:** Participants follow a hypothesis-testing strategy, but when their current hypothesis is absent from the picture display, they resort to the closest match. Concretely, if their current hypothesis is that *blicket* means *dog*, but no dog is present in the display learners would not randomly select any other possible meaning, but would rather select the closest match, which in this experiment will be another animal. (3) **Partial representations:** Participants entertain partial representations: they may encode the semantic category of a word (for example: *animal*), in the same way as they may encode the grammatical features of this word (e.g., syntactic category, gender, animacy, etc.) without encoding any meaning hypotheses. In following learning instances, participants would then (randomly) select one member of the encoded category.

Hypotheses (2) and (3) contrast with (1), as for these two options, participants would thus select a distractor from the correct category more often than chance, but this would not be mediated by memory for the previous learning situation itself. Experiment 3 disentangles between these possible interpretations of the improvement observed between Experiments 1 and 2.

Experiment 3 – Encoding context consistency

Experiment 3 was set up to replicate Experiment 2 with three *artificial* categories of objects with no *a priori* coherence (e.g., {apple, dog, flower, hat}) instead of “natural” categories. Note that despite the lack of semantic coherence among these objects, categories could nonetheless emerge here due to the repeated and consecutive co-occurrence of the four objects that constitute each of them.

Although these categories are clearly artificially induced, the process of category induction may in fact not be unnatural. Specifically, under the right circumstances many sets of objects, however unrelated they may appear to be, can co-occur. For example in the kitchen, you may simultaneously see an apple, a dog, a vase with flowers and a hat hung on the wall. These items are not transparently related but all of them may be simultaneously found in the kitchen, possibly for quite different reasons. Thus, in the absence of semantic relations between the objects of an *artificial* category in Experiment 3, the coherence may be induced by their co-occurrence on 4 consecutive trials (once for each of the word that refers to them). The consistent display here plays the role of the kitchen in the example above.

If participants fail to use this contextual consistency, then they should behave like participants in Experiment 1, where none of the objects within a trial was semantically related to the other. This would favor hypotheses (2) and (3) which attribute the improvement observed in Experiment 2 to semantic consistency. By contrast, if the artificial categories improve cross-situational learning compared to Experiment 1, this would favor hypothesis (1), which relies on consistency in general, and not on a tendency to resort to a semantically close selection (as of hypothesis 2) or on partial representations (remembering ‘*animal*’ instead of specific animals; as per hypothesis 3).

Method

Participants. Forty adults were recruited from Amazon Mechanical Turk (12 females, $M=34$ years; 39 native speakers of English). Four participants were excluded from our analysis because over 20% of their responses felt outside the 1-30 seconds response time window (See Experiment 1 – Analysis) ($n=2$) or because they participated in previous experiments ($n=2$).

Stimuli, Design. The design was similar to Experiment 2. We used a novel set of objects in order to minimize the potential semantic associations among them within each of the 3 artificial categories. Categories were defined as follows: {apple, dog, flower, hat}, {pants, chair, pan, teddy bear}, {leaf, snake, watch, book}.

The first block follows the same design as in Experiment 2 but the position on the screen for each object within the trials of the same category was fixed. For example, considering the set {apple, dog, flower, hat}, these objects appeared in the same position (albeit with different images) on the screen in all four learning instances for the four target words associated with them. This should raise the awareness that the situation is constant. Thus, a *dog* might be the left-most object for 4 consecutive trials, but the image used on each trial will change.

Procedure and analysis. The procedure and analysis are identical to those in Experiments 1 and 2.

Results

We replicated the two main results of Experiments 1 and 2. First, we modeled participants' accuracy in Experiment 3 with a mixed logit model using a predictor Block (1 to 5) with subjects and words as random effects on intercepts and a random slope for the effect of Block with subjects (model 1). Participants demonstrated a gradual learning of word-referent pairs across learning instances as evidenced by a significant effect of Block on accuracy (Figure 2; $\beta = 0.20$, $z = 3.48$, $p < 0.001$). Second, we modeled our measure of information retrieval in block 2 with subjects and words as random effects on intercepts (model 2). Participants retrieved more information from the first exposure to a word than expected by chance (Figure 4: 231 data points, $\beta = 0.68$, $z = 4.68$, $p < 0.001$).

We compared the three experiments along these two dimensions. First, we modeled participants' accuracy in block 1 and 2 for the three experiments with the predictor Block (1 vs. 2) used in model 1 and an additional predictor Experimental condition (Experiment 1, Experiment 2, Experiment 3) and its interaction with Block. There was no significant interaction between Block and Experimental condition (Experiment 3 vs. Experiment 1; $\beta = -0.20$, $z = -1.02$, $p = 0.3$) and (Experiment 3 vs. Experiment 2; $\beta = 0.21$, $z = 1.02$, $p = 0.3$, see Figure 2). Second, we modeled our information retrieval measure in block 2 for the three experiments similarly to model 2 with a predictor Experimental condition (Experiment 1, Experiment 2, Experiment 3). Participants in Experiment 3 were significantly less likely to choose a previously seen, but not selected referent than participants in Experiment 2 ($\beta = 0.47$, $z = 2.14$, $p < 0.05$), but they were significantly more likely to do so than participants in Experiment 1 ($\beta = -0.42$, $z = -2.11$, $p < 0.05$, see Figure 4).

Overall, these results demonstrate that participants in this experiment retrieved the systematic co-occurrence of seemingly unrelated objects to degree intermediate between participants in Experiments 1 and 2. This shows that participants use contextual information from consistent contexts to inform word learning, and they do so to a greater extent if contexts furthermore share a semantic relation.

Discussion

Participants used the artificial categories presented in the first learning instance to guide their choice of the word's referent in subsequent instances. Crucially, this effect was preserved although none of the objects presented in the first learning instance shared a "natural" property. This rules out the possibility that the results from the previous experiment could be due entirely to an under-specification of a selection (e.g., *animal* instead of *dog*) or to a tendency to resort to a semantically-close choice (e.g., from *dog* to *cat*) when the previously hypothesized referent was not available. Instead, our results favor the hypothesis that contextual consistency helps encoding situations in both Experiment 2 and 3.

Nonetheless, participants in Experiment 3 were less likely to resort to previously encountered referents than participants in Experiment 2. One reasonable explanation may be that encoding an artificial relation is more demanding than encoding a natural relation: while participants in Experiment 2 could remember a label readily available to

characterize the relation among objects (“animal”, “clothes” or “dishes”), participants in Experiment 3 had to encode the category as a plain list of objects. Hence, learners may have encoded contextual information in both experiments but the format of the relevant information varies from one experiment to the other and this could recruit different memory resources.

Experiment 3 showed that contextual consistency, and not only semantic consistency, helped learners resort to possible word meaning hypotheses. However this effect could be explained by two possible representations of context in memory. (a) Internal to word-meaning mappings: one-to-many word-meaning mappings may be more or less easier to remember and a coherence between the possible meanings may indirectly boost an active memory for these mappings. As a result, multiple hypotheses for a word are better remembered when these hypotheses form a coherent group, but context is not necessarily stored in memory as such. (b) External to word-meaning mappings: contextual information could be directly accessible as an independent source of information, i.e., learners could remember the situation in which they heard a word *in addition* to the single or multiple hypotheses they entertain for this word. In this case, contextual information can be used actively to constrain subsequent learning instances.

Experiment 3 did not distinguish between an internal vs. an external representation of context since contextual representation was confounded with word-meaning representations. In Experiment 4, we propose to disentangle these two possibilities and assess whether context is represented *per se* in memory.

Experiment 4 – Context representation in memory

Experiment 4 investigates whether the effect of context observed in Experiment 2 and 3 is the result of an internal or an external representation of context. Much like Experiment 2 and 3, objects in the first block of this experiment were grouped into three sets. Two of these sets contained objects from a single natural category (animals and clothes) as in Experiment 2, henceforth “natural sets”. By contrast, the third set was hybrid: it contained two (new) animals and two (new) pieces of clothes. For a word whose referent belongs to a natural set, participants could encode a natural category (e.g., animal), as in Experiment 2. However, some objects from this natural category occurred in the hybrid set and should not be considered possible referents for this word after the first learning instance (contrary to Experiment 2).

We propose to reproduce a memory illusion effect identified in earlier work (Roediger III & McDermott, 1995) showing that participants asked to remember a list of words are likely to *mis-report* a word as being part of this list if there is a natural relation between the word and the list. For example, participants incorrectly recall the word *sleep* as a member of a list such as *bed, pillow, night*. Applied to our word learning task, lists can be thought of as sets of objects seen in the first block (e.g., *dog, cat, snake, cow*). If context is encoded as an additional source of information (hypothesis b), participants are in the same situation as in the memory illusion experiment and we expect to reproduce the same illusion. Participants should be more likely to map a target word in the natural sets onto a distractor object from the appropriate natural category than from the other category. But crucially, this bias for the appropriate category should

be observed even when we compare *only* distractors from the hybrid set, which had never appeared with the target word before. If context is not encoded independently and the effect occurs at the level of the lexicon (hypothesis a), then there is no immediate expectation with respect to this illusion.

Method

Participants. 119 adults were recruited from Amazon Mechanical Turk (47 females, $M=36$ years; 116 native speakers of English). 24 participants were excluded from our analysis because they participated in previous experiments ($n=14$), because they indicated that they took notes during the task ($n = 4$) or because their RT patterns were highly irregular, in a fashion similar to participants who indicated that they took notes (e.g., 5-10 times faster from one block to another, $n = 6$).

Stimuli, Design. The design was similar to Experiment 2. We formed 2 natural sets: animals {*cat, cow, snake, rabbit*}, and clothes {*pants, tie, hat, socks*} and 1 hybrid set of images mixing objects from each natural category {*dog, rat, shirt, shoe*}. The hybrid set served as a reservoir of objects that could reveal the illusion when used as distractors. The hybrid set was always presented first.

We generated the learning trials following the constraints described in Experiment 1. However, our planned analysis focused on responses in the second block such that the target would be from one of the natural sets but responses would be a distractor from the hybrid set H. Hence, in order to have more data points of interest, we assigned the learning instance with the maximal number of distractors belonging to H (among the four learning instances otherwise distributed randomly in blocks 2 to 5) to the second block. To limit the frequency of objects from H in block 2, we did this for target objects in natural sets, but the opposite for target objects in H (which trials were not of interest). As a result, participants saw on average 5 instances of the objects in the hybrid set during block 2 (instead of 4 instances before).

Procedure and analysis. The procedure and analysis are identical to those in Experiment 1, 2 and 3.

Results

We selected learning instances from block 2 for words belonging to the two natural sets of objects. We looked at the artificial set of objects to compare the proportion of responses that belong to the set S of distractors from the same category and to the set D of distractors from a different category. Figure 6 shows the proportion of responses in S and D minus the probability of selecting them by chance (the cardinal of S and D divided by 4). Note that we selected trials where neither set S nor set D were empty (482 data points; chance level for S or D was either .25 or .50).

We modeled the proportion of responses in the artificial set of objects by a predictor Distractor type (Same category vs. Different category). Observations of the results led us to add a predictor Semantic category (Animals vs. Clothes) to the model, as well as its interaction with Distractor type. The random structure included subjects and words as random effects on intercepts and no random slope was justified. We applied an offset corresponding to chance to the model.

We observe a main effect of Distractor type ($\beta = 0.45$, $z = 2.7$, $p < 0.01$) showing that participants were 61% ($\text{logit}^{-1}(0.45)$) more likely to choose a distractor object from the same category as the target than from another category, even if this object did not co-occur with the word in the previous trial.³

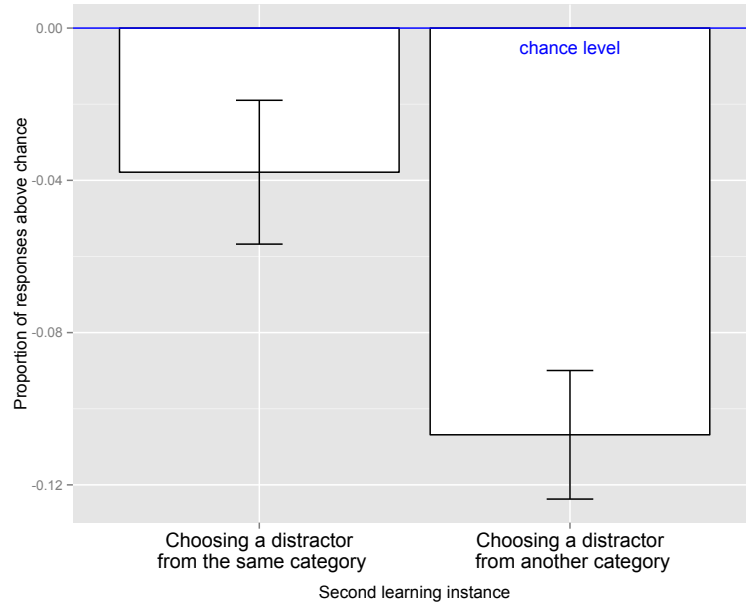


Figure 6: Experiment 5. Proportion of responses falling in the hybrid set as whether responses are from the semantic category of the target word (left bar) or from another category (right bar) minus the probability of selecting them by chance.

Discussion

Participants are more likely to select a distractor from the semantic category of the target than a possible referent from another category. Crucially, this effect occurs even though none of the distractors were present in the first learning instance for that word. This illusion is reminiscent of memory illusions observed for word lists, and thus provides an indirect argument for the fact that learning situations are stored in memory *per se*.

Our results thus suggest that a situation in which a novel word occurs can be stored and bound to this word during word learning. Others have argued that, even in adults, the information that is retrieved about a word is the accumulation of all the situations in which that word has been encountered (Perfetti & Hart, 2002). Although our results are compatible with such a proposal, they are at present restricted to cross-situational word learning stages and provide no evidence that early word representation *is* the set of learning contexts in which this word was encountered. In the general discussion, we discuss the broader implications of our results for the role and

³ Additionally there is a significant interaction between Distractor type and Semantic category ($\beta = 1.07$, $z = 3.19$, $p < 0.01$). Participants were significantly more likely to choose a distractor from the same type as the target for words referring to clothes than for words referring to animals. This could be due to the fact that in this task the memory illusion may be stronger for one category than for the other (e.g., because the animal category may be more salient than the cloth category making it more subject to illusions).

representation of contextual information during the development of lexical representations.

General Discussion

The present paper examined the impact of the context on word learning mechanisms. In four experiments, we showed that learners can simultaneously retrieve multiple candidates for the meaning of a word and that manipulating the contextual properties of the set of plausible candidates could boost the amount of information retrieved. Specifically, our results show that cross-situational learning benefits from higher-order properties of a word-learning situation: the semantic relation between the possible referents (Experiment 2) as well as contextual consistency (Experiment 3). Moreover, this effect is subject to memory illusions, in a way that suggests that the effect of context found above is the result of an attempt to store contextual information directly in memory (Experiment 4).

Learning strategies

Most of the accounts of cross-situational learning have concentrated on the amount of information the learner stores for each learning instance. We introduced two learning strategies at the opposite end of the continuum: an accumulative learning account, in which the learner encodes one-to-many word-meaning mappings, and a hypothesis-testing account, in which the learner remembers a single word-meaning association. While computational models have emphasized the importance of defining the number of hypotheses entertained at each point in time (Yu & Smith, 2012), we add a new parameter showing that learners could also encode a different kind of information, context, to increase the amount of prior experience they could retrieve. Our results argue against an extreme version of the hypothesis-testing account where learning operates *only* through a single hypothesis for each word. Instead we suggest that cross-situational learning is informed by the type of learning context.

One may imagine other learning strategies in more intermediate continuum positions to accommodate the finding that learners encode more than a single meaning hypothesis. For instance, Koehne, Trueswell and Gleitman (2013) proposed multiple-hypothesis tracking strategy, according to which learners may memorize not only one hypothesis, but all past hypotheses for a given word. Previous research on cross-situational learning has also suggested that learners do not attend equally to all possible meanings for a word and use several additional strategies to prune the set of possible meanings (mutual Exclusivity: Yurovsky & Yu, 2008; attention to stronger associations: Yu & Smith, 2012).

Overall, investigations about word learning strategies concentrated on the possible forms of relations learners could entertain between a word and possible referents. Here we propose that some contextual information is memorized and can boost word learning in realistic situations.

Implications for learning words in the real world

Learners relied on previously experienced information more efficiently when this information was packaged conveniently. That is, cross-situational learning was improved not only by a natural relation between possible referents (Experiment 2), but

also by an artificial relation between objects solely induced by their repetitively joint presentation (Experiment 3). Of course, real life situations are much more complex learning environments than the situations in the word-learning paradigm we used in Experiment 1 (Medina et al., 2011): here the level of referential ambiguity is relatively low (four possible referents), only one word is presented at a time, and the true referent is always present in all word occurrences. Further simplification of the task may hence seem inappropriate. However, the specific simplifications we introduced in Experiments 2 and 3 in fact make the task more ecologically valid. In daily life, learners navigate through situations they may be interested in and find coherent. This could help them remember various properties of these situations (a kitchen, a zoo, a pantry, etc.). In Experiments 2 and 3 we introduced such coherence and showed that it has a specific impact on their strategy and performance for learning new words.

Interestingly, a recent computational approach looking at environment regularities showed that coherent activity contexts such as eating, bathing or other regular activities could help simplify the learning problem (Roy, Frank & Roy, 2012). Our results align with this view, showing that more complex information from the broader context in which a word has been uttered is part of the learning problem faced by the child. The role of the learning environment on word learning requires attention in future research.

The representation of lexical meaning during learning

One important issue in the acquisition of word meaning involves the kind of representations children form about words. In other words, what do learners encode about a word when they first hear it? The full understanding of a word requires that learners know its word form, its meaning, its syntactic properties but also information about contexts in which this word may occur. Recent evidence has shown that even infants in the first year of life have already acquired some knowledge for basics words (Bergelson & Swingley, 2012; Bergelson & Swingley, 2013). However, there is growing evidence that children do not fast-map a dictionary-like definition at the first encounter of the word. Instead, word learning, including verb learning, seems to be a slow process gradually emerging through the accumulation of syntactic, semantic and pragmatic fragmental evidence (Bion, Borovsky & Fernald, 2013; Gelman & Brandone, 2010; Yuan & Fisher, 2009). However it is currently unclear what this partial knowledge might be. The present results suggest that, alongside linguistic features (e.g., phonological form, syntactic category), non-linguistic features such as semantic category (Experiment 2) and situations in which the word occurred (Experiments 3 and 4) may be encoded and part of an early word representation. Non-linguistic relations between words are a crucial component of the organization of the lexicon. Work on lexical priming has evidenced that young 21-month-olds already possess a structured knowledge of familiar words based on non-linguistic information such as semantic and associative relations (Arias-Trejo & Plunkett, 2013). As models of lexical development suggest (Steyvers & Tenenbaum, 2005), such a semantic organization of the lexicon is the product of the mechanisms by which word-meaning associations are constructed throughout learning. This suggests that semantic and contextual relations may be encoded from the earliest step of lexical acquisition (see Wojcik & Saffran, 2013 for evidence that toddlers can encode similarities among referents when learning words).

However, to our knowledge, no cross-situational study investigated the role of the learning context in word learning. Such studies could shape our understanding of early word representation but also shed light on the content and structure of adults' mature lexical entries.

Summary

Overall our findings suggest that learners store in memory the learning situation in which they hear a novel word and use this information to constrain their word-meaning hypotheses. We first proposed a new way to analyze classical word learning experiments through an information retrieval measure. We then modified the classical word learning paradigm to evaluate whether realistic features of the world could inform word learning strategies. Our results show that prior experience is better used when it consists of coherent contexts, and real world situations may well be coherent contexts in the relevant sense. We conclude that such paradigms, however simple, could and should be used to further study the structure, richness and poverty of the representations that constitute the early developing lexicon.

References

- Akhtar, N., & Montague, L. (1999). Early lexical acquisition: The role of cross-situational learning. *First Language*, 19(57), 347–358.
- Arias-Trejo, N., & Plunkett, K. (2013). What's in a link: Associative and taxonomic priming effects in the infant lexicon. *Cognition*, 128(2), 214–227.
- Baayen, R.H., Davidson, D.J. and Bates, D.M. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 390-412.
- Bates, D., & Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes. R package version 0.99875-6.
- Bergelson, E., & Swingley, D. (2012). At 6 to 9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the USA*, 109, 3253-3258.
- Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition*, 127, 391-397.
- Bion, R. A. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, 126(1), 39–53.
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference*, 15, 17–29.
- Gelman, S. A., & Brandone, A. C. (2010). Fast-mapping placeholders: Using words to talk about kinds. *Language Learning and Development*, 6(3), 223–240.
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 66, 325–331.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.
- Koehne, J., Trueswell, J.C., & Gleitman, L.R. (2013). 'Multiple Proposal Memory in Observational Word Learning.' Proceedings of the 35th Annual meeting of the Cognitive Science Society. Austin, TX: Cognitive Science Society.
- Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, 108(22), 9014–9019.
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. The

- MIT Press.
- Perfetti, C.A., & Hart, L. (2002). The lexical quality hypothesis. In L. Vehoeven. C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam/Philadelphia: John Benjamins.
- Quine, W. V. O. (1964). *Word and object* (Vol. 4). MIT press.
- Roediger, H.L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24(4), 803–814.
- Roy, B. C., Frank, M. C., & Roy, D. (2012). Relating activity contexts to early word learning in dense longitudinal data. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*.
- Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61(1), 39–91.
- Smith, K., Smith, A. D. M., & Blythe, R. A. (2011). Cross-Situational Learning: An Experimental Study of Word-Learning Mechanisms. *Cognitive Science*, 35(3), 480–498.
- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106(3), 1558–1568.
- Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic bulletin & review*, 8(2), 203–220.
- Steyvers, M., & Tenenbaum, J. B. (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive science*, 29(1), 41–78.
- Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1536), 3617–3632.
- Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, 66(1)
- Vlach, H. A., & Sandhofer, C. M. (2011). Developmental differences in children’s context-dependent word learning. *Journal of Experimental Child Psychology*, 108(2), 394–401.
- Vouloumanos, A., & Werker, J. F. (2009). Infants’ learning of novel words in a stochastic environment. *Developmental Psychology*, 45(6), 1611–1617.
- Wickham, H. (2009) ggplot2: elegant graphics for data analysis. Springer New York.
- Wojcik, E. H., & Saffran, J. R. (2013). The Ontogeny of Lexical Networks: Toddlers Encode the Relationships Among Referents When Learning Novel Words. *Psychological Science*.
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5), 414–420.
- Yu, C., & Smith, L. B. (2012). Modeling cross-situational word–referent learning: Prior questions. *Psychological Review; Psychological Review*, 119(1), 21.
- Yuan, S., & Fisher, C. (2009). “Really? She Blinked the Baby?” Two-Year-Olds Learn Combinatorial Facts About Verbs by Listening. *Psychological Science*, 20(5), 619–626..
- Yurovsky, D., & Yu, C. (2008). Mutual exclusivity in cross-situational statistical learning. In *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 715–720).