Meaning Matters: When Does the Quality of Retrieved Contextual Information Influence the Feeling of Knowing?

A thesis

submitted by

Stacey J. Dubois

In partial fulfillment of the requirements
for the degree of

Master of Science

in

Psychology

TUFTS UNIVERSITY

May 2011

Adviser: Ayanna K. Thomas

Abstract

Research has demonstrated that the amount of accessible information related to an unrecalled target affects feeling-of-knowing (FOK) judgments (Koriat, 1993). In some situations, FOK judgment magnitude is not only related to the amount (quantity) but also the correctness (quality) of retrieved contextual information (e.g. Thomas, Bulevich, & Dubois, 2010). The present study examined the conditions under which the correctness of contextual information influences FOKs. We hypothesized that both quantity and quality of retrieved contextual information would influence FOK judgments in situations where to-be-remembered stimuli were inherently meaningful and where meaningful contextual information was retrieved. In three experiments, we varied meaningfulness of to-be-remembered items both intrinsically and extrinsically. In Experiments 1 (word pairs) and 2 (picture pairs), the quality of retrieved contextual information influenced FOK judgments when semantic attributes were retrieved. However, the quality of retrieved contextual information did not influence mean FOKs when participants encoded inherently meaningless stimuli (Exp. 3).

Keywords: Feeling of knowing (FOK), accessibility, partial information, context, meaning, evaluation

Acknowledgments

A special thanks to my adviser, Ayanna K. Thomas, for her guidance and support as I conducted this research, as well as for her comments and suggestions during my preparation of the manuscript. I would also like to thank the other two members of my Master's committee, John Bulevich and Richard Chechile, for their insights and advice.

Table of Contents

Meaning Matters: When Does the Quality of Retrieved Contextual Information Influence the Feeling of Knowing?

In standard feeling-of-knowing (FOK) paradigms, participants are asked to predict the likelihood of future recognition for items that are currently unrecallable. Though FOK judgments are far from perfect, studies have shown that they are above-chance predictors of future explicit memory performance (e.g. Hart, 1965, 1967; Blake, 1973; Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982; Schacter, 1983; Metcalfe, 1986; Koriat, 1993, 1995). Despite this finding, numerous studies have suggested that participants do not have direct access to the target item while making FOK judgments. Rather, feeling of knowing is inferred from a variety of cues, such as knowledge about the general topic (e.g. Glenberg, Sanocki, Epstein, & Morris, 1987; Costermans, Lories, & Ansay, 1992; Maki & Serra, 1992), familiarity with the cue or parts of the question (e.g. Reder & Ritter, 1992; Schwartz & Metcalfe, 1992; Metcalfe, Schwartz, & Joaquim, 1993; Koriat & Levy-Sadot, 2001), and the retrieval of information related to the target (e.g. Blake, 1973; Eysenck, 1979; Schacter & Worling, 1985; Koriat, 1993).

According to the accessibility heuristic (Koriat, 1993), FOK judgments are driven by an inferential process in which estimates of future recognition are influenced by the amount of relevant partial information (e.g. first letter of the target item, the number of syllables) generated during the memory search. Research has demonstrated that FOK judgments are positively correlated with the amount of retrieved partial information (letters recalled or target valence), regardless of whether or not the partial information is accurate (Koriat, 1993; see also Koriat, 1995; Koriat & Levy-Sadot, 2001). The primary conclusion drawn from this body of work is that FOKs are influenced by the *quantity* of accessible partial information, and the quality, or correctness, of that information is irrelevant.

However, recent studies have demonstrated that correctness of information associated with an unrecallable target does influence both the magnitude and accuracy of metacognitive judgments. For example, Dunlosky, Rawson, and Middleton (2005) found that the correctness of information generated during pre-judgment recall attempts influenced subsequent term-specific judgments. In this study, participants in the “pre-judgment recall” condition attempted to define four key terms before predicting the likelihood of future definition recall, and these metacognitive judgments were higher when participants had retrieved accurate partial information. In addition, Thomas, Bulevich, and Dubois (2010) found that younger adults provided higher FOK judgments when they retrieved the correct valence of the target item, as opposed to an incorrect target valence. Further, Thomas et al. (2010) demonstrated that FOK accuracy also improved when participants had an unlimited amount of time with which to make FOK judgments. Thomas et al. hypothesized that participants may use that time to evaluate the partial information and search for additional relevant information.

In light of these findings, we ask not *if* but *when* does the quality of retrieved partial information influence the feeling of knowing? We will argue that this depends on whether or not meaningful information—features and attributes that are central to the item's conceptual representation in memory—is involved in the process. Both Dunlosky et al. (2005) and Thomas et al. (2010) used inherently meaningful stimuli and solicited meaningful (semantic/affective) partial information, while the to-be-remembered stimuli in Koriat's (1993) first three experiments were meaningless strings of letters (e.g. FKDR) and partial information was defined as the number of letters recalled. It is possible that an item's meaningfulness is tied closely to metacognitive monitoring and accuracy for that item, but no work has addressed the issue of whether information devoid of conceptual meaning is associated with impaired metacognition. Stimuli meaningfulness might affect whether the inferential feeling-of-knowing process moves beyond accessibility to a later evaluation stage based on the quality, or accuracy, of the retrieved partial information.

In the present study, we sought to examine the specific conditions under which the quality of retrieved partial information might affect FOK judgment magnitude. We suggest that the quality of retrieved partial information is likely to matter only in situations where *both* the to-be-remembered stimuli and the retrieved partial information are meaningful. Across three experiments, we varied the availability of meaningful contextual details in two ways: by using stimuli that varied in inherent meaningfulness (intrinsic manipulation) and by using levels-of-processing instructions at encoding (extrinsic manipulation). Participants encoded concrete word pairs (Exp. 1), nameable picture pairs (Exp. 2), or ambiguous picture pairs (Exp. 3) incidentally by answering a perceptual (shallow) or semantic (deep) orientation question for each pair. During the cued recall/FOK phase, participants had a chance to provide both perceptual and semantic contextual information related to the target item (we will hereafter be using 'contextual' information in lieu of 'partial' information, in order to describe more accurately the broad range of details and attributes that come to mind during a retrieval attempt). Therefore, we were able to examine the effects of both intrinsically- and extrinsically-instantiated conceptual processing of the to-be-remembered items, as well as the influence of the meaningfulness of retrieved contextual information, on FOK judgment magnitude.

Intrinsic vs. Extrinsic Meaning

Koriat's (1997) cue-utilization framework classifies metamemorial cues into three categories: intrinsic, extrinsic, and mnemonic. Intrinsic cues pertain to properties of the to-be-remembered stimuli itself, such as word frequency, semantic relatedness, and conceptual meaning. Extrinsic cues refer to the conditions of learning (e.g. presentation rate) and/or the operations employed by the participant during encoding (e.g. levels of processing manipulations). Mnemonic cues are idiosyncratic and experience-based; this class refers to cue familiarity and accessibility of relevant contextual information, as well as other cues such as memory for previous recall attempts and fluency of processing at encoding. Koriat suggested that intrinsic properties of items are often given priority over extrinsic conditions of learning when judgments of learning (JOLs) are made (Koriat, 1997). We varied the intrinsic meaningfulness of our stimuli; participants encoded concrete word pairs in Experiment 1 and nameable picture pairs in Experiment 2 (both inherently meaningful), while participants in Experiment 3 encoded ambiguous figure pairs (inherently meaningless).

According to Blake (1973), retrieved contextual information is only useful in predicting future recognition (i.e. formulating accurate FOK judgments) when it comprises attributes that help to distinguish the correct response from the distractor items. Research that suggests that participants have access to individual features of a memory trace even when complete recall fails. Specifically, Koriat, Levy-Sadot, Edry, and de Marcas (2003) found that participants could remember the semantic attributes of words that they could not recall; further, the phenomenological experience associated with this access was diagnostic of the accuracy of the partial information. Koriat et al. (2003) asked participants to provide judgments for the target item on three semantic-differential dimensions (good-bad, strong-weak, active-passive) when they failed to recall the item or made a commission error. Results indicated that participants were successful in judging the polarity for each of the three dimensions, suggesting that they had access to the semantic attributes of the unrecalled target word. We argue that meaningful stimuli are composed of a number of defining features that can be recalled, even when complete recall fails, and that those features should influence FOK magnitude. The finding that partial recall is possible even when whole item recall fails can be taken as support for feature-based models of semantic memory. Such models posit that the meaning of words is represented in memory as a list of features, some of which are essential to the word's meaning (defining features) (Smith, Shoben, & Rips, 1974; Koriat et al., 2003). A logical hypothesis, then, would be that semantic features are useful pieces of contextual information when it comes to discriminating between targets and lures, because they are, in essence, fragments of the item's conceptual representation in memory. The intrinsic meaningfulness of to-be-remembered items affects the likelihood that such attributes can be recalled after a retrieval failure, and the type of attribute recalled, in turn, may be an important factor in whether the accuracy of such contextual information influences FOK judgments.

In addition to being inherently meaningless, the nonwords used in Koriat's (1993) study might have evoked more perceptual (shallower) processing than the English word pairs used in the Thomas et al. (2010) study, and this might have affected the relevance of contextual information quality to FOK judgments. Therefore, we also instantiated meaning through an extrinsic levels-of-processing manipulation at encoding. Deeper levels of processing result in better recall (e.g. Craik & Tulving, 1975), which implies that a richer, more detailed memory trace is formed under deep, as opposed to shallow, encoding conditions. We therefore hypothesized that semantic debris (contextual information) would come to mind more often when items had been encoded with a focus on meaning as opposed to visual characteristics. In addition, we hypothesized that participants' FOK judgments would be sensitive to the extrinsic manipulation, because research has demonstrated that the depth at which information is processed at encoding affects participants' metacognitive judgments (Lupker, Harbluk, and Patrick, 1991). Specifically, Lupker et al. (1991) demonstrated that participants in a deep encoding condition not only performed better on the initial cued recall task than participants in the shallow encoding condition, but they also assigned higher FOK judgments to items in the deep condition as compared to ones in the shallow condition. Participants in the deep encoding condition also showed better relative prediction accuracy (as measured by gamma correlations) than participants in the shallow encoding condition—that is, the FOK judgments made by participants who had encoded the word pairs with sentence generation were more predictive of final task performance (explicit cued stem completions) than participants who had encoded the word pairs by counting vowels (Lupker et al., 1991).

The Present Study

In Experiments 1 and 2, we used inherently meaningful stimuli and predicted that FOK judgment magnitude would be influenced by the quality of retrieved contextual information only when that information is semantic in nature. We examined our hypothesis by (1) putting participants in a situation where they would be more likely to retrieve semantic debris (deep encoding) and (2) soliciting multiple types of contextual information (semantic and perceptual). This is the first study in the FOK literature to solicit more than one attribute, allowing us to compare the unique contributions of each type of attribute to FOK judgments. In addition, soliciting multiple attributes enabled us to assess whether the effects of quantity found in previous studies (e.g. Koriat, 1993) can best be explained by a “some-vs.-none” monitoring process or an “additive” monitoring process. For example, the feeling of knowing may be higher when any contextual information is retrieved as opposed to no contextual information, or it is possible that FOK judgments will increase along with the number of attributes retrieved (our predictions align with the latter possibility).

While Koriat (1993) found that the amount of retrieved partial information influenced FOK judgments, the partial information provided by participants in his first three experiments were actual pieces of the to-be-remembered stimuli (letters of the letter strings). In contrast, we solicited relevant contextual information—features related to the target item—that did not include fragments of the target itself. For Experiments 1 and 2, we hypothesized that the quality of retrieved contextual information would influence FOK judgments only when that information was meaningful—i.e. when the contextual information consisted of semantic attributes inherently tied to the target item as opposed to perceptual attributes arbitrarily manipulated between items.

General Method

Participants

We collected data from 142 undergraduates (ages 17-23) in this study (46 in Experiment 1, 31 in Experiment 2, and 65 in Experiment 3). Participants were recruited through the student pool maintained by Tufts University or volunteered by responding to an advertisement posted on the TuftsLife website, and they received partial course credit or $10 per hour as compensation for their participation, respectively. Only individuals with normal color vision were eligible to participate.

Materials

Computer programs were coded and run using E-Prime v.1.1 software (Psychology Software Tools Inc., Pittsburg, PA), and participants completed the feeling-of-knowing (FOK) task on Dell desktop computers. We presented participants with word pairs in Experiment 1, concrete picture pairs in Experiment 2, and ambiguous figure pairs in Experiment 3 (these stimuli sets will be described in more detail when we address the specifics of each experiment's methodology). The following section details the general procedure for Experiments 1 and 2; we made significant changes to the methodology for Experiment 3, and these will be described later.

Procedure: Experiments 1 and 2

Participants provided informed consent prior to participating in the experiment. The study description indicated that we were investigating how information processing differences affect both the way information is stored and how accessible that information becomes for later use. Up to four participants completed the task simultaneously in a lab with four available computers. Each experiment consisted of four main phases: encoding, retention interval, cued recall/FOK judgments, and recognition. Participants read the instructions for each phase on the computer, and the experimenter provided clarifications when necessary.

We manipulated levels of processing at encoding (as defined by Craik & Lockhart, 1972) within participants. During the encoding phase, participants viewed word pairs (Exp. 1) or picture pairs (Exp. 2) one pair at a time on the computer screen, and they answered an orienting question for each pair. In each experiment, one orienting question promoted "shallow" processing of the stimuli, while the other promoted "deep" processing of the stimuli (Craik & Lockhart, 1972). We did not explicitly instruct participants to remember any of the word pairs for a future test of memory; rather, they simply answered one question (shallow or deep) per word pair, and therefore the stimuli were encoded incidentally.

We divided word/picture pairs into two equivalent blocks of items within each experiment, and encoding depth for each block was counterbalanced across participants. Presentation of the stimuli pairs was randomized during the encoding phase. On each trial, participants viewed an encoding question for three seconds, and then a cue-target pair appeared on the screen. After a 1-second delay, a response box appeared for participants to answer the encoding question. As soon as the participant provided a response, the program moved to the next trial. After participants completed all of the trials in the encoding phase, they played Tetris on the computer for five minutes.

The cued recall/feeling-of-knowing (FOK) phase of the experiment followed the retention interval. In this phase, we presented participants with only the cues from each cue-target pair and asked them to produce the target by typing a word into the response box. Once eight seconds had elapsed without a response from the participant, or once a response had been provided, we gave participants an opportunity to provide contextual information relating to the target. The two contextual information questions were designed to correspond to the two encoding depths—one question solicited perceptual (shallow) contextual information and the other solicited semantic (deep) contextual information. Participants could respond to both questions for each target item, though they were told to select the "abstain" response when they did not have a particular feeling about the attribute in question. Presentation order for the contextual information questions was counterbalanced across participants. After responding to both questions, participants rated their feeling of knowing for the target item—that is, we asked participants to predict their chances of recognizing the target out of four choices when presented with the cue in a later phase of the experiment. Responses to the FOK question ("What are your chances of recognizing the correct target?") were provided on an 11-point Likert scale (0 = *I definitely will NOT be able to recognize the target word/image*; 10 = *I definitely WILL be able to recognize the target word/image*).

Participants then completed a forced choice recognition test. Participants viewed the cues from each stimuli pair (presented in a random order) and selected the corresponding target item from four alternatives. Participants responded using the number keys on the keyboard. Following the experiment, participants were debriefed and thanked for their participation.

Experiment 1

Method

Materials. We generated 60 unrelated cue-target pairs from the normed category word lists of Van Overschelde, Rawson, and Dunlosky (2004). We selected 12 target items from each of the following five categories: a four-footed animal, an article of furniture, a part of the human body, a fruit, and an article of clothing. We selected 20 cue words from these same five categories, and we selected the remaining 40 cue words from the following categories: a kitchen utensil, a carpenter's tool, a natural earth formation, a sport, a weather phenomenon, a musical instrument, a transportation vehicle, a flower, and a gardener's tool (Van Overschelde et al., 2004).

We examined the English-language frequencies of the cues and targets with The English Lexicon Project's website (Balota et al., 2007, http://elexicon.wustl.edu/default.asp). By comparing logs of the HAL study frequencies, which ranged in value from 0 (low frequency) to 17 (high frequency) (Lund & Burgess, 1996), we determined that target items across categories had comparable, mid-range frequencies ($M_{total} = 8.36$). However, we were unable to obtain frequency data for three words in the furniture category ('loveseat,' 'futon,' and 'recliner'). We also determined that cues had mid-range frequencies comparable to those of the target items ($M_{total} = 8.28$).

We constructed 60 cue-target pairs (Appendix A) consisting of 20 "same-category" pairs (e.g. Kiwi-Lemon) and 40 "different-category" pairs (e.g. Plateau-Lime). The same-category pairs included four pairs from each of the five target categories, and the different-category pairs included eight targets from each of the five target categories. We used the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998) to examine forward and backward association strength, and we determined that all of the cue-target pairs were unrelated.

We constructed the recognition test by pulling distractor items from the list of 120 cues and targets. There were a few constraints, however: each cue and target item appeared as an answer choice exactly twice, the four alternatives for any given cue were unique in their category membership (e.g. only one option could be a fruit, or an animal, etc), and each set of alternatives always contained at least one red and one blue word. Once again, we used the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998) to examine forward and backward association strength between the cues and distractor items, as well as between the targets and distractor items, to ensure that all possible word combinations formed unrelated pairs.

Procedure. The design of Experiment 1 was a 2 (Encoding Orientation: shallow vs. deep) x 2 (Contextual Information Question: font color vs. category) within-subject factorial.

The procedure was carried out as described in the General Methods section, with the following particulars. During the encoding phase, we presented participants with sixty word pairs, one at a time, and asked them to answer one of the following orienting questions for each pair: “Are these words written in the same font type?” or “Which word do you find more pleasant?” The font type question was designed to promote “shallow” processing of the word pairs, while the pleasantness question was designed to promote “deep” processing of the word pairs (Craik & Lockhart, 1972). We presented cues and targets in one of four distinct font styles (“42,” “Joint by Pizza Dude,” “SF Minced Meat,” and “Prodotto in Cina”; all were downloaded from the internet), and each word was written in either red or blue font. The various combinations of font style, font color, pair type (same-category vs. different-category), and target category were balanced across the two stimuli blocks (which each contained 30 pairs).

During the cued recall/FOK phase, cues were presented in the same font style and font color as seen during encoding. The perceptual (shallow) contextual information question was, “What was the font color of the target?” and the semantic (deep) contextual information question was, “Was the target in the same category as the cue?” For the shallow question, participants could respond “red,” “blue,” or “abstain,” and for the deep question, participants could respond “yes,” “no,” or “abstain.” Finally, during the recognition test, all cues, targets, and distractor items were presented in the same font style and font color in which they were presented during the encoding phase.

Results

Alpha was set at 0.05 for all statistical analyses. Items for which participants produced the correct response during the cued recall phase of the experiment (5.65%) were excluded from the following analyses because FOK judgments for these items virtually always equaled 10 (Figure 2) and the items were recognized with very high accuracy on the final test. Overall recognition was 0.42, which is significantly greater than chance recognition of 0.25 [$t(45) = 10.88$]. Thus, the task difficulty was optimal because it maximized the number of observations we could include for the FOK analyses (see Figure 1 for the overall distribution of FOK judgments) and still resulted in above-chance performance on the final recognition test.

Encoding Orientation. We calculated two feeling-of-knowing (FOK) judgment means for each participant, one for items that had been encoded deeply and the other for items that had been encoded shallowly. An analysis of encoding depth revealed that, across participants, significantly higher FOK judgments were given to items that had been encoded deeply (M = 4.81) than items that had been encoded shallowly (M = 2.70) [$t(45) = 11.75$, $d = 1.36$]. For recognition scores, items that had been encoded deeply were recognized at significantly higher rates (M = 0.53) than items that had been encoded shallowly (M = 0.32), [$t(45) = 8.12$, $d = 1.50$], thus demonstrating a standard levels-of-processing effect (e.g. Craik & Tulving, 1975). It should be noted that, although the mean for shallowly encoded items was low (M = 0.32), recognition performance for those items was still significantly greater than chance (0.25) [$t(45) = 3.62$].

We computed gamma correlations for each participant in order to examine the relative prediction accuracy of their feeling-of-knowing judgments. Gamma correlations are the statistical measurement of choice in metacognitive research for analyzing the relationship between predictions or postdictions and criterion performance (Nelson, 1984). We computed two gamma correlations (which can range in value from -1 to 1) for each participant, one for deeply encoded items and one for shallowly encoded items. A paired samples t-test indicated that participants were significantly better at predicting future recognition performance for deeply encoded items ($M_\gamma = 0.42$, $N = 45$) than shallowly encoded items ($M_\gamma = 0.07$, $N = 45$) [$t(44) = 4.17$, $d = 0.97$]. The gamma correlation for items encoded shallowly ($M_\gamma = 0.07$) did not differ significantly from zero [$t(44) = 1.30$, p >

0.05, *NS*]. Thus, depth of processing appears to influence not only subjective assessments and objective performance, but also the relative predictive accuracy of feeling-of-knowing judgments.

Quantity of Retrieved Contextual Information. To examine the effects of contextual information *quantity* in the present study, FOK judgment means were computed for three groups of items: (a) items for which participants chose the "abstain" response for both contextual information questions (font color and category), (b) items for which participants answered one of the contextual information questions and abstained for the other, and (c) items for which participants answered both contextual information questions. For each of these groups, two means were calculated for each participant—one for items encoded shallowly and one for items encoded deeply. Accuracy of the retrieved contextual information was irrelevant to this analysis.

A repeated measures 2 (Encoding Orientation: shallow, deep) x 3 (Contextual Information Quantity: abstained for both, answered one, answered both) ANOVA was conducted on participants' mean FOK judgments (Figure 3). There were significant main effects of both Encoding Orientation [$F(1, 36) = 72.32$] and Contextual Information Quantity [$F(2, 72) = 160.10$]. Mean FOK judgments were higher for items that had been encoded deeply ($M = 4.33$) as opposed to shallowly ($M = 3.09$). Mean FOK judgments also increased as a function of the amount of contextual information retrieved. Planned comparisons revealed that mean FOK judgments were lower when participants abstained for both contextual information questions ($M = 1.83$) than when they answered one contextual information question (*M* = 3.90) [*t*(44) = 10.38, *d* = 1.34]. FOK judgments were higher when participants provided answers for both contextual information questions (*M* = 5.43) as opposed to just one contextual information question [*t*(44) = 6.99, *d* = 1.48]. There was also a significant Encoding Orientation x CI Quantity interaction [*F*(2, 72) = 10.34], indicating that the effect of encoding was smaller when participants abstained for both contextual information questions than when they either answered one question or answered both questions.

To examine the effects of the quantity of retrieved contextual information on objective performance, a repeated measures 2 (Encoding Orientation: shallow, deep) x 3 (Contextual Information Quantity: abstained for both, answered one, answered both) ANOVA was conducted on participants' mean recognition scores. There were significant main effects of both Encoding Orientation [*F*(1, 36) = 29.12] and Contextual Information Quantity [*F*(2, 72) = 5.96]. Mean recognition scores were higher for items that had been encoded deeply (*M* = 0.53) as opposed to shallowly (*M* = 0.34). Mean recognition scores were also better when participants retrieved some contextual information as opposed to none. Planned comparisons revealed that mean recognition scores were lower when participants abstained for both contextual information questions (*M* = 0.36) than when they answered one contextual information question (*M* = 0.47) [*t*(44) = 3.27, *d* = 0.22]. However, there was no difference in mean recognition scores when participants provided answers for both contextual information questions (*M* = 0.48) as opposed to just one contextual information question [$t(44) = 0.48$, $p > 0.05$, *NS*]. There were no other significant effects.

Quality of Retrieved Contextual Information. In order to examine the effects of contextual information *quality* (accuracy) on participants' mean FOK judgments, we confined our analyses to the trials for which participants provided a response for *both* contextual information questions (i.e. they did not abstain). After scoring the retrieved contextual information for accuracy, we computed FOK means by participant for each of the following groups: only font color correct (shallow, deep), only category correct (shallow, deep), and both questions correct (shallow, deep). When a mean could not be computed, the mean for that group was substituted for the missing data, as recommended by Cohen, Cohen, West, and Aiken (2003) (Table E1). It should be noted that, because this analysis was based on accurately retrieved contextual information, the number of observations that were used to generate FOK means differed as a function of condition. On average, there were fewer observations when participants only answered the font color question correctly ($M_{shallow} = 3.88$; $M_{deep} = 3.00$). More responses were given when participants answered the category question correctly ($M_{shallow} = 5.66$; $M_{deep} = 10.53$). One participant was excluded from this analysis entirely because there were no trials for which this participant answered both contextual information questions.

We conducted a 2 (Encoding Orientation: shallow, deep) x 3 (Question Accuracy: only font color correct, only category correct, both correct) repeated measures ANOVA on participants' mean FOK judgments (Table E1). There were significant main effects of both Encoding Orientation [$F(1, 44) = 74.87$] and Question Accuracy [$F(2, 88) = 31.52$]. Mean FOK judgments were higher for items that had been encoded deeply ($M = 6.13$) as opposed to shallowly ($M = 4.64$). Planned comparisons revealed that mean FOK judgments were lower when only the font color question was correct ($M = 4.54$) as opposed to when only the category question was correct ($M = 5.49$) [$t(44) = 5.29$, $d = 1.22$]. However, mean FOK judgments were at their highest when both contextual information questions were correct ($M = 6.14$) [category only vs. both correct: $t(44) = 3.06$, d = 1,.41]. There were no other significant effects.

Secondary Analysis for the Quality of Retrieved Contextual Information. Koriat (1993) noted that when partial recall is solicited (such as in feeling-of-knowing and tip-of-the-tongue studies), data analysis must sometimes depart from traditional statistical methods. This is known as the 'fragmentary data problem,' and Brown and McNeill (1966) state that "the best thing to do with fragmentary data is to report them very fully and analyze them in several different ways" (p. 328). This is what we aim to do here. We face the 'fragmentary data problem' for all of our experiments, and, although we were able to conduct the more 'traditional' quality analysis provided above, we also conducted an analysis where we collapsed the data across participants (similar to Brown and McNeill, 1966) in order to provide the most complete picture of our data as possible.

For the instances in which participants answered only one contextual information question, a 2 (Encoding Orientation: shallow, deep) x 2 (Contextual Information Question: font color, category) x 2 (Question Accuracy: correct, incorrect) ANOVA was conducted using mean FOK ratings as the dependent variable. There was a significant main effect of Encoding Orientation [$F(1, 644) = 27.00$, $p < 0.001$], indicating that participants gave significantly higher FOK ratings to items that had been encoded deeply ($M = 5.03$) as opposed to shallowly ($M = 3.51$). There was also a significant main effect of Question type [$F(1, 644) = 26.79$, $p < 0.001$], indicating that participants gave significantly higher FOK ratings when answering the category contextual information question ($M = 4.86$) than the font color contextual information question ($M = 3.33$). The main effect of Accuracy approached significance [$F(1, 644) = 3.77$, $p = 0.053$], suggesting that participants trended toward giving higher FOK ratings when they answered either contextual information question correctly ($M = 4.61$) as opposed to incorrectly ($M = 3.71$).

The omnibus analysis of variance also revealed a significant Question x Accuracy interaction [$F(1, 644) = 5.59$, $p = 0.018$]. Because the three-way (Encoding x Question x Accuracy) interaction did not reach significance, we collapsed the data across encoding depth to interpret the Question x Accuracy interaction (Table E2, Figure 4). Planned comparisons indicated that, when participants answered the font color contextual information question, their FOK ratings were similar in magnitude regardless of whether they answered that question correctly ($M = 3.22$) or incorrectly ($M = 3.43$) [$t(227) = 0.84$, $p > 0.05$, *NS*]. However, when participants answered the category contextual information question, their FOK ratings were significantly higher when they answered that question correctly ($M = 5.09$) as opposed to incorrectly ($M = 4.03$) [$t(421) = 3.74$,

$p < 0.001$, $d = 0.45$]. Thus, when participants answered only one contextual information question, the quality of the retrieved contextual information influenced mean FOK ratings, but only when the contextual information was semantic in nature.

Experiment 1 Discussion

Experiment 1 produced a standard levels-of-processing effect on memory accuracy (e.g. Craik & Tulving, 1975) in that items encoded “deeply” (with a semantic encoding orientation) were recognized at higher rates than items encoded “shallowly” (with a perceptual encoding orientation). In addition, we found a levels-of-processing effect for both mean FOK judgments and relative prediction accuracy (mean gamma correlations), which is consistent with the findings of Lupker, et al. (1991). Deep encoding therefore not only improved objective performance, but also resulted in higher metacognitive judgments and better prediction accuracy. Finally, encoding depth also influenced the type of attribute that participants provided when complete recall failed. Paired sample t-tests indicated that, for items that had been encoded shallowly, participants answered the font color (perceptual) contextual information question (466 of 1361 trials, or 34.34%) just as often as they answered the category (semantic) question (486 of 1361 trials, or 35.71%) [$t = 1.15$, $p > 0.05$, *NS*]. However, for items that had been encoded deeply, participants answered the category (semantic) question (821 out of 1243 trials, or 66.05%) significantly more often than they answered the font color (perceptual) question (648 of 1243 trials, or 52.13%) [$t = 9.59$, $p <$

0.05]. Thus, deep encoding made participants more likely to retrieve semantic debris, as predicted.

Consistent with Koriat's (1993) accessibility model, the quantity of retrieved contextual information influenced participants' FOK judgments—specifically, the more contextual information retrieved, the higher participants rated their feeling of knowing (an 'additive' effect of quantity). A similar pattern emerged for objective performance; recognition performance improved when some contextual information was retrieved as opposed to none. While Koriat (1993) emphasized this effect of partial information quantity over partial information quality, Thomas et al. (2010) suggested that the correctness of retrieved contextual information might also influence feeling-of-knowing judgments. The results of the present study indicate that participants provided higher FOK judgments when they answered both questions but only the category question was correct as opposed to only the font color question was correct. Further, our secondary analysis (in which we collapsed observations across participants) indicated that, when participants answered only the font color contextual information question, the accuracy of their responses did not affect FOK ratings. However, when participants answered only the category contextual information question, the quality of the retrieved contextual information *did* matter—FOK ratings were significantly higher when participants answered the question correctly as opposed to incorrectly. Thus, when participants are accessing semantic contextual information, FOK judgments seem to reflect an influence of the quality of that information, and not just its accessibility. This is the first study to provide evidence that the type of retrieved contextual information has an effect on the feeling of knowing.

Experiment 2

In Experiment 2, we examined conditions where the quality of perceptual information should be more important to FOK judgments. Research has demonstrated that memory for pictorial details is enhanced when images are encoded perceptually as opposed to conceptually (Marks, 1991). Further, Intraub and Nicklos (1985) reported a reverse levels-of-processing effect when participants encoded images with physical or semantic orienting questions. They found that images were recalled at higher rates when they had been encoded perceptually (e.g. "Is this angular?") as opposed to conceptually (e.g. "Is this edible?")—the opposite pattern as typically found with lexical stimuli. Intraub and Nicklos (1985) dubbed this unexpected trend the "physical superiority effect," and they suggested that, because semantic information is rapidly extracted from images (as cited in Intraub & Nicklos, 1985), the semantic orienting questions might have encouraged processing that was redundant with the automatic processes, thus resulting in less effortful encoding or a less detailed memory trace. The physical encoding questions, on the other hand, might have directed participants' attention to aspects of the scene that were not processed automatically, thus resulting in more elaborate or distinctive memory traces and better retention (Intraub & Nicklos, 1985).

In light of this evidence, we hypothesized that the quality of retrieved contextual information might matter for both semantic *and* perceptual details when using pictorial stimuli. Experiment 2 sought to determine: (a) whether the quality of contextual information would affect participants' feeling-of-knowing judgments for pictorial stimuli, (b) if so, whether quality would matter only for semantic contextual information or also for perceptual contextual information, and (c) how these variables might interact with encoding orientation.

Method

Materials. We constructed 48 picture pairs using concrete images from Rossion and Pourtois' (2004) normed set of "Snodgrass and Vanderwart 'Like' Objects." (see Appendix B for examples). We selected six target images to represent each of the following eight categories: mammals, birds, clothing, fruit, furniture, instruments, transportation, and tools. We selected 16 cue images from these same eight categories, and we selected the other 32 cue images from the remainder of the image set. Of the 48 picture pairs, there were 16 "same-category" pairs (e.g. Gorilla-Zebra) and 32 "different-category" pairs (e.g. Pen-Eagle). The same-category pairs included two pairs from each of the eight target categories, and the different-category pairs included four targets from each of the eight target categories (Appendix B). We used the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998) to examine forward and backward association strength between the cue-target labels (e.g. Tree-Hammer), and we determined that none of the picture pairs were related.

We constructed the recognition test by pulling distractor images from the list of 96 cues and targets. There were a couple of constraints, however: each cue and target image appeared as an answer choice exactly twice, and the four alternatives for any given cue were unique in their category membership (e.g. only one option could be a bird, or a tool, etc). Once again, we used the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998) to examine forward and backward association strength between the cues and distractor items, as well as between the targets and distractor items, to ensure that all possible combinations formed unrelated pairs.

Procedure. The design of the Experiment 2 was a 2 (Encoding Orientation: perceptual vs. semantic) x 2 (Contextual Information Question: texture vs. category) within-subject factorial.

The procedure of Experiment 2 was the same as Experiment 1, with a few modifications. During the encoding phase, participants were presented with 48 picture pairs, one at a time, and asked to answer one of the following encoding questions for each pair: “Which object is more textured?” or “Could these objects be classified in the same category?” The texture question was designed to promote perceptual processing of the picture pairs, while the category question was designed to promote semantic processing of the picture pairs. The encoding depth of each block of 24 picture pairs was counterbalanced across participants.

In the feeling-of-knowing (FOK) phase, participants were presented with only the cue image of a cue-target pair and asked to produce the label for the target image (e.g. to type “raccoon” if the target image had been a raccoon). Participants then had an opportunity to provide contextual information related to the target. The perceptual contextual information question was, “Was the target more textured than the cue?” and the semantic contextual information question was, “Was the target in the same category as the cue?” Participants could respond “yes,” “no,” or “abstain.” After responding to both questions, participants were asked to rate their feeling of knowing for the target, as in Experiment 1.

The recognition phase of the experiment was executed as described previously, but the procedure for Experiment 2 included an additional, post-recognition phase. Because judgments of “texture” are subjective, it was necessary to collect these judgments from each participant for all 48 pairs so that we would later be able to score perceptual contextual information accuracy on a participant-by-participant basis. Accordingly, participants completed a second encoding phase (with the opposite counterbalance of their initial encoding phases) after the recognition test. Thus, each participant answered *both* orienting questions for all 48 picture pairs.

Experiment 2 Results

Alpha was set at 0.05 for all analyses. Items for which participants produced the correct response during the cued recall phase of the experiment (14.11%) were again excluded from the following analyses (see Figure 6 for the distribution of FOK judgments for correctly recalled items; Figure 5 illustrates the distribution of FOK judgments for incorrectly recalled items). As in Experiment 1, overall recognition ($M = 0.60$) was significantly better than chance (0.25) [$t(30) = 11.76$].

Encoding Orientation. We calculated two feeling-of-knowing (FOK) judgment means for each participant, one for items that had been encoded perceptually and the other for items that had been encoded semantically.

Participants assigned similar FOK judgments to items that had been encoded perceptually (*M* = 4.26) as opposed to semantically (*M* = 4.07) [$t(30) = 1.35$, $p > 0.05$, *NS*]. We calculated two mean recognition scores for each participant, one for items that had been encoded perceptually and the other for items that had been encoded semantically. An analysis of encoding orientation revealed that, across participants, recognition accuracy was similar for items encoded perceptually (*M* = 0.58) as opposed to semantically (*M* = 0.62) [$t(30) = 1.69$, $p > 0.05$, *NS*]. There was also no effect of encoding orientation on relative prediction accuracy. A paired samples t-test on mean gamma correlations revealed that participants predicted future recognition performance with similar accuracy for shallowly encoded items ($M_\gamma = 0.32$, $N = 31$) and deeply encoded items ($M_\gamma = 0.34$, $N = 31$) [$t(30) = 0.21$, $p > 0.05$, *NS*].

Quantity of Retrieved Contextual Information. To examine the effects of contextual information *quantity*, FOK judgment means were computed for three groups of items: (a) items for which participants chose the “abstain” response for both contextual information questions (texture and category); (b) items for which participants answered one of the contextual information questions and abstained for the other; and (c) items for which participants answered both contextual information questions. For each of these groups, two means were calculated for each participant—one for items encoded with a focus on texture and one for items encoded with a focus on picture category. Accuracy of the retrieved contextual information was irrelevant to this analysis.

A repeated measures 2 (Encoding Orientation: perceptual, semantic) x 3 (Contextual Information Quantity: abstained for both, answered one, answered both) ANOVA was conducted on participants' mean FOK judgments (Figure 7). There was a significant main effect of Contextual Information Quantity [$F(2, 38) = 127.83$], indicating that mean FOK judgments increased as the amount of retrieved contextual information increased. Planned comparisons revealed that mean FOK judgments were lower when participants abstained for both contextual information questions ($M = 1.28$) than when they answered one contextual information question ($M = 3.88$) [$t(24) = 8.69$, $d = 1.41$]. FOK judgments were higher when participants provided answers for both contextual information questions ($M = 5.28$) as opposed to just one contextual information question [$t(26) = 5.77$, $d = 1.34$]. There were no other significant effects.

To examine the effects of the quantity of retrieved contextual information on objective performance, a repeated measures 2 (Encoding Orientation: perceptual, semantic) x 3 (Contextual Information Quantity: abstained for both, answered one, answered both) ANOVA was conducted on participants' mean recognition scores. There were significant main effects of both Encoding Orientation [$F(1, 19) = 6.32$] and Contextual Information Quantity [$F(2, 38) = 4.96$]. Mean recognition scores were higher for items that had been encoded semantically ($M = 0.61$) as opposed to perceptually ($M =0.56$). Mean recognition scores were also better when participants retrieved two pieces of contextual information as opposed to one piece or no contextual information. Planned comparisons revealed that mean recognition scores were similar when participants abstained for both contextual information questions (*M* = 0.47) or answered one contextual information question (*M* = 0.55) [$t(24) = 1.11$, $p > 0.05$, *NS*]. However, mean recognition scores were higher when participants answered both contextual information questions (*M* = 0.67) as opposed to just one [$t(26) = 2.42$, $d = 0.26$]. There were no other significant effects.

Quality of Retrieved Contextual Information. To examine the effects of contextual information *quality* (accuracy), we conducted a 2 (Encoding Orientation: perceptual, semantic) x 3 (Question Accuracy: texture correct, category correct, both correct) repeated measures ANOVA on participants' mean FOK judgments (when a mean could not be computed, the mean for that group was substituted for the missing data, as recommended by Cohen et al., 2003) (Table E3). There was a significant main effect of Question Accuracy [$F(2, 60) = 10.49$]. Mean FOK judgments were lower when only the texture question was correct (*M* = 4.56) as opposed to when only the category question was correct (*M* = 5.52) [$t(30) = 2.89$, $d = 1.84$] or both contextual information questions were correct (*M* = 5.67) [$t(30) = 5.84$, $d = 1.06$]. There was no difference between mean FOK judgments when the category question was correct or both questions were correct [$t(30) = 0.61$, $p > 0.05$, *NS*]. There were no other significant effects. It should be noted that, once again, the number of observations that were used to generate FOK means differed as a function of condition. On average there were fewer observations when participants only answered the texture question correctly (*Mshallow* = 1.94; *Mdeep* = 2.15). There were more observations when participants answered the category question correctly (*Mshallow* = 3.38; *Mdeep* = 3.72).

Secondary Analysis for the Quality of Retrieved Contextual Information. As with Experiment 1, in an effort to be thorough, we also decided to collapse observations across participants in Experiment 2. We scored the accuracy of participants' responses to the "texture" contextual information questions according to their responses from the encoding phases (one counterbalance was completed at the beginning of the experiment and the other was completed post-recognition).

For the instances in which participants answered only one contextual information question, a 2 (Encoding Orientation: perceptual, semantic) x 2 (Contextual Information Question: texture, category) x 2 (Question Accuracy: correct, incorrect) ANOVA was conducted using mean FOK ratings as the dependent variable. The omnibus analysis of variance revealed a significant Question x Accuracy interaction [$F(1, 264) = 4.86$, $p = 0.028$]. Because the three-way (Encoding x Question x Accuracy) interaction did not reach significance, we collapsed the data across encoding orientation to interpret the Question x Accuracy interaction. Planned comparisons indicated that, when participants answered the texture contextual information question, their FOK ratings were similar in magnitude regardless of whether they answered that question correctly ($M = 3.46$) or incorrectly ($M = 3.87$) [$t(116) = 1.16$, $p > 0.05$, *NS*]. However, when participants answered the category contextual information question, their FOK ratings were higher when they answered that question correctly ($M = 4.27$)

as opposed to incorrectly (M = 3.14) [$t(152) = 1.87$, $p = 0.063$, $d = 0.43$], though this comparison was only marginally significant. Thus, when participants answered only one contextual information question, the quality of the retrieved contextual information influenced mean FOK ratings, but only when the contextual information was semantic in nature. These findings are consistent with Experiment 1.

Experiment 2 Discussion

Using concrete picture pairs, Experiment 2 provided additional support for the conclusions of Experiment 1: participants had stronger feelings of knowing when they answered only the category question correctly as opposed to only the texture question correctly. We had hypothesized that perceptual attributes might matter more for pictures than for words, but this was not the case—as with the word pairs in Experiment 1, only the quality of the semantic attribute influenced FOK judgments. In addition, the results of the Experiment 2 secondary analysis (in which observations were collapsed across participants) were consistent with those of Experiment 1; when only one contextual information question was answered, FOK judgments were higher when the category question was answered correctly as opposed to incorrectly, but mean FOKs did not differ as a function of the quality of the perceptual contextual information question. Finally, Experiment 2 also found an effect of contextual information quantity; the more contextual information retrieved by the participant, the higher the FOK judgment.

Unlike Experiment 1, however, encoding orientation did not have a significant effect on FOK means, recognition performance, or relative prediction accuracy in Experiment 2. Encoding depth did affect the type of attribute participants provided when recall failed, though. Paired sample t-tests indicated that, for items that had been encoded perceptually (with a focus on texture), participants answered the perceptual contextual information question (506 of 672 trials, or 75.30%) significantly more often than they answered the semantic question (475 of 672 trials, or 70.68%) [$t = 2.7$, $p < 0.05$]. This contrasts with Experiment 1, in which we found no differences in the type of attribute provided for items that had been encoded perceptually. For items that had been encoded semantically (with a focus on object category), the opposite pattern emerged: participants answered the semantic question (413 of 607 trials, or 66.39%) significantly more often than they answered the perceptual question (346 of 607 trials, or 57.00%) [$t = 5.67$, $p < 0.01$].

Interestingly, Marks (1991) suggested that, for pictorial stimuli, the effectiveness of a given encoding orientation (in terms of retention) is dependent upon the nature of the test. He found that conceptual encoding facilitates both recall and recognition of pictures' names, but that this benefit also comes at a cost—namely, participants' ability to recognize specific visual details of the stimuli (Marks, 1991). If we look at Experiment 2 cued recall performance—a task that required participants to produce the label for the target image, rather than recognize the image—we see that approximately 66% of items correctly recalled during the cued recall phase (N = 138 out of 210) had been encoded *semantically*. This seems to be consistent with the findings of Marks (1991), in that a semantic encoding orientation appeared to benefit our participants in terms of label cued recall but not image recognition, and this provides a plausible explanation for the lack of encoding orientation effects on recognition performance in Experiment 2.

Experiment 3

The goal of Experiment 3 was to examine whether the quality of retrieved contextual information would affect FOK judgments when intrinsic meaning was removed from the stimuli but extrinsic meaning was still manipulated via levels of processing. Thus, the stimuli used in Experiment 3 were pairs of ambiguous figures—pictorial stimuli that are not inherently meaningful but could potentially take on meaning for the participants if presented along with a semantic orienting question at encoding. We hypothesized that, even when participants had encoded the images with a "meaningful" (semantic/conceptual) orientation, the quality of retrieved contextual information would not influence participants' feeling of knowing because the figures themselves lack intrinsic meaning (unlike the concrete word pairs and picture pairs used in Experiments 1 and 2).

Method

Materials. I created forty ambiguous figures from basic shapes using the liquid vector tool in Real-DRAW Pro, a shareware program available on the Internet (http://www.mediachance.com/realdraw/index.html). Each ambiguous figure was colored red, blue, or green. Pairs were constructed so that one figure was visibly bigger than the other, and, for half of the pairs, the smaller figure could "fit" completely inside of the larger figure if superimposed. Figures varied in whether their contours were primarily smooth or jagged, and pairs always consisted of two differently colored figures (see Appendix C for example pairs).

We used 20 stimulus pairs in Experiment 3 (as opposed to 60 word pairs in Experiment 1 and 48 picture pairs in Experiment 2) because pilot testing highlighted the difficulty of the memory task for participants. Ambiguous figures are not amenable to verbal labels; thus, participants could not complete a traditional cued recall task during the FOK phase. Instead, participants were presented with an array of all the target figures, numbered 1-20 on a paper handout, and asked to select the correct target figure when presented with the cue figure. Three versions of the target array handout were created for this "closed set" cued recall task, each featuring a different arrangement of the figures to account for any possible order effects (i.e. to reduce the possibility that location within the array would influence whether or not an item was selected). We counterbalanced the array version given to each participant across the three encoding conditions. Though technically a recognition task, pilot testing indicated that the "closed set" cued recall task resulted in hit rates similar to those observed with the standard cued recall tasks in Experiments 1 and 2. Finally, pilot participants performed at chance on the final recognition task when presented with a four-alternative forced-choice test; therefore, we presented experimental participants with a three-alternative forced-choice test instead.

Distractor figures for the recognition test were pulled from the list of 20 targets. Because all 20 cues had been paired with targets of a different color (i.e. a red cue was never paired with a red target), none of the answer options on a given recognition trial included a figure that was the same color as the cue. For twelve of the recognition trials, one of the distractors was the same color as the target figure (for example, the three answer options might include the *red* target, a *red* distractor, and a blue distractor). On the remaining eight trials, neither distractor was the same color as the target (e.g. *red* target, blue distractor, blue distractor). Each set of answer options also included at least one “smaller” and one “larger” figure (excluding one trial, which—due to experimenter error—contained three “smaller” figures). Each target figure appeared at least twice during the recognition phase, once as the correct response and then again as a distractor. Targets varied in the number of times they appeared as a distractor, with a range of 1-3. This variety of color and size combinations was utilized in order to minimize the possibility that participants would use any one particular rule to help them eliminate lures during the recognition test.

A post-experimental questionnaire (Appendix D) asked participants how they viewed/perceived the ambiguous figures during the encoding phase (e.g. did they focus on perceptual characteristics, or did they try to think of objects that the figures resembled?), as well as whether they used specific strategies during the recognition task. This questionnaire allowed us to verify that participants in the “dangerousness” encoding condition (see Procedure for description) were thinking about the figures differently than in the other two encoding conditions (i.e. they weren’t basing their dangerousness judgments solely on how many points a figure possessed, for example). In addition, review of the completed questionnaires revealed that participants did not seem to rely on any one specific strategy for eliminating lures on the recognition task.

Procedure. The design of the experiment was a 3 (Encoding Orientation: dangerousness, fit, points) x 3 (Contextual Information Question: dangerousness, size, color) mixed factorial with repeated measures on the last factor. The procedure of Experiment 3 was similar to Experiments 1 and 2, with some modifications.

During the study phase, participants were presented with 20 ambiguous figure pairs, one at a time, and asked to answer one of the following encoding questions for each pair: (1) "Which figure has more points?", (2) Could one of these figures fit inside the other?", and (3) "Which figure is more dangerous?" Encoding condition was manipulated between subjects and figure pairs were presented in a random order. The "points" question was designed to promote perceptual (or "shallow") processing of the figures—it could be answered by simply counting the number of points on each figure. The "fit" question was designed to promote perceptual processing of the figures as well, but at a more effortful level (because participants had to mentally superimpose the smaller figure onto the larger one). The "dangerousness" question was designed to promote meaningful (or "deep") processing of the ambiguous figures. Although ambiguous figures are not inherently meaningful, they can still be evaluated on the basic dimension of dangerousness. Participants might not be able to label the ambiguous figures verbally, but they can still indicate which figure "feels" more dangerous or threatening—a type of meaningful processing that aids survival. Indeed, adaptive memory research has demonstrated that processing information in terms of its survival relevance aids in long-term retention; in fact, it appears to be a “deeper” form of encoding than semantic processing or self-referential processing (two of the most often used “deep” levels-of-processing manipulations) (Nairne & Pandeirada, 2008).

There was no filler task due to overall task difficulty. In the closed set cued recall/FOK phase, participants were randomly presented with each of the 20 cue figures on the computer screen, one at a time, and asked to select the target figure from the array handout. Participants were given 12 seconds to type the target figure’s corresponding number into a response box on the computer. Once a response had been provided or time had expired, participants were then presented with the three contextual information questions: “What was the color of the target?” (red, blue, green, abstain); “Was the target larger than the cue?” (yes, no, abstain); “Was the target more dangerous than the cue?” (yes, no, abstain). The order of the three contextual information questions varied across participants, as the computer program randomly selected one of three predetermined orders. After responding to the contextual information questions, participants were asked to rate their feeling of knowing (“What are your chances of recognizing the correct target figure out of three choices?”) on a 0-10 Likert scale.

The final recognition phase of the experiment was executed as described for Experiments 1 and 2, although participants in Experiment 3 completed a three-alternative forced-choice recognition test instead of a four-alternative forced-choice test. This was followed by a post-recognition phase (as in Experiment 2) in which participants judged the ambiguous figure pairs on the dimensions that they had not encountered during the initial encoding phase. For example, participants who were in the “points” condition at initial encoding completed the “fit” and “dangerousness” encoding programs after their final recognition test. Our rationale for this procedure aligns with the one described for Experiment 2; because judgments of “dangerousness” are subjective, it was necessary to collect responses from all participants on this dimension so that we would later be able to score contextual information accuracy on a participant-by-participant basis.

Finally, participants completed the post-experimental questionnaire. They were then debriefed and thanked for their participation.

Experiment 3 Results

Alpha was set at 0.05 for all analyses. As in Experiments 1 and 2, items for which participants produced the correct response during the closed set cued recall phase of the experiment (4.54%) were excluded from the following analyses (see Figure 9 for the distribution of FOK judgments for items that were correctly recalled; Figure 8 illustrates the distribution of FOK judgments for items that were incorrectly recalled). As in Experiments 1 and 2, overall recognition (M = 0.51) was significantly better than chance (0.333) [$t(64) = 9.68$].

Encoding Orientation. Encoding orientation was manipulated between participants, and we conducted a one-way ANOVA (Encoding Orientation: dangerousness, fit, points) on the dependent variables of mean FOK judgments, mean recognition scores, and mean gamma correlations. Encoding orientation did not influence the magnitude of FOK judgments [$F(2, 64) = 0.11$, $p > 0.05$, *NS*], participants’ objective performance [$F(2, 64) = 0.31$, $p > 0.05$, *NS*], or participants’ prediction accuracy [$F(2, 62) = 3.01$, $p > 0.05$, *NS*]. One-sample t-tests revealed that prediction accuracy was not statistically different from chance (0) for the Dangerousness condition (M_γ = 0.01, N = 21) [t(20) = 0.14, p > 0.05, *NS*] or the Fit condition (M_γ = 0.11, N = 21) [t(20) = 1.08, p > 0.05, *NS*], but the gamma correlation mean was significantly worse than chance for the Points condition (M_γ = -0.17, N = 21) [t(20) = 2.46]. This finding does not appear to be important theoretically—it is most likely an artifact produced by a couple of extreme participant means.

Quantity of Retrieved Contextual Information. Because there was no overall effect of encoding orientation on mean FOK judgments, we did not include it as a factor in the following analyses. To examine the effects of contextual information *quantity*, FOK judgment means were computed for four groups of items for each participant: (a) items for which participants chose the "abstain" response for all three contextual information questions; (b) items for which participants answered one of the contextual information questions and abstained for the other two; (c) items for which participants answered two of the contextual information questions and abstained for the third; and (d) items for which participants answered all three contextual information questions. Accuracy of the retrieved contextual information was irrelevant to this analysis.

We conducted a repeated measures (Contextual Information Quantity: answered one question, answered two questions, answered three questions) ANOVA on participants' mean FOK judgments (Table E4). There was a significant effect of CI Quantity [F(2, 80) = 38.18], indicating that mean FOK judgments increased as the amount of retrieved contextual information increased

(Figure 10). Planned comparisons indicated that mean FOK judgments were lower when participants answered only one contextual information question (*M* = 1.92) than when they answered two contextual information questions (*M* = 3.34) [$t(32) = 4.07$, $d = 2.05$]. In addition, mean FOK judgments were highest when participants answered all three contextual information questions (*M* = 4.63) [answer three vs. answer two: $t(51) = 5.40$, $d = 1.71$]. Though listed in the Table E4, we excluded the "abstained all questions" means from the analysis because missing observations for this group significantly reduced the number of participants that could be included in the repeated measures ANOVA (it should be noted, however, that the effect of CI Quantity was still significant when all four levels were included). We also ran this ANOVA with participants' mean recognition scores as the dependent variable, but the effect of CI Quantity was not significant [$F(2, 82) = 1.20$, $p > 0.05$].

Analyses for the Quality of Retrieved Contextual Information. For Experiment 3, we could only analyze the effects of the quality of retrieved contextual information on FOK judgments if we collapsed the data across participants (this was due to the limited number of observations per each class of items). In our first analysis of contextual information quality, FOK means were calculated, across participants, for three groups of items: abstained for all three questions, answered all three questions incorrectly, and answered all three questions correctly. It should be noted that the number of observations contributing to each mean varied between the three groups (abstained for all three questions, N = 316; answered all three incorrectly, N = 51; answered all three correctly, N = 75). We conducted a one-way (Contextual Information Quality: abstain all, all incorrect, all correct) ANOVA on the FOK means. There was a significant effect of CI Quality [$F(2, 439) = 159.94$]. Post-hoc analyses indicated that FOK means were significantly lower when participants abstained for all three contextual information questions ($M = 0.94$) than when they answered all three questions incorrectly ($M = 5.04$) [$t(365) = 16.17$] or when they answered all three questions correctly ($M = 3.97$) [$t(389) = 13.08$] (Figure 11). Interestingly, the FOK mean for the "all incorrect" ($M = 5.04$) group of items was actually higher than the FOK mean for the "all correct" ($M = 3.97$) group [$t(124) = 2.37$]. We also ran this one-way ANOVA on the dependent variable of participants' mean recognition scores for these three groups of items and found a significant effect of CI Quality [$F(2, 442) = 3.31$] (Table E5). Post-hoc analyses indicated that recognition scores were higher when participants either abstained for all three contextual information questions ($M = 0.51$) or answered all three questions correctly ($M = 0.53$) as opposed to when they answered all three questions incorrectly ($M = 0.33$) [$t(368) = 2.46$ and $t(125) = 2.33$]. The means for the "abstain all" and "all correct" groups of items did not differ [$t(391) = 0.37$, $p > 0.05$, *NS*].

In our second analysis of contextual information quality, we selected just the trials for which participants had answered all three contextual information questions but only answered one of them correctly. Then we computed FOK means for three groups of items: only color question correct ($M = 5.31$, N = 36), only size question correct ($M = 4.50$, N = 46), and only dangerousness question correct (*M* = 4.87, N = 85). A one-way (Question Accuracy: only color correct, only size correct, only dangerousness correct) ANOVA on participants' FOK means revealed that the effect of Question Accuracy was not significant [$F(2, 164) = 1.36$, $p > 0.05$, *NS*]. We also did not find a significant effect when we ran this same ANOVA on the dependent variable of participants' mean recognition scores [$F(2, 167) = 1.94$] (Table E5).

Experiment 3 Discussion

As in Experiment 2, there were no effects of encoding orientation on mean FOK judgments or recognition accuracy in Experiment 3. Participants' FOK judgments were at chance in predicting future recognition performance for the Fit and Danger conditions, but mean gamma correlations were significantly below chance for the Points condition (most likely a random finding). Consistent with both Experiments 1 and 2, there was an effect of the quantity of retrieved contextual information on mean FOK judgments, indicating that participants reported higher FOKs when they provided a greater amount of contextual information (i.e. when they answered more of the CI questions).

When we collapsed the data across participants, mean FOK judgments were significantly lower when no contextual information was provided as opposed to when all three contextual information questions were answered (again highlighting the effect of quantity). However, there was no effect of contextual information quality on FOK means—that is, answering those three questions correctly did not lead to higher FOK judgments than when the three questions were answered incorrectly (the reverse was actually found, though this makes no sense theoretically and was likely due to chance; a follow-up experiment will address this issue). Interestingly, when participants abstained for all three questions, their recognition scores were above chance levels, yet when they answered all three questions incorrectly, recognition fell to chance. This further demonstrates that the quantity of retrieved contextual information was driving FOK judgments in Experiment 3, and participants seemed completely unable to evaluate the quality of that information. Our second analysis of contextual information quality indicated that, unlike our findings for Experiments 1 and 2, the type of attribute retrieved by participants did not influence participants' feeling of knowing; FOK judgments were equivalent regardless of whether participants answered a perceptual question (color or size) correctly or the semantic question (dangerousness) correctly.

Together, these findings support our hypotheses for Experiment 3; for ambiguous figure pairs, participants' FOK judgments were influenced by the quantity of retrieved contextual information (as predicted by the accessibility model), but not the quality of that information. Even when "semantic" contextual information was provided (i.e. participants answered the dangerousness question), the accuracy of the response did not influence FOKs, presumably because the stimuli used in Experiment 3 were inherently meaningless. As a result, participants were either unable to evaluate retrieved contextual information for its quality, or the quality of that information was simply not predictive of future recognition performance (i.e. it was not a useful piece of attribute information).

General Discussion

The present study was designed to elucidate the conditions under which the quality of retrieved contextual information influences the episodic feeling of knowing. While Koriat's (1993) accessibility account posited that the quantity of retrieved partial information—correct or incorrect—drives feeling-of-knowing (FOK) judgments, Dunlosky et al. (2005) and Thomas et al. (2010) each showed that the quality of retrieved contextual information can influence metacognitive judgments in certain situations. We hypothesized that the quality (accuracy) of retrieved contextual information would only influence FOK judgments when the to-be-remembered items were inherently meaningful and the contextual information provided at retrieval was semantic in nature. The results of the present study support this hypothesis.

Meaning Matters

In Experiments 1 and 2, participants encoded concrete word pairs and nameable picture pairs, respectively. When participants answered both contextual information questions, mean FOK judgments were higher when participants provided only a correct semantic attribute than when they provided only a correct perceptual attribute. Further, secondary quality analyses (with observations collapsed across participants) indicated that, when participants answered only one contextual information question, mean FOK judgments were influenced by the quality of the retrieved contextual information only when that information was semantic in nature. Critically, these findings cannot be attributed solely to the levels-of-processing manipulation at encoding; although we found standard levels-of-processing effects in Experiment 1, encoding orientation did *not* influence FOK judgments or recognition scores in Experiment 2. Because we found similar effects of contextual information quality between the two studies, we can attribute our findings to intrinsic, rather than extrinsic, manipulations. What did the stimuli used in Experiment 1 have in common with the stimuli used in Experiment 2? Concrete word pairs and nameable picture pairs are both inherently meaningful and comprise semantic attributes that can be recalled even when total recall fails. Thus, though extrinsic manipulations can influence FOK judgments and affect the availability of certain types of contextual information during the retrieval attempt, we can conclude that the intrinsic meaningfulness of the stimuli, combined with the type of attribute retrieved, was critical in determining whether the quality of the retrieved contextual information would influence FOK judgments.

Why should the meaningfulness of retrieved contextual information matter? In a feature-based view of semantic memory, inherently meaningful information such as words and concrete pictures are stored in memory as a list of features, some of which are essential to the target item's conceptual representation (Koriat et al., 2003). Thus, partial recall of semantic features is useful in discriminating between targets and lures because those attributes are essential components of the item's overall representation. Whereas semantic attributes (such as category membership) are central to an item's conceptual representation in memory, certain perceptual characteristics (e.g. font color) can be arbitrarily assigned to individual items and therefore cannot be considered defining features/attributes. When we removed intrinsic meaning from the equation by using ambiguous figure pairs as the stimuli for Experiment 3, we found that the quality of retrieved contextual information did not matter—rather, FOK judgments were driven solely by the quantity of retrieved contextual information (consistent with the accessibility model's predictions).

Experiment 3 Revisited

Although the findings of Experiment 3 appear to support our hypothesis that the inherent meaningfulness of the stimuli is important in determining whether the quality of retrieved contextual information will influence the feeling of knowing, we must consider other potential explanations for our lack of quality effects with the ambiguous figure pairs. First, it is possible that our participants didn't effectively encode the ambiguous figure pairs at all—in which case we wouldn't expect to find any effects of quality because the targets wouldn't actually be stored in memory. It is possible that participants were simply guessing during the closed-set cued recall phase of Experiment 3 and that their correct responses reflect chance hit rates rather than actual recall. Although final recognition performance was above chance, it is possible that participants were able to eliminate lures using rules or logical reasoning, and therefore the recognition means might not be a valid indicator of whether items were initially encoded. Participants' responses to the Experiment 3 questionnaire (Appendix D) did not reveal any reliable patterns in self-reported recognition test strategies, but this does not provide concrete evidence through which we could dismiss the above concern.

Accordingly, we graphed the distributions of FOK judgments (collapsed across participants) for all three experiments in order to compare the distributions for items that were incorrectly recalled during the FOK phase (Figures 1, 5, and 8) with the distributions for items that were correctly recalled during the FOK phase (Figures 2, 6, and 9). In all three experiments, items that were correctly retrieved during the FOK phase were dropped from further analyses because it was assumed that these items would nearly always be given an FOK judgment of 9 or 10 and that they would be recognized at ceiling on the final recognition test. Our FOK distributions support the former assumption for Experiment 1 (Figure 2) and Experiment 2 (Figure 6); looking at the graphs, it is obvious that there is a very large negative skew to both distributions. However, this was not the case for Experiment 3 (Figure 9)—here, the distribution of FOK judgments seems to suggest that participants were not confident in their responses even when they had selected the correct target figure from the array. Thus, participants in Experiment 3 may have simply been guessing rather than actually retrieving the item from memory, and this could be due to inadequate learning during the encoding phase. We are currently addressing this issue in a follow-up experiment that aims to ensure that participants will encode the ambiguous figure pairs satisfactorily during the first phase of the experiment.

It is also possible that our extrinsic manipulation of meaningfulness failed in Experiment 3. Our “dangerousness” orienting question, which we selected based on recent evidence that survival-relevant processing results in a retention benefit beyond even some of the “deepest” levels of processing (see Nairne &

Pandeirada, 2008), might not have actually facilitated "deep" processing in Experiment 3. Although we had specific theoretical motivations for using this encoding orientation, the relevant adaptive memory literature has traditionally used lexical stimuli, so it is possible that our ambiguous figure pairs were not encoded in the "meaningful" way that we had intended. Perhaps a pleasantness question at encoding would have produced different results. Thus, while our findings suggest that the inherent meaningfulness of the stimuli and the meaningfulness of the retrieved contextual information both influence feeling-of-knowing judgments, more work should be conducted with inherently meaningless stimuli to rule out alternative explanations for our findings.

Rather than using a levels-of-processing manipulation with ambiguous figure pairs, which appear to be very difficult for participants to encode with only one stimulus presentation, one solution might be to use a multiple-trial learning procedure in which participants are taught to associate inherently meaningless stimuli with semantic information. For example, native English-speaking individuals could be presented with Japanese kanji characters and taught to associate these simple pictures with their actual Japanese meanings. Or, participants could be presented with letter strings like the ones Koriat (1993) used in his first two experiments (e.g. FKDR) and instructed to generate a phrase for the item as if it were an acronym (e.g. Fluffy Kittens Dislike Rats). By varying the number of learning trials, it might be possible to compare whether the quality of retrieved contextual information influences FOK judgments for inherently meaningless stimuli that have become weakly associated with semantic information vs. inherently meaningless stimuli that have actually *acquired* meaning for the participant through repeated presentations. This would be an interesting follow-up to the present study and would further elucidate the impact of conceptual meaning on the mechanisms underlying the feeling of knowing.

Accessibility and Evaluation

Like Koriat (1993), we found that the quantity of retrieved contextual information influences FOK judgments—specifically, the more 'debris' generated during the retrieval attempt, the higher participants rated their feeling of knowing. Koriat posited that the accessibility—the amount and intensity—of this information is what drives FOK judgments, regardless of whether it that information is correct or incorrect. Yet we've demonstrated that, for inherently meaningful stimuli, retrieving a correct piece of semantic contextual information results in a higher FOK judgment than retrieving a correct perceptual attribute, and FOK judgments appear to be sensitive only to the quality of semantic contextual information (as demonstrated by our secondary analyses, in which we collapsed observations across participants). Thus, the accessibility model cannot account for our findings with regard to the quality of retrieved contextual information.

The mechanisms underlying the feeling of knowing are complex, and future research must continue to disentangle the relative contributions of the quantity and quality of retrieved contextual information on episodic feeling of knowing tasks. Future research should also attempt to place the influences of target accessibility within a larger framework of inferential decision making—one that incorporates contributions from other sources, including general knowledge of the topic and cue familiarity. Taking into account previous work in this area, we would like to propose such a framework here—a multi-stage inferential model of the mechanisms underlying the feeling of knowing. While such a model may be preliminary in its design, it will allow us to consider how multiple factors might interact to produce the phenomenology of the feeling of knowing.

The first stage in the model takes into account the contributions of general knowledge of the topic and cue familiarity. Research has demonstrated that when a speeded FOK judgment (i.e. one that's made prior to any retrieval attempts) is solicited, participants rely on quick heuristics—for example, they can make plausibility judgments (e.g. Reder, 1982) or base their judgments on the familiarity of parts of the question (e.g. Reder & Ritter, 1992) or its fluency of processing (see Schwartz, 1994 for a more thorough review of early sources of information for metacognitive judgments). Koriat and Levy-Sadot (2001) argued that the cue familiarity account (Reder, 1987, as cited in Koriat & Levy-Sadot, 2001) and the accessibility account (Koriat, 1993) of the feeling of knowing aren't necessarily mutually exclusive. Rather, Koriat and Levy-Sadot (2001) suggested that the influences of cue familiarity appear early in the process and that this information can actually determine whether retrieval is attempted (i.e. if cue familiarity is low, a participant might end up with a feeling of *not* knowing and therefore wouldn't bother to search their memory for the target). We agree with this position and suggest that metacognitive monitoring processes will only move past this initial stage if cue familiarity is high or there is another source of motivation to attempt retrieval.

The next stage in our proposed model takes into account the contributions of target accessibility. When a speeded FOK judgment is not required and participants have time to attempt retrieval, the accessibility (amount and intensity) of retrieved contextual information likely exerts an initial influence on participants' subjective feeling of knowing, as proposed by Koriat (1993). The process, however, need not end here. Thomas et al. (2010) demonstrated that FOK accuracy improved when participants had an unlimited amount of time to make their FOK judgments, and they hypothesized that participants may have used this time to evaluate the retrieved partial information and search for additional relevant information. Thus, the final stage in our proposed model is an evaluation stage. Perhaps the quantity/accessibility of retrieved contextual information is the main determinant of FOK judgments when participants are under a time constraint or the stimuli are inherently meaningless. Encoding inherently meaningful stimuli might move the entire process beyond mere accessibility of retrieved contextual information to a later evaluation stage (either conscious or below conscious awareness) where the quality of retrieved contextual information can exert an influence on FOK judgments.

The findings of the present study support this hypothesis; in Experiments 1 and 2, the quality of retrieved contextual information mattered when that information was meaningful, indicating that participants had some ability to evaluate the accuracy of that information. In Experiment 3, however, participants encoded inherently meaningless pairs of ambiguous figures, and the *quantity* of retrieved contextual information influenced their FOK judgments, but not the quality of that information. Thus, stimuli meaningfulness may be one important factor in whether the feeling of knowing moves beyond its initial reliance on the accessibility of relevant contextual information to an actual evaluation of that information. For meaningful stimuli, the quality of semantic debris becomes more relevant to the retrieval process, and therefore the participant may have more motivation to evaluate those attributes for correctness.

The next step in this line of research would be to conduct an empirical investigation of the proposed framework to determine whether these processes operate in stages, as is proposed here, or whether it is possible that certain processes might run parallel to one another. Regardless of the actual time sequence of these processes, the findings of the present study suggest that the feeling of knowing is influenced not only by the accessibility of retrieved contextual information, but also by the quality of retrieved contextual information.

Appendix A

<u>Exp. 1 Same-Category Word Pairs</u>

Fruit:
1. Banana – Pineapple
2. Kiwi – Lemon
3. Cherry – Mango
4. Peach – Melon

Furniture:
1. Ottoman – Cabinet
2. Sofa – Dresser
3. Bookcase – Stool
4. Futon – Chair

Animals:
1. Donkey – Rat
2. Pig – Turtle
3. Elephant – Deer
4. Squirrel – Fox

Clothing:
1. Gloves – Belt
2. Hat – Shoe
3. Bra – Pants
4. Shirt – Underwear

Body Part:
1. Butt – Teeth
2. Wrist – Stomach
3. Knee – Lungs
4. Toe – Muscle

<u>Exp. 1 Different-Category Word Pairs</u>

Fruit:
1. Hail – Pear
2. Rugby – Grape
3. Tennis – Plum
4. Trumpet – Apple
5. Cliff – Tangerine
6. Mixer – Strawberry
7. Plateau – Lime
8. Tsunami – Orange

Furniture:
1. Airplane – Desk
2. Blizzard – Shelf
3. Bowling – Couch
4. Wrestling – Loveseat
5. Canyon – Bed
6. Flute – Table
7. Hammer – Recliner
8. Hose – Lamp

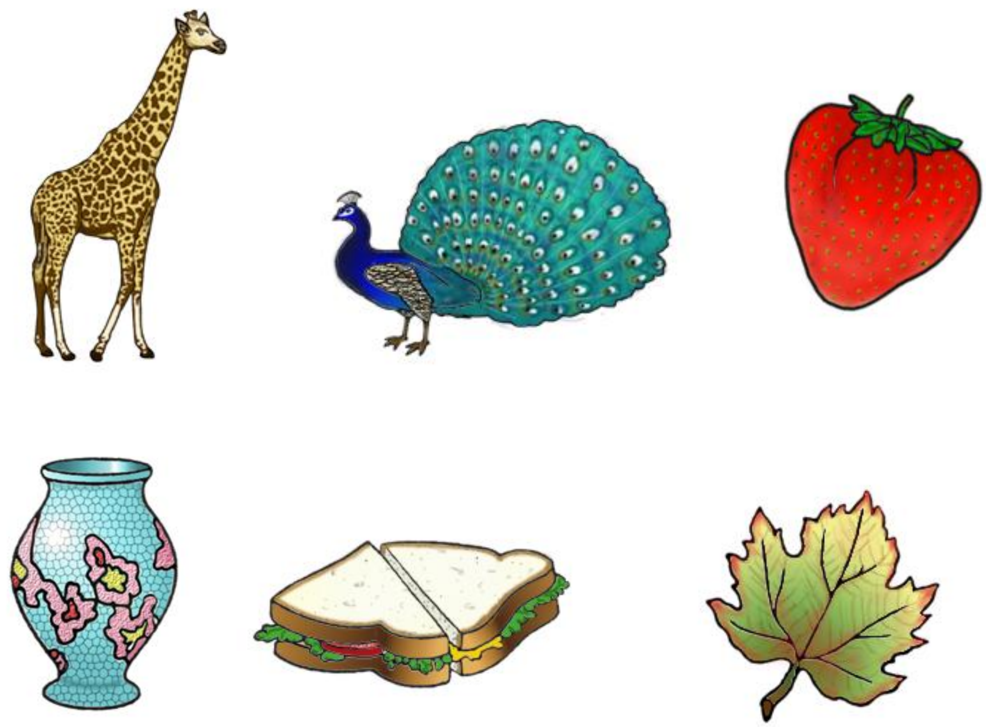
Animals:
1. Bucket – Lizard
2. Skiing – Cow
3. Swimming – Rabbit
4. Violin – Moose
5. Chisel – Sheep
6. Earthquake – Lion
7. Motorcycle – Tiger
8. Sunshine – Goat

Clothing:
1. Pencil – Socks
2. Seeds – Jacket
3. Taxi – Blouse
4. Volcano – Sweater
5. Bicycle – Coat
6. Grass – Shorts
7. Hurricane – Skirt
8. Knife – Jeans

Body Parts:
1. Drums – Hip
2. Helicopter – Ankle
3. Lily – Thigh
4. Ruler – Torso
5. Drill – Elbow
6. Orchid – Ear
7. Spoon – Shoulder
8. Volleyball – Nails

Appendix B

Example "Snodgrass-like" Images from Rossion and Pourtois (2004)



<u>Exp. 2 Same-Category Picture Pairs</u>

Mammals:

1. Giraffe – Raccoon
2. Gorilla – Zebra

Birds:

1. Peacock – Duck
2. Swan – Rooster

Clothing:

1. Sock – Jacket
2. Glove – Blouse

Fruit:

1. Banana – Pineapple
2. Watermelon – Pear

Furniture:

1. Bed – Stool
2. Desk – Couch

Instruments:

1. Harp – Drum
2. Accordion – Flute

Transportation:

1. Bus – Sailboat
2. Airplane – Motorcycle

Tools:

1. Saw – Pliers
2. Chisel – Screwdriver

Exp. 2 Different-Category Pairs

Mammals:
1. Racket – Pig
2. Plug – Tiger
3. Vase – Kangaroo
4. Well – Cow

Birds:
1. Snowman – Ostrich
2. Ball – Penguin
3. Clothespin – Owl
4. Pen – Eagle

Clothing:
1. Leaf – Shoe
2. Candle – Tie
3. Toaster – Pants
4. Watering Can – Vest

Fruit:
1. Kite – Strawberry
2. Umbrella – Cherry
3. Button - Grapes
4. Whistle – Lemon

Furniture:
1. Comb – Table
2. Sandwich – Dresser
3. Toothbrush - Chair
4. Wheel – Lamp

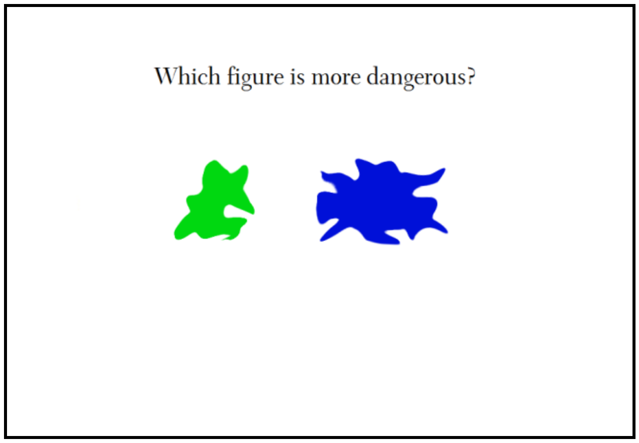
Instruments:
1. Sun – Piano
2. Balloon – Guitar
3. Flower – Trumpet
4. Light bulb – Violin

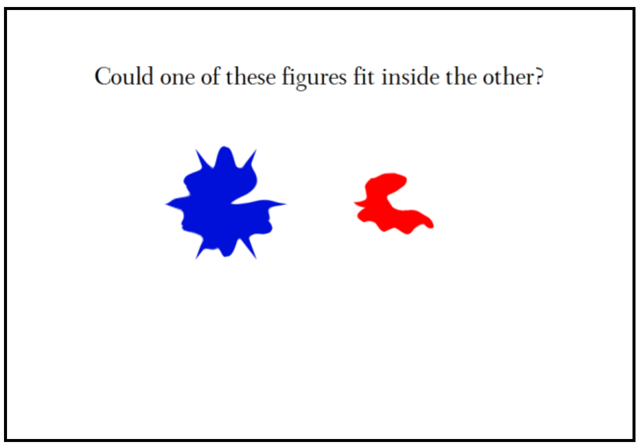
Transportation:
1. Garbage can – Helicopter
2. Key – Train
3. Purse – Car
4. Iron – Bicycle

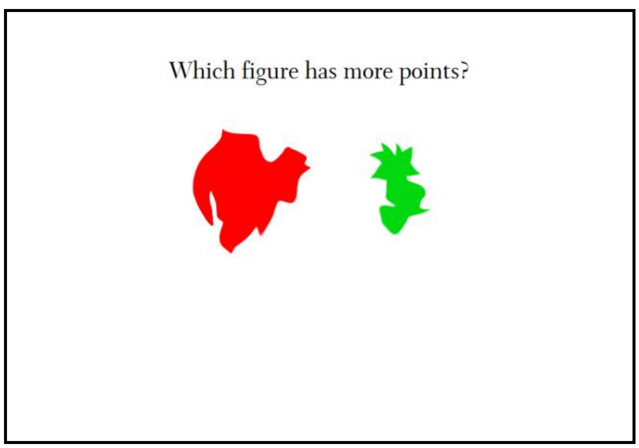
Tools:
1. Tree – Hammer
2. Helmet – Wrench
3. Cake – Screw
4. Basket – Ax

Appendix C

Example Experiment 3 Stimuli Pairs:







Appendix D

Experiment 3 Questionnaire:

Participant # ____________ Array # __________ Date________________

Final Questionnaire

Please read each question carefully and answer to the best of your ability.

1. In the **FIRST PHASE** of the experiment, on which dimension were you asked to judge the ambiguous figures? (circle one)

   POINTINESS FIT DANGEROUSNESS

2. Briefly describe how you made these judgments in the **FIRST PHASE** of the experiment (i.e. describe your thought process/decision criteria):

3. On a scale of 0 – 10, please rate how difficult you thought it was to make this type of judgment (circle a number):

   0 1 2 3 4 5 6 7 8 9 10

   Not at all difficult Extremely difficult

4. Please indicate which of the following statements best describes the way you viewed the ambiguous figures while making judgments in the **FIRST PHASE** of the experiment:

   ☐ I paid attention to the physical features of the figures (i.e. shape, size, etc).

   ☐ I tried to think of the ambiguous figures as real objects and/or I thought about objects that they reminded me of.

   ☐ Other (please elaborate): ______________________________________

   ____

4. On the **final recognition test**, you were asked to pick the target figure out of four answer choices. Briefly describe how you made your decisions (i.e. any strategies you may have used):

Appendix E

Table E1

Exp. 1 FOK Means by Encoding Orientation and Contextual Information Quality

	Shallow Encoding		Deep Encoding	
Contextual Information Accuracy	*M* (*SE*)	*N**	*M* (*SE*)	*N**
Only Font Color Correct	3.74 (0.20)	45 (12)	5.33 (0.26)	45 (12)
Only Category Correct	4.60 (0.24)	45 (1)	6.39 (0.20)	45 (0)
Both Questions Correct	5.59 (0.28)	45 (10)	6.68 (0.24)	45 (7)

*Note. *Number of mean replacements indicated in parentheses*

Table E2

Exp. 1 FOK Means by Contextual Information Quality – One Question Answered

	Answered Font Color Question		Answered Category Question	
Contextual Info Accuracy	*M* (*SE*)	*n*	*M* (*SE*)	*n*
Incorrect	3.43 (0.19)	113	4.03 (0.24)	94
Correct	3.22 (0.17)	116	5.09 (0.14)	329

Table E3

Exp. 2 FOK Means by Encoding Orientation and Contextual Information Quality

Contextual Information Accuracy	Perceptual Encoding		Semantic Encoding	
	M (*SE*)	*N**	*M* (*SE*)	*N**
Only Texture Correct	4.23 (0.33)	31 (15)	4.90 (0.31)	31 (11)
Only Category Correct	5.77 (0.42)	31 (5)	5.28 (0.40)	31 (2)
Both Questions Correct	5.73 (0.32)	31 (2)	5.62 (0.29)	31 (2)

Note. *Number of mean replacements indicated in parentheses*

Table E4

Exp. 3 FOK Means by Contextual Information Quantity

Contextual Information Quantity	FOK Mean	
	M (*SE*)	*N*
Abstained All	1.07 (0.19)	42
Answered One	1.92 (0.21)	50
Answered Two	3.34 (0.26)	54
Answered Three	4.63 (0.22)	58

Table E5

Exp. 3 Recognition Scores by Contextual Information Quality (Collapsed Across Participants)

	Recognition Score	
Contextual Information Quality	*M* (*SE*)	*N*
Abstained All	0.51 (0.03)	318
All Incorrect	0.33 (0.07)	52
All Correct	0.53 (0.06)	75
Only Color Correct	0.36 (0.08)	36
Only Size Correct	0.55 (0.07)	47
Only Dangerousness Correct	0.54 (0.05)	87

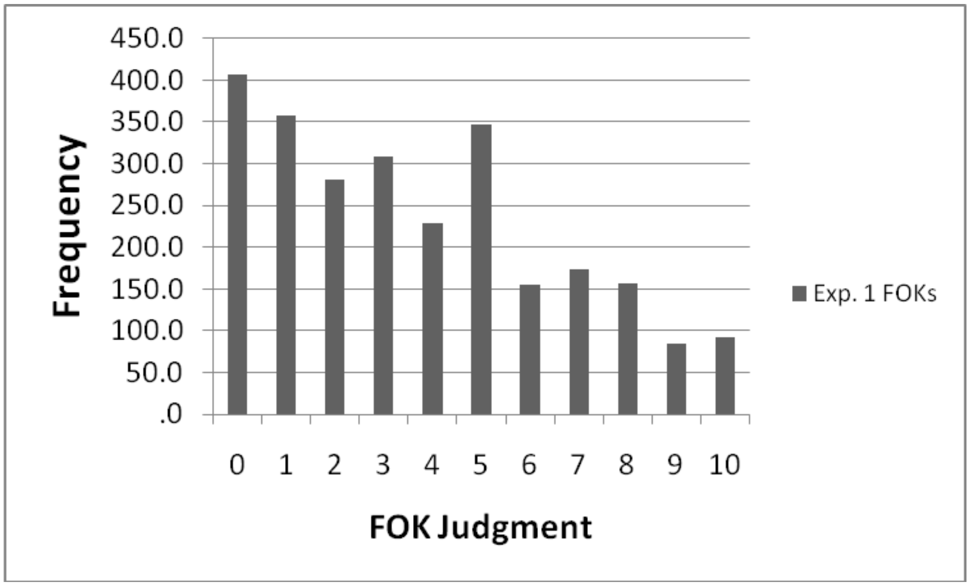


Figure 1. Experiment 1: Distribution of FOK judgments (collapsed across participants) that includes all items for which cued recall was incorrect.

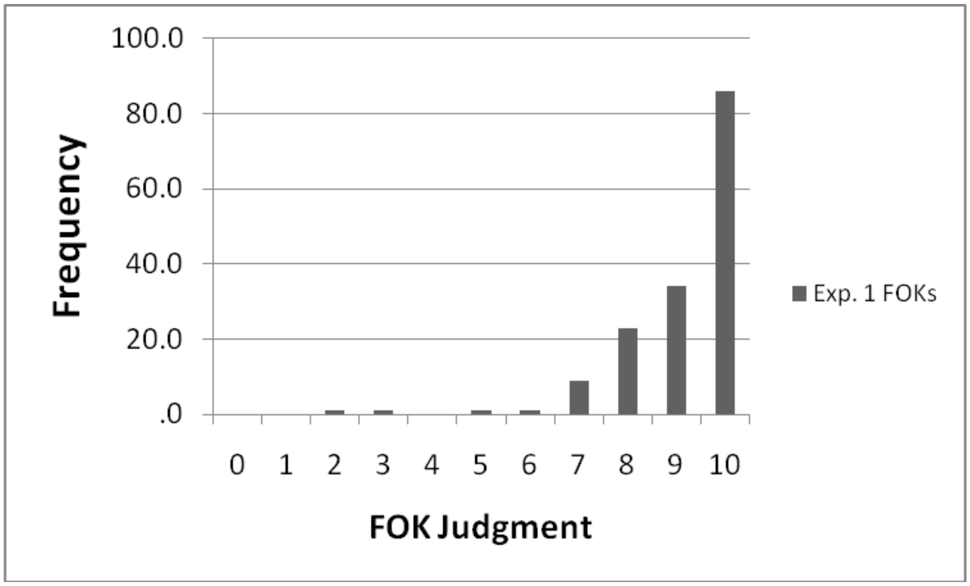


Figure 2. Experiment 1: Distribution of FOK judgments (collapsed across participants) that includes all items for which cued recall was correct.

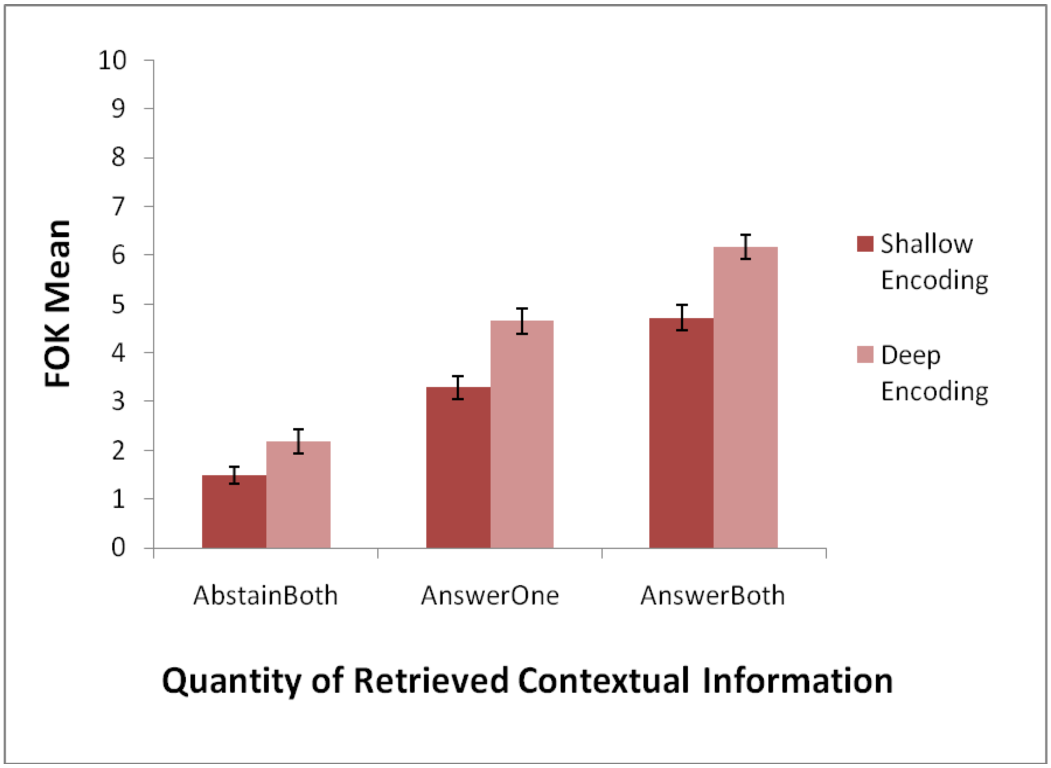


Figure 3. Experiment 1: Mean feeling-of-knowing (FOK) judgments as a function of the quantity of retrieved contextual information.

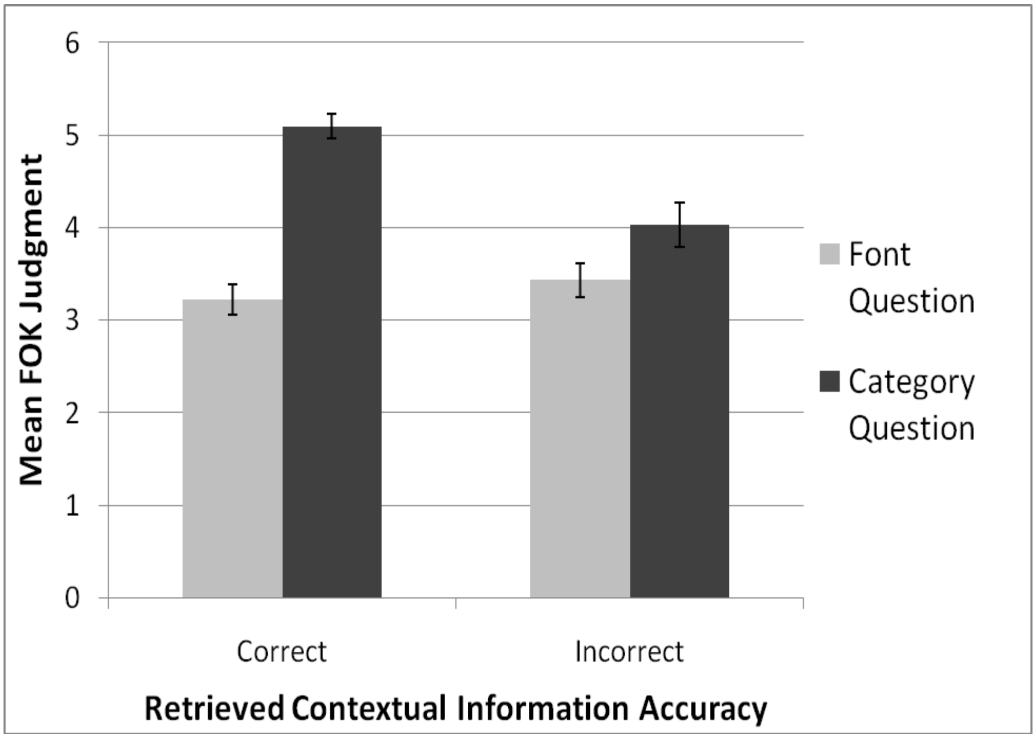


Figure 4. Experiment 1: Mean feeling-of-knowing (FOK) ratings (collapsed across participants) as a function of the quality of retrieved contextual information, for situations in which only *one* contextual information question was answered.

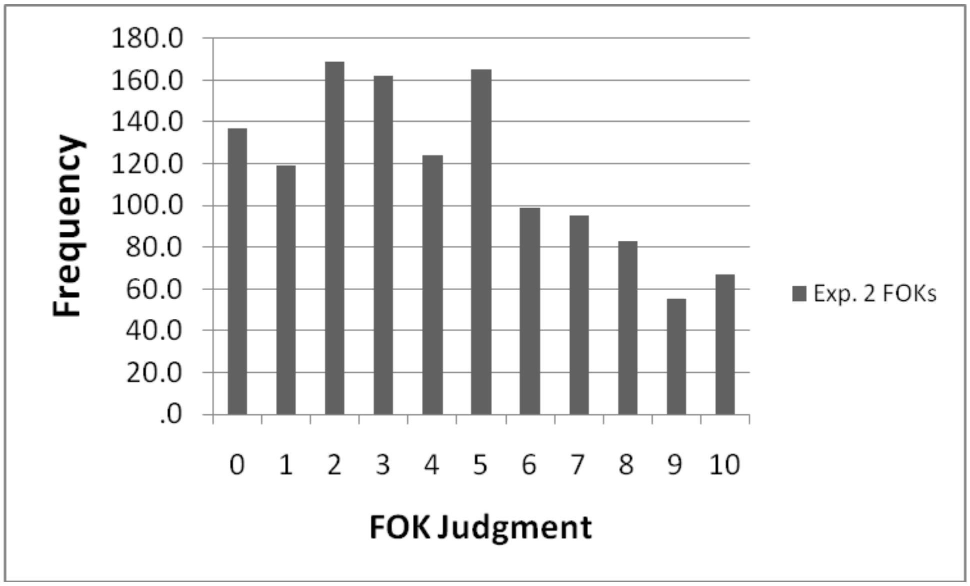


Figure 5. Experiment 2: Distribution of FOK judgments (collapsed across participants) that includes all items for which cued recall was incorrect.

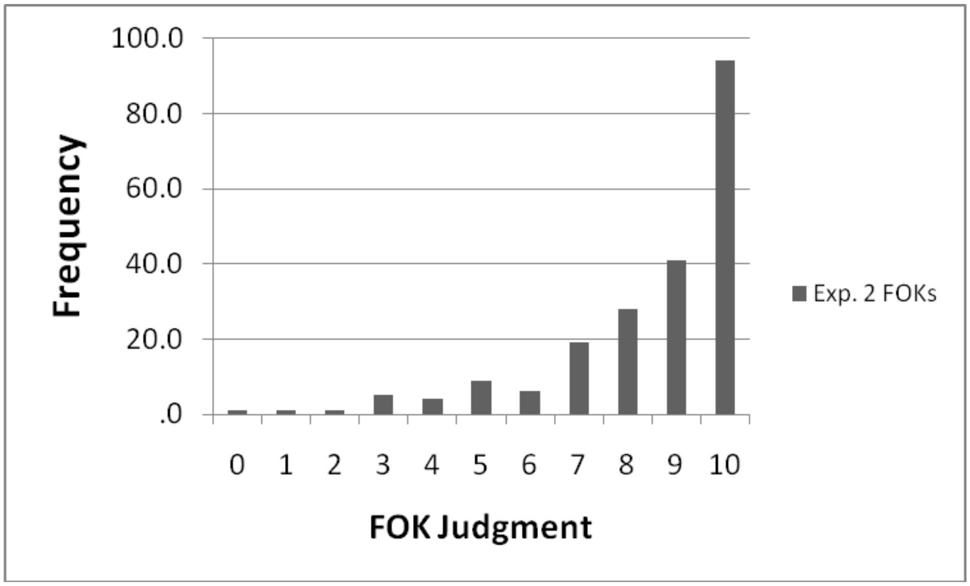


Figure 6. Experiment 2: Distribution of FOK judgments (collapsed across participants) that includes all items for which cued recall was correct.

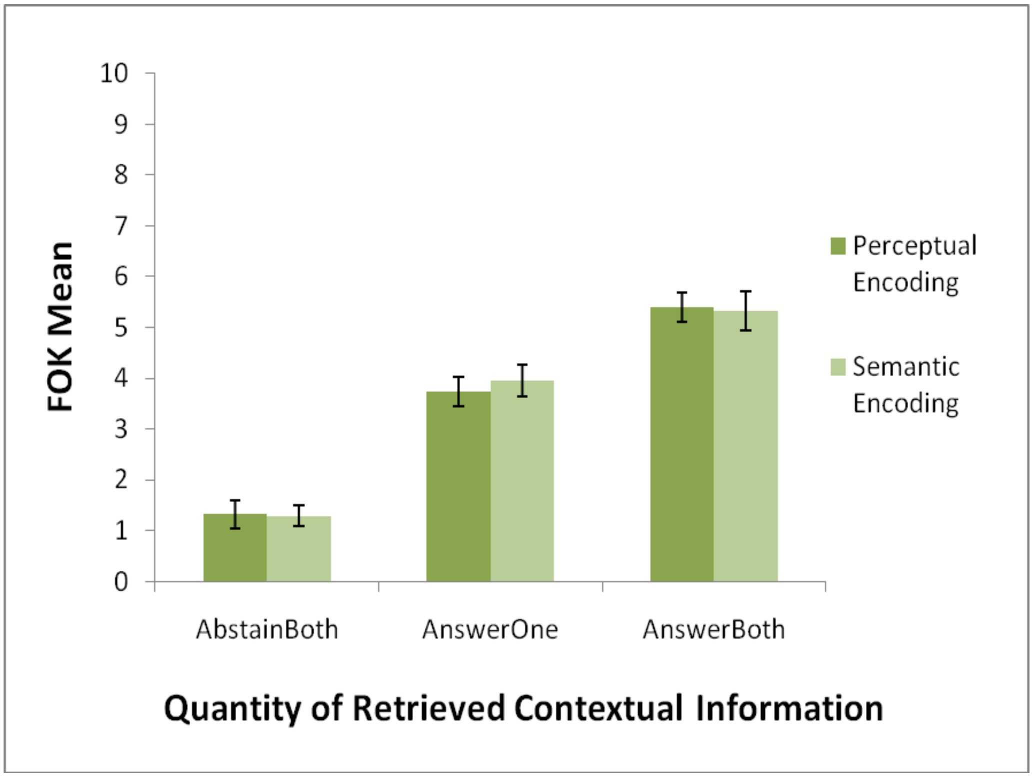


Figure 7. Experiment 2: Mean feeling-of-knowing (FOK) judgments as a function of the quantity of retrieved contextual information.

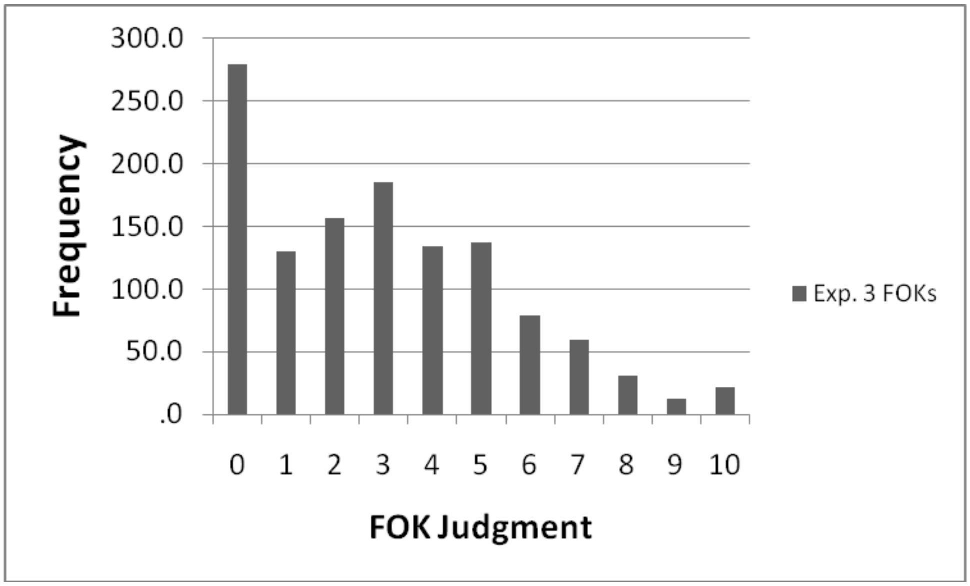


Figure 8. Experiment 3: Distribution of FOK judgments (collapsed across participants) that includes all items for which cued recall was incorrect.

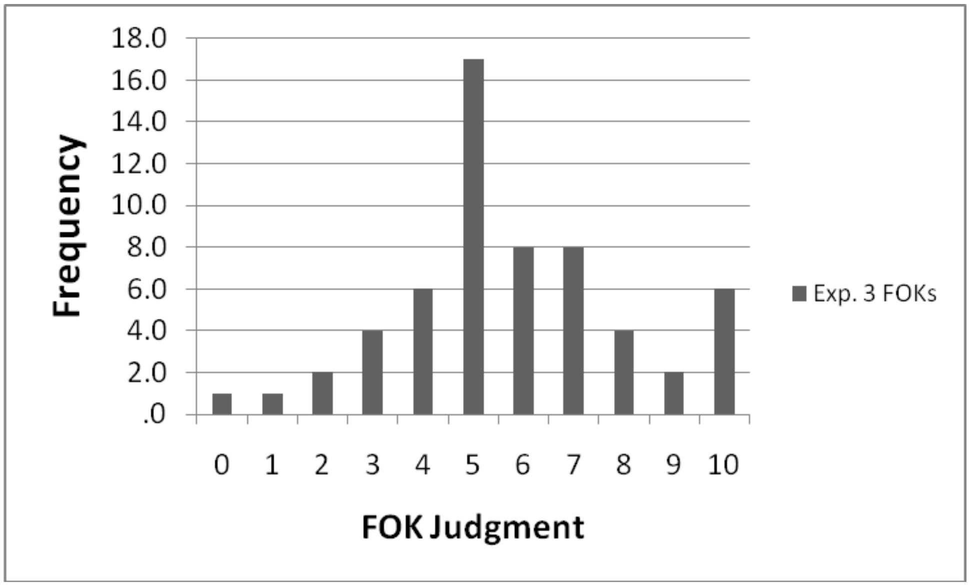


Figure 9. Experiment 3: Distribution of FOK judgments (collapsed across participants) that includes all items for which cued recall was correct.

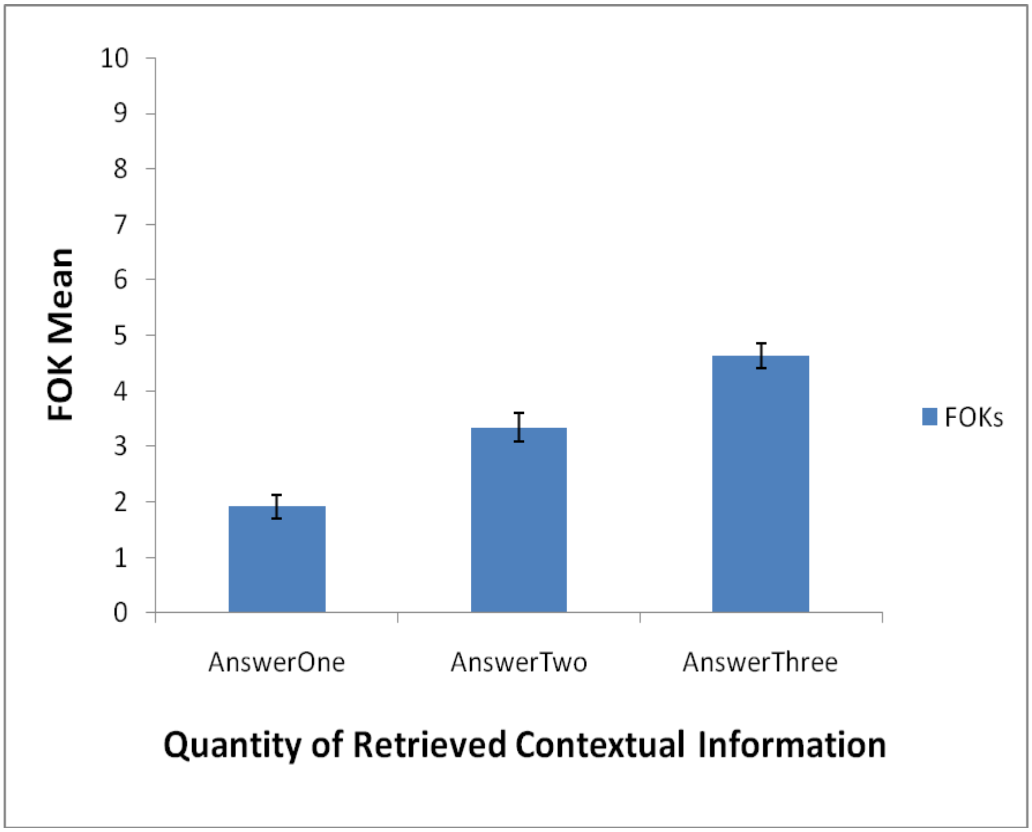


Figure 10. Experiment 3: Mean feeling-of-knowing (FOK) judgments as a function of the quantity of retrieved contextual information.

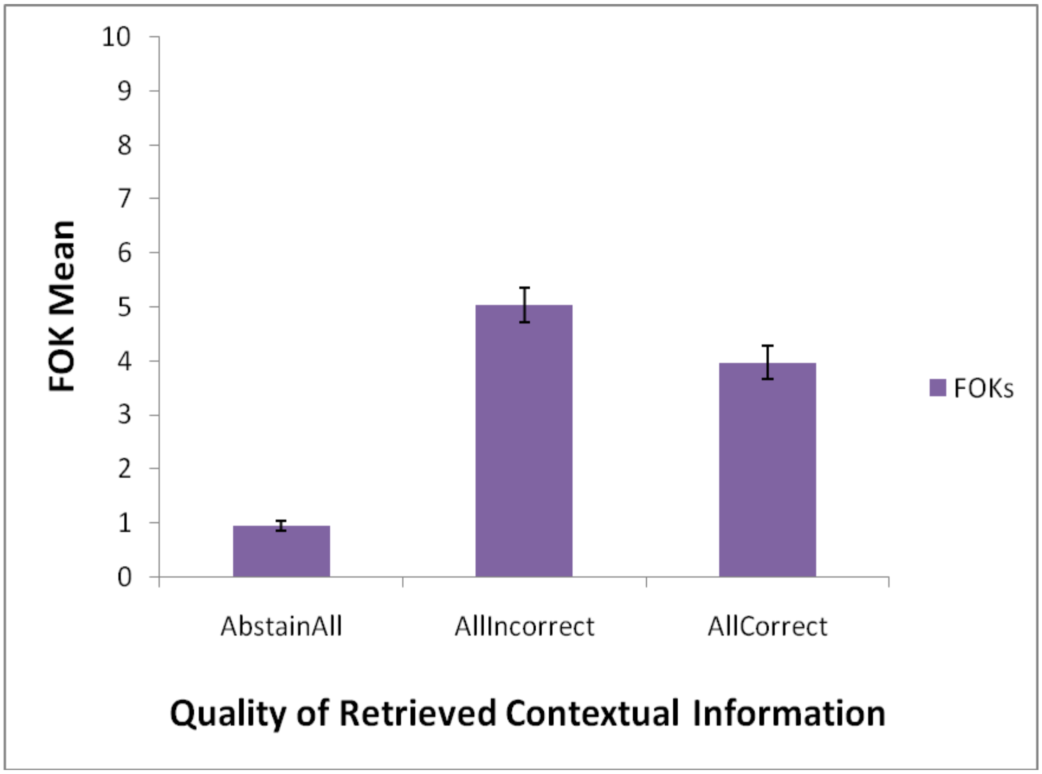


Figure 11. Experiment 3: Mean feeling-of-knowing (FOK) judgments as a function of the quality of retrieved contextual information.

References

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445-459. Retrieved from http://elexicon.wustl.edu/default.asp.

Barsalou, L. W. (1992). Cognitive psychology: An overview for cognitive scientists. Hillsdale, NJ, US: Lawrence Erlbaum Associates Publishers.

Blake, M. (1973). Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. *Journal of Verbal Learning and Verbal Behavior, 12*(3), 311-319.

Brown, R., & McNeill, D. (1966). The 'tip of the tongue' phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325-337.

Burgess, C., & Lund, K. (2000). The dynamics of meaning in memory. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (pp. 117–156). Mahwah, NJ: Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Costermans, J., Lories, G., & Ansay, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(1), 142-150.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*(6), 671-684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268-294.

Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language, 52*. 551-565.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General, 116*(2), 119-136.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*(4), 208-216.

Hart, J. T. (1967). Second-try recall, recognition, and the memory-monitoring process. *Journal of Educational Psychology, 58*(4), 193-197.

Intraub, H., & Nicklos, S. (1985). Levels of processing and picture memory: The physical superiority effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(2), 284-298.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*(4), 609-639.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, *124*(3), 311-333.

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 34-53.

Koriat, A., Levy-Sadot, R., Edry, E., & de Marcas, S. (2003). What do we know about what we cannot remember? Accessing the semantic attributes of words that cannot be recalled. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1095-1105.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers*, *28*(2), 203-208.

Lupker, S. J., Harbluk, J. L., & Patrick, A. S. (1991). Memory for things forgotten. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 897-907.

Maki, R. H., & Serra, M. (1992). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 116-126.

Marks, William. (1991). Effects of encoding the perceptual features of pictures on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(3), 566-577.

Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 288-294.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 851-861.

Nairne, J. S., & Pandeirada, J. N. S. (2008). Adaptive memory: Remembering with a stone-age brain. *Current Directions in Psychological Science, 17*(4), 239-243.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Retrieved from http://www.usf.edu/FreeAssociation/.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109-133.

Nelson, T.O., Gerler, D., & Narens, L. (1984). Accuracy of feeling-of knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General, 113*(2), 282-300.

Nelson, T. O., Leonesio, R. J., Shimamura, A. P., Landwehr, R. S., & Narens, L. (1982). Overlearning and the feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 279-288.

Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, *89*(3), 250-280.

Reder, L. M., & Ritter, F.E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 435-451.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object set: The role of surface detail in basic-level object recognition. *Perception, 33*, 217-236.

Schacter, D.L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(1), 39-54.

Schacter, D.L., & Worling, J.R. (1985). Attribute information and the feeling-of-knowing. *Canadian Journal of Psychology, 39*(3), 467-475.

Schwartz, B.L. (1994). Sources of information in metamemory: Judgments of learning and feeling of knowing. *Psychonomic Bulletin & Review, 1*(3), 357-375.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 1074-1083.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81*(3), 214-241.

Thomas, A.K., Bulevich, J.B., & Dubois, S.J. (2010). Value of context affects the episodic feeling of knowing when young adults are given enough time and older adults are given enough direction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the norms. *Journal of Memory and Language*, *50*(3), 289-335.