An analysis of the determinants of the feeling of knowing

Ayanna K. Thomas [a,*], John B. Bulevich [b], Stacey J. Dubois [a]

[a] *Tufts University, United States*
[b] *Richard Stockton College of New Jersey, United States*

ARTICLE INFO

ABSTRACT

Research has demonstrated that feeling-of-knowing (FOK) judgments are affected by the amount of accessible information related to an inaccessible target. Further, studies have demonstrated that, in some situations, FOK judgment magnitude is not only related to the amount (quantity) of accessed features, but also the correctness of those features (Thomas, Bulevich, & Dubois, 2011). The present study examined the conditions under which the correctness of features would influence FOK judgment magnitude. We hypothesized that accuracy of retrieved features would influence FOK judgments, but only in situations where semantically meaningful information was accessible. In three experiments, we manipulated accessibility of semantic information. In all experiments, the quantity, or amount of retrieved partial information had a greater impact on FOK judgments than the accuracy of that information. However, in situations where semantic information was accessible, accuracy of retrieved semantic features also influenced FOK judgment magnitude, and later recognition.

Published by Elsevier Inc.

1. Introduction

Human memory is an extraordinarily dynamic and flexible system. People cannot only retrieve a stored image, but also estimate the likelihood of successful retrieval (Tulving & Madigan, 1970). Research examining this *feeling-of-knowing* (FOK) has resulted in a general consensus that these estimations are inferred from a variety of cues. Cues such as knowledge about the general topic (e.g. Connor, Balota, & Neely, 1992; Costermans, Lories, & Ansay, 1992; Glenberg, Sanocki, Epstein, & Morris, 1987; Maki & Serra, 1992), familiarity with the parts of the question (e.g. Koriat & Levy-Sadot, 2001; Metcalfe, Schwartz, & Joaquim,1993; Reder & Ritter, 1992; Schwartz & Metcalfe, 1992), attributes or information related to the target (e.g. Blake, 1973; Eysenck, 1979; Koriat, 1993; Schacter & Worling, 1985), the conditions of learning (Lupker, Harbluk, & Patrick, 1991), or, more generally, properties associated with the target (Koriat, 1997), have all be shown to influence FOK judgments. However, because of the inferential nature of FOK judgments, there is discrepancy between FOK estimations and actual memory performance.

The present study investigated the nature of this discrepancy. We hypothesize that FOK judgments do not always accurately predict whether information will later be remembered, because the cues used to make those FOK judgments may sometimes be undiagnostic of later retrievability. In the present study, we examined the hypothesis that conceptually meaningful cues will be more diagnostic of future retrievability and will result in higher feeling-of-knowing predictions than perceptual cues. In typical feeling-of-knowing (FOK) studies, participants are asked to predict the likelihood of future recognition for items which are momentarily inaccessible (e.g. Blake, 1973; Hart, 1965, 1967; Koriat, 1993, 1995; Metcalfe, 1986; Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982; Schacter, 1983). The FOK is made after a failed retrieval attempt.

Researchers have proposed that the FOK judgment process is parasitic on the process of retrieval. Whenever memory is searched, a variety of cues and/or attributes associated with a target come to mind. These may include activation from the

---

terms in the question, structural, contextual, and semantic attributes, or fragments of the target (see Durso & Shore, 1991; Lovelace, 1987; Read & Bruce, 1982). When retrieval fails, these attributes can be used to judge whether the target can be recalled or recognized at a later point (see Koriat, 1993). Thus, if few or no information is retrieved, then the FOK judgment should be low. According to the accessibility model of the feeling-of-knowing (cf., Koriat, 1993), the accuracy of these attributes does not directly influence FOK judgment magnitude. Rather, FOKs are made based on the amount, or quantity of retrieved attributes (Koriat, 1993). The more information accessed, the higher the FOK (Koriat, 1993, 1995; Koriat & Levy-Sadot, 2001). FOKs are relatively accurate primarily because the attributes accessed are more likely to point to the actual target in memory, and be correct. The present study investigated the hypothesis that conceptual attribute accuracy will affect FOK judgment magnitude. That is, we examine whether FOKs will be higher when correct as compared to incorrect attributes are retrieved. When attributes are perceptual, the accuracy of those attributes will be less critical to FOK judgment magnitude. In these cases, FOK judgment magnitude will be influenced by the amount, or quantity of retrieved attributes, but the accuracy of those cues will not influence FOK judgments.

Our primary hypothesis is based in part on recent studies that have demonstrated that accuracy of information associated with an inaccessible target influenced 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 (2011) 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. (2011) 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 utilizing a controlled recollective process. Finally, research has demonstrated that information about semantic attributes of words were accessible even when word recall failed (Koriat, Levy-Sadot, Edry, & de Marcas, 2003). Specifically, Koriat et al. (2003) asked participants to provide ratings 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. 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.

The metacognitive decisions made in both Dunlosky et al. (2005) and Thomas et al. (2011) were associated with conceptual or meaningful stimuli. The attributes used to make those predictions were semantic in nature. Finally, when retrieval of the target failed, Koriat et al. (2003) demonstrated that participants were still able to accurately assess a variety of conceptual attributes associated with the un-recallable target. In the present study, we hypothesized that when participants are able to access conceptual attributes, the accuracy of those attributes will influence FOK judgment magnitude. We argue that semantic stimuli are composed of numerous conceptual features that can be recalled (Underwood, 1969) even when recall of the target fails. Research suggests that when recall fails, these attributes can be recalled correctly (Koriat et al., 2003). Thus, in situations where people have access to semantic attributes, the accuracy of those attributes should directly influence FOK judgments.

Semantic attributes are inherent to verbal material. Various dimensions of this general inherent attribute have been shown to affect both metamemorial predictions and memory performance. For example, metamemory predictions have been shown to be influenced by noun concreteness (Begg, Duft, Lalonde, & Melnick, 1989; Groninger, 1979; Koriat, 1997; Koriat & Bjork, 2005, 2006; Moulin, Perfect, & Jones, 2000), and by associative strength between cue-target word pairs (Koriat & Bjork, 2005). These cues, inherent to verbal material, are perceived to disclose some information regarding the difficulty associated with learning and later remembering (Koriat, 1997). People may perceive concrete nouns to be more memorable than abstract nouns. That perception is not only correct, but appropriately influences metamemorial predictions. Research has also demonstrated that extrinsic factors, such as the conditions of learning or the operations employed by participants during learning can impact metamemorial assessment. For example, the depth at which information is processed at encoding has been shown to affect participants' FOK judgments. 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 ratings to items in the deep condition as compared to ones in the shallow condition. We hypothesized that deep encoding processes would affect the accessibility of semantic attributes. Thus, access to the associated semantic attributes may be what supports the finding of greater FOK accuracy after deep processing. In the present study we examined whether semantic features that were inherent to the studied material would be similarly influential on FOKs as semantic features that may have been extrinsically accentuated by the conditions of learning. To explore these questions we examined situations in which semantic features could be easily extracted from the stimuli, as well as situations where semantic features were emphasized by the encoding situation.

Across three experiments, we varied the accessibility of semantic features in two ways: by using stimuli that varied in inherent meaningfulness, and by using levels-of-processing (LOP) instructions at encoding. Participants encoded concrete word pairs, nameable picture pairs, or ambiguous picture pairs. Incidental encoding procedures were used in the first two experiments in order to compare deep to shallow encoding. In Experiment 3, intentional encoding procedures were used. During the FOK phase, participants had a chance to provide both perceptual and semantic attributes related to the unrecalled target. Therefore, we were able to examine the effects of intrinsically and extrinsically instantiated conceptual processing of the to-be-remembered items as well as the meaningfulness of retrieved attributes on FOK magnitude.

We hypothesized that attribute recall accuracy would influence FOK judgments when those attributes were semantic in nature. That is, FOK judgments would be higher when participants retrieved correct semantic attributes. In addition, we hypothesized that a focus at encoding on deeper or more conceptual processing, would affect the accessibility of specific attributes. Numerous studies have demonstrated that deeper levels of processing result in better recall (e.g. Craik & Tulving, 1975). This finding implies that a richer, more detailed memory trace is formed under deep, as opposed to shallow, encoding conditions. Thus, we hypothesized that deeper encoding processes would result in better access to semantic attributes, even in conditions where to-be-remembered stimuli were inherently perceptual. Finally, we hypothesized that in conditions where semantic attributes were not associated with targets, amount of attribute information would predict FOK judgment magnitude; however the accuracy of those attributes would be irrelevant.

2. Experiment 1

In Experiment 1, participants studied unrelated word pairs under either deep or shallow encoding conditions. After study, participants were given a cued-recall test. When target recall failed, participants were asked about several attributes associated with targets, and asked to provide FOK judgments. Specifically, participants were asked to recall whether the font of the target was the same as that of the cue (a perceptual attribute). In addition, participants were asked to recall whether the category of the target was the same as that of the cue (a conceptual attribute). After attribute assessment, participants were asked to judge how likely they would be to recognize the target on a later recognition test of memory (the FOK question). We hypothesized that participants would provide high FOK judgments in association with correctly remembered conceptual attributes. Further, we hypothesized that the effects of attribute accuracy would be stronger in the deep encoding condition, where participants were encouraged to attend to and extract inherent semantic features.

2.1. Method

2.1.1. Participants

We analyzed data from 46 undergraduates (ages 18–23). 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 as compensation for their participation, respectively. Only individuals with normal color vision were eligible to participate.

2.1.2. Materials

Data were collected using E-Prime v.1.1 software (Psychology Software Tools Inc., Pittsburgh, PA), on Dell desktop computers. For Experiment 1, 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 constructed 60 cue-target pairs (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. Words were selected to be moderately frequent in the English language (as determined by The English Lexicon Project's website, Balota et al., 2007, http://elexicon.wustl.edu/default.asp) and to be low in both forward and backward associative strength (determined by the University of South Florida Free Association Norms, Nelson, McEvoy, & Schreiber, 1998).

For the recognition test, we used a four-alternative forced-choice test. Distractors were selected from the total set of cues and targets used during the study phase. Each cue and target item appeared as an answer choice twice. In addition, 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. Cues were unrelated (as measured by both forward and backward associative strength) to answer choices on each recognition trial.

2.1.3. Procedure

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

We manipulated levels of processing at encoding within participants. 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, 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). 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 the word pairs into two equivalent blocks of items, and encoding depth for each block was counterbalanced across participants. Presentation of the stimuli pairs was randomized during encoding. On each trial, participants viewed an encoding question for 3 s, and then a cue-target pair appeared on the screen. After a 1-s delay, a response box appeared for participants to answer the encoding question. After participants completed all of the trials in the encoding phase, they played Tetris on the computer for 5 min.

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. Cues were presented in the same font style and font color as seen during encoding. After either 8 s or a response from the participant, we gave participants an opportunity to provide contextual information relating to the target. The perceptual (shallow) question was, "What was the font color of the target?" and the semantic (deep) 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." 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 feature in question. Presentation order for the questions was counterbalanced. 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 a 10-point Likert scale (1 = *I definitely will NOT be able to recognize the target word/image*; 10 = *I definitely WILL be able to recognize the target word/image*).

Finally, participants 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. 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, and participants responded using the number keys on the keyboard. Following the experiment, participants were debriefed and thanked for their participation.

2.2. 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 almost always equaled 10 and the items were always recognized on the final test.

2.2.1. Encoding orientation

We calculated two feeling-of-knowing (FOK) rating 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 ratings 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* = 0.73) than items that had been encoded shallowly (*M* = 0.52), *t*(45) = 8.12, *d* = 1.50, thus demonstrating a standard levels-of-processing effect (e.g. Craik & Tulving, 1975).

2.2.2. The relationship between FOKs and accuracy of attributes

Our first hypothesis was that FOK judgment magnitude would be higher when conceptual attribute memory, as opposed to perceptual attribute memory, was correct. To evaluate this hypothesis we calculated mean FOKs for conditions in which only the font attribute question was answered correctly, for which only the category question was answered correctly, and for which both questions were answered correctly. Items for which participants abstained on either attribute question were excluded from this analysis. In addition, items for which participants answered both questions incorrectly were not included in this analysis, because there were too few observations of this kind within each condition. One participant was excluded from this analysis entirely because there were no trials for which this participant answered both questions. To determine whether FOK judgment magnitude would be affected by encoding condition and type of correct attribute retrieved, we conducted a 2 (Encoding Orientation: shallow, deep) × 3 (Question Type: only font color correct, only category correct, both correct) repeated measures ANOVA on participants' mean FOK judgments.[1] There were significant main effects of both Encoding

---

[1] While these statistical procedures are atypical, they are often used in the FOK literature. The primary goal of this and other research in this area is to examine the relationship between attribute memory and FOK judgments. This is conceptually similar to examining the relationship between item memory and FOK judgment magnitude. As such, we examined FOK judgment magnitude as a function of the accuracy of attributes, just as we would examine FOK judgment magnitude as a function of the accuracy of retrieved items. The question of interest is whether FOK judgments are higher when attributes are correct, just as the question could be whether FOK judgments are higher when item memory is correct.

Table 1
Exp. 1 FOK means by encoding orientation and attribute accuracy. Standard deviation, samples for each mean are also presented.

Contextual information accuracy	Shallow encoding		Deep encoding	
	M (SD)	n	M (SD)	n
Only font correct	3.74 (1.35)	33	5.37 (1.86)	33
Only category correct	4.59 (1.62)	44	6.41 (1.31)	45
Both questions correct	5.66 (1.86)	35	6.69 (1.59)	38

Orientation, $F(1,44) = 74.87$, $p < .001$, and Question Type, $F(2,88) = 31.52$, $p < .001$. As Table 1 demonstrates, 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 question was correct ($M = 4.54$) as opposed to when only the category question was correct ($M = 5.49$), $t(44) = 5.29$, $d = 1.22$. Mean FOK judgments were highest when both questions were correct ($M = 6.14$) [category only vs. both correct: $t(44) = 3.06$, $d = 1.41$]. There were no other significant effects, $F < 1$. It is important to note that traditional statistical procedures could not be applied to these results because of the "fragmentary data problem" characteristic of studies on partial recall (see Brown, 1991; Brown & McNeill, 1966). Specifically, each cell is based on a different combination of participants and items.[2] It should also be noted that, because this analysis was based on accurately retrieved attributes, the number of observations that were used to generate FOK means differed as a function of condition. On average, there were fewer observations per participant when only the font color question was answered correctly ($M_{shallow} = 3.88$; $M_{deep} = 3.00$), as compared to the category question ($M_{shallow} = 5.66$; $M_{deep} = 10.53$).

In our second analysis we examined whether FOK judgment magnitude would be affected by accuracy of retrieved attribute. In order to conduct this analysis, we computed average FOKs for instances where participants answered the font question correctly or incorrectly, but did not answer the category question. Similarly, we computed average FOKs for instances where participants answered the category question correctly or incorrectly, but did not answer the font question. Again, we could not compute instances for which participants answered both questions incorrectly, because of limitations in number of observations. Four participants were excluded from this analysis entirely because there were no trials for which these participants answered only one question. A 2 (Encoding Orientation: shallow, deep) × 2 (Question Accuracy: correct, incorrect) × 2 (Question Type: font, category) within-subjects ANOVA found main effects of Encoding and Question Type, [$F(1,41) = 47.89$, $p < .001$; $F(1,41) = 25.32$, $p < .001$]. As with the previous analysis, FOK judgments were higher when participants engaged in deep encoding ($M = 4.45$) as compared to shallow encoding ($M = 3.47$). In addition, as Table 2 demonstrates, average FOKs were higher when participants answered the category question ($M = 4.43$) as compared to the font question ($M = 3.49$). Finally the interaction between Question Type and Question Accuracy was significant, $F(1,41) = 6.67$. FOK judgments were similar whether the font attribute was remembered correctly or incorrectly; however, FOK judgments were significantly higher when the category question was remembered correctly ($M = 4.81$) as opposed to incorrectly ($M = 4.05$). As with the first accuracy analysis, in the present analysis, observations per participant differed for question type, question accuracy, and encoding. On average there were more observations for the category question ($M = 3.04$) as compared to the font question ($M = 2.16$). There were more observations for deep encoding ($M = 2.75$) as compared to shallow encoding ($M = 2.44$). Finally, there were more observations when participants provided correct feature information ($M = 3.31$) as compared to incorrect information ($M = 1.89$).

2.2.3. The relationship between FOKs and recognition

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. Gamma correlations are the statistical measurement of choice in metacognitive research for analyzing the relationship between predictions and criterion performance (Nelson, 1984). 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, $t(44) = 1.30$, $p > .05$. 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.

2.3. Experiment 1: Discussion

The primary focus of the present study was to examine the relationship between accuracy of attributes of unrecallable targets and FOK judgment magnitude. We found that participants were more likely to provide correct category information as compared to correct font information. We also found that more correct attribute information was provided after deep

---

[2] While it may seem unorthodox to use a dependent variable (i.e., accuracy of attribute information) as a variable in an ANOVA, this is a common practice in FOK research. The primary motivation of the analysis is to examine how FOK judgment magnitude changes depending on whether attribute information associated with the target in question is correct or incorrect. This is analogous to examining any kind of metacognitive assessment associated with correct and incorrectly remembered information. We see similar practices for studies using remember/know (Mather, Henkel, Johnson, 1997) confidence (Holmes, Waters, Rajaram, 1998) and FOK (Koriat, 1993; Thomas et al., 2011).

Table 2
Exp. 1 FOK means by encoding orientation, question type, and question accuracy. Standard deviation, samples for each mean, and number of means replaced, are also presented.

Question type and accuracy	Shallow		Deep	
	M (SD)	n	M (SD)	n
Font incorrect	3.86 (1.98)	31	3.18 (1.64)	28
Font correct	4.06 (1.72)	27	2.86 (1.52)	31
Category incorrect	4.45 (2.15)	30	3.64 (1.62)	28
Category correct	5.42 (1.94)	38	4.22 (1.92)	35

encoding than after shallow encoding. The present experiment extends the previous findings by demonstrating that the effect of accuracy may be contingent on how the information is encoded and what type of attribute information is retrieved. That is, when participants deeply encoded paired associates, accuracy of features was more important to average FOKs than when participants focused on perceptual information at encoding.

Results from Experiment 1 are consistent with Koriat et al. (2003) in which participants were shown to have access to semantic attribute information even in the absence of target information. The finding that FOK judgments were higher when correct semantic information was retrieved is suggestive of an evaluation mechanism. That is, participants may have deliberately sought additional information in order to ascertain the validity of the semantic attributes. However, Experiment 1 does not provide direct support for this evaluation mechanism. As an alternative, participants may simply access correct semantic features more quickly than incorrect semantic features or perceptual features. Higher FOKs may have directly resulted because of ease of accessibility. Shallow processing was designed to influence access to perceptual attributes; however, research suggests that for inherently meaningful stimuli, conceptual processing may take precedence over perceptual processing (Booth & Waxman, 2002; Gelman, 2003; Gelman & Markman, 1986). Thus, even in a shallow encoding manipulation, the initial processing of semantic attributes of inherently meaningful stimuli may have disrupted the effective processing of perceptual attributes. Experiment 2 was designed to improve access to perceptual attributes and determine whether correctness of those attributes would prove important to FOK judgment magnitude.

3. Experiment 2

In Experiment 2, we examined whether the accuracy of perceptual features would be more important to FOK judgment magnitude when participants studied material that were rich in perceptual detail. According to the materials appropriate processing (MAP) framework, different types of materials will encourage encoding of different types of information (McDaniel, Einstein, Dunay, & Cobb, 1986). By presenting pictures, we sought to influence access to perceptual attributes. Thus, we hypothesized that the accuracy of perceptual features would be more important to FOK judgment magnitude when the inaccessible target was a picture. Our hypothesis was supported by research demonstrating that memory for pictures was often enhanced when participants encode perceptually as opposed to conceptually (Intraub & Nicklos, 1985; Marks, 1991). That is, access to perceptual attributes, as compared to conceptual attributes, has a greater influence on picture memory. It follows then that access to perceptual attributes will also have a greater influence on FOK judgment magnitude. Experiment 2 sought to determine: (1) whether the accuracy of retrieved attributes (perceptual and conceptual) would affect participants' feelings of knowing in judgments for pictorial stimuli, (2) if so, whether accuracy of semantic features or perceptual features would be more influential, and (3) how these variables might interact with encoding orientation.

3.1. Method

3.1.1. Participants
We analyzed data from 31 undergraduates (ages 18–23) for Experiment 2. Participants were recruited as described for Experiment 1.

3.1.2. Materials
We constructed 48 picture pairs using concrete images from Rossion and Pourtois' (2004) normed set. 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. We used the University of South Florida Free Association Norms (Nelson et al., 1998) to examine forward and backward association strength between the cue-target labels (i.e. Tree–Hammer), and we determined that verbal labels for pictures pairs were low in associative strength. That said, pairs from the same category were more closely related than pairs from different categories.

Answer options on the four-alternative forced-choice test had served as cues and targets in the study phase of the experiment. A within-category lure was used on every recognition trial. Thus, each recognition trial consisted of the target, a lure

Table 3
Exp. 2 FOK means by encoding orientation and feature accuracy. Standard deviation, samples for each mean, and number of means replaced, are also presented.

Contextual information accuracy	Perceptual encoding		Semantic encoding	
	M (SD)	n	M (SD)	n
Only Texture Correct	4.22 (1.82)	31 (15)	5.19 (1.51)	31 (11)
Only Category Correct	5.77 (2.34)	31 (5)	5.28 (2.21)	31 (2)
Both Questions Correct	5.73 (1.81)	31 (2)	5.62 (1.81)	31 (2)

from the same category as the cue, and two additional lures that belonged to different categories than the cue. Once again, we used the University of South Florida Free Association Norms (Nelson et al., 1998) to examine forward and backward association strength between the cues and lures, as well as between the targets and lures, to ensure that all possible combinations formed pairs low in associative strength.

3.1.3. Procedure

Experiment 2 was a 2 (Encoding Orientation: perceptual vs. semantic) x 2 (Attribute Question: texture vs. category) within-subject factorial design. The procedure of Experiment 2 was similar to that used in Experiment 1, except for the following. 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 (see Intraub & Nicklos, 1985), 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 features related to the target. The perceptual feature question was, "Was the target more textured than the cue?" and the semantic feature 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 was identical to that described in Experiment 1; however, in this experiment, we also included a post-recognition phase. Because judgments of "texture" are subjective, it was necessary to collect these judgments from each participant for all 48 pairs in order to evaluate the accuracy of this specific attribute question. 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.

3.2. Results

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. As in Experiment 1, overall recognition (M = 0.60) was significantly better than chance (0.25), t(30) = 11.76.

3.2.1. Encoding orientation

We calculated two feeling-of-knowing (FOK) rating means for each participant, one for items that had been encoded perceptually and the other for items that had been encoded semantically. Unlike in Experiment 1, average FOK ratings were not affected by encoding orientation, t(30) = 1.35, p > .05. In the semantic condition, participant's FOK ratings averaged 4.07. Ratings averaged 4.26 in the perceptual condition. Neither cued recall performance t(30) < 1, nor recognition performance was affected by encoding condition, t(30) = 1.69, p > .05. Average recognition hit rates in the perceptual condition was 0.58 and in the semantic condition was 0.62.

3.2.2. The relationship between FOKs and accuracy of attributes

To evaluate the hypothesis that accuracy of specific attributes associated with unrecallable targets affected FOK judgment magnitude, we conducted a 2 (Encoding Orientation: perceptual, semantic) × 3 (Question Accuracy: texture correct, category correct, both correct) repeated measures ANOVA on participants' mean FOK judgments. As in Experiment 1, pairs for which participants abstained on either attribute question were eliminated from this analysis. In addition, there were too few observations within each condition to compute average FOKs in association two incorrect attribute answers. As Table 3 illustrates, there was a significant main effect of Question Accuracy, F(2,60) = 7.13, p < .01. Mean FOK judgments were lower only when the texture question was correct (M = 4.56) as opposed to only when 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 = .55. There were no other significant effects, Fs < 1. It should be noted that, once again, the number of observations that were used to generate FOK means differed as a function of condition and question answered. On average there

were fewer observations when participants only answered the texture question correctly ($M_{shallow}$ = 1.94; M_{deep} = 2.15). There were more observations when participants answered the category question correctly ($M_{shallow}$ = 3.38; M_{deep} = 3.72).

We scored the accuracy of participants' responses to the texture 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 question, a 2 (Encoding Orientation: perceptual, semantic) × 2 (Contextual Information Question: texture, category) × 2 (Question Accuracy: correct, incorrect) ANOVA was conducted using mean FOK ratings as the dependent variable. The ANOVA revealed a significant Question x Accuracy interaction, $F(1,25)$ = 4.86. Planned comparisons indicated that, when participants answered the texture question, their FOK ratings were similar in magnitude regardless of whether they answered that question correctly (M = 3.46) or incorrectly (M = 3.87), $t(24)$ = 1.16, p > 0.05. However, when participants answered the category question, their FOK ratings were higher when they answered that question correctly (M = 4.87) as opposed to incorrectly (M = 3.14), $t(24)$ = 2.87, d = 0.73. Thus, when participants answered only one question, the accuracy of the retrieved feature influenced mean FOK ratings, but only when the feature was semantic in nature. These findings are consistent with Experiment 1. No other effects were significant, Fs < 1.

3.3. Experiment 2: Discussion

In Experiment 2, we tested whether correct recall of perceptual attributes would influence FOKs if to-be-remembered stimuli contained more salient perceptual details. In contrast to our predictions, we found that FOKs were higher when category information was correctly remembered as compared to perceptual information, replicating the findings of Experiment 1. In addition, we found that quality of perceptual attributes had little influence on FOK judgment magnitude. These results suggest that perceptual attributes may be less relevant for FOK judgment magnitude than conceptual attributes. Alternatively, our manipulation of perceptual attribute salience may not have been sufficiently strong. While pictures encompass a richer set of perceptual details than words, the picture pairs used in the present experiment were those of nameable concrete nouns. With easy access to conceptual/semantic attributes, perceptual attributes may have played a smaller role on FOK decisions. Perceptual attribute salience may have also been negatively impacted by type of test used in the FOK judgment phase.

Common in the FOK literature, and employed in Experiment 2, was a cued recall test. This type of test may have encouraged participants to search for conceptual, or semantic, features, rather than perceptual features. That is, while participants studied pictures rich in perceptual attributes, when initially tested, they were required to retrieve the semantic labels associated with those pictures. In employing a FOK procedure that has been standardized in the literature, we may unwittingly have biased the retrieval of attributes associated with unrecallable targets. According to Transfer Appropriate Processing (TAP; Morris, Bransford, & Franks, 1977), manipulations that encouraged a particular type of processing should improve memory when the processing invited by the test is congruent with that of study. Initial retrieval, during the cued recall phase, may have invited conceptual, or semantic, processing, influencing access to conceptual attributes rather than perceptual attributes. Accuracy of conceptual attributes remained important to both FOK judgments and recognition performance in this experiment, because while the studied materials were perceptually rich, they were also conceptually rich, as each picture pair depicted nameable concrete nouns. The accuracy of perceptual attributes may have played a larger role had the initial retrieval had been one that encouraged a search for perceptual features.

4. Experiment 3

By using a cued-recall task that required participants to semantically label pictures during the attribute assessment and FOK phase, participants may have automatically engaged in a search of conceptual attributes. This search may have minimized the impact of perceptual features. The goal of Experiment 3 was to examine whether the accuracy of retrieved perceptual attributes would affect FOK judgments under conditions where inherent meaning of stimuli was minimized, and when testing was based on perceptual features as opposed to conceptual ones. In order to test this hypothesis, we developed stimuli that lacked inherent semantic attributes. Participants were presented with object pairs, and then given a closed-set cued-recall test. Specifically, for each cue, participants had to select the target from the entire array of presented objects. We hypothesized that the retrieval processes involved in this task would capitalize on perceptual matching as opposed to conceptual generation. If target retrievability failed, participants were then asked two perceptual attribute questions and provided FOK judgments. A final three-alternative forced choice test followed. We predicted that in the absence of useful conceptual attributes, the value of perceptual attributes would become apparent, and the accuracy of those attributes would exert an influence of FOK judgment magnitude. Finally, in Experiment 3 we eliminated the levels-of-processing manipulation, as are goal was to minimize the accessibility of any possible conceptual attributes.

4.1. Method

4.1.1. Participants
Twenty-four participants were included in Experiment 3.

4.1.2. Materials

Forty ambiguous figures were created from basic shapes using the liquid vector tool in Real-DRAW Pro, (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 A 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. 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 selected 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 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. In addition, review of the completed questionnaires revealed that participants did not systematically rely on any one specific strategy for eliminating lures on the recognition task.

4.1.3. Procedure

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 simply study the pairs. Participants were told that they would have to remember the target when given the cue. 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 s 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 two attribute information questions: "What was the color of the target?" (red, blue, green, abstain); "Was the target larger than the cue?" (yes, no, abstain). The order of the questions varied across participants. After responding to the attribute questions, participants were asked to rate their feeling of knowing.

The final recognition phase of the experiment was executed as described for Experiments 1 and 2, with the exception that participants in Experiment 3 completed a three-alternative forced-choice recognition test instead of a four-alternative forced-choice test.

4.2. Experiment 3: Results

Alpha was set at 0.05 for all analyses. Items for which participants produced the correct response during the closed cued recall phase of the experiment (4.54%) were excluded from the following analyses. As in Experiments 1 and 2, overall recognition ($M = 0.51$) was significantly better than chance (0.33), $t(23) = 9.68$, $p < .001$.

4.2.1. The relationship between FOKs and accuracy of attributes

To evaluate the hypothesis that accuracy of specific attributes associated with unrecallable targets affected FOK judgment magnitude, we conducted a One- way ANOVA (Question Accuracy: size correct, color correct, both correct) on participants' mean FOK judgments. This analysis did not yield a significant main effect, $F < 1$. As in the previous experiments, pairs for which participants abstained on either attribute question were eliminated from this analysis. In addition, a computation of FOKs for when participants were wrong on both partial information questions could not be computed. To examine the effect of accuracy of attribute information on FOK judgment magnitude we compared average FOK judgments associated with correctly remembered and incorrectly remembered attribute information. An FOK average was calculated based on

judgments given when the size attribute question was answered correctly *and* the color attribute question was not answered (M = 3.60). Similarly, we compared average FOKs when the color attribute question was answered correctly and the size attribute question was not answered (M = 4.05). A 2 (Attribute: Size, Color) × 2 (Accuracy: Correct, Incorrect) within subjects ANOVA did not find effects of attribute or accuracy, and did not find an interaction between attribute and accuracy, Fs < 1.70.

Because accuracy of attributes had no effect on FOK judgment magnitude, in this experiment, we also examined the effect of quantity, or amount, of retrieved attributes on FOK judgment magnitude. To examine the effects of feature *quantity*, FOK judgment means were computed for three groups of items for each participant: (a) items for which participants chose the "abstain" response for all questions; (b) items for which participants answered one of the questions; and (c) items for which participants answered both questions A repeated measures ANOVA on participants' mean FOK judgments found a significant effect of Quantity, $F(2,34)$ = 14.47, p < .001, indicating that average FOK judgments increased as the amount of retrieved attribute information increased. Planned comparisons indicated that average FOK judgments were lower when participants answered only one question (M = 1.83) than when they answered two questions (M = 3.48), $t(26)$ = 3.68, d = 2.05. In addition, average FOK judgments were lowest when participants did not answer any questions (M = 1.14), $t(21)$ = 2.38, d = .40.

4.3. Experiment 3: Discussion

In Experiment 3 we tested whether accuracy of perceptual attribute information would relate to FOK judgment magnitude. We hypothesized, and found, that the accuracy of perceptual attributes was irrelevant to FOK judgment magnitude. Importantly, quantity of perceptual attributes did affect FOK judgment magnitude. Average FOK judgments were significantly lower when no attributes were retrieved as opposed to when all attributes were retrieved. As it relates to recognition, when participants abstained on both attribute questions, their recognition scores were above chance levels, yet when they answered both questions incorrectly, recognition fell to chance. That is, when average recognition was the dependent measure in an analysis of quantity, a significant effect of attribute quantity was found, $F(2,34)$ = 3.31, p < .01. Post-hoc analyses indicated that recognition scores were higher when participants either abstained on both questions (M = 0.51) as compared to when they answered all three questions incorrectly, (M = 0.33), $t(22)$ = 2.46, d = .75. This further demonstrated that the quantity of retrieved features influenced FOK judgments in Experiment 3, but these perceptual features were undiagnostic of later recognition performance.

The inverse relationship between attribute quantity and recognition may be analogous to a verbal overshadowing effect (Schooler & Engstler-Schooler, 1990). Research has demonstrated that when people are required to describe perceptual stimuli, later recognition for those stimuli is often impaired. Melcher and Schooler (1996) suggested that the effect of verbal-overshadowing resulted, because participants were forced to use verbal labels that were not meaningfully associated with the perceptual targets. In Experiment 3, the perceptual attributes assessed likely were not meaningfully associated with the targets. If meaningful association was engendered, we may have seen an effect of accuracy of perceptual attributes on FOK judgment magnitude as well as recognition performance. Analogously, when expert wine-tasters were asked to describe wines, and given a later recognition test, they did not demonstrate a verbal-overshadowing effect; however, novice tasters, performed less well on the recognition test after verbal description, than after a non-verbal filler task. Expert tasters likely had established meaningful verbal labels to associate with non-verbal characteristics; therefore verbal labels did not interfere with perceptual recognition (Melcher & Schooler, 1996). We explore these ideas further in the General Discussion.

5. General discussion

The present study was designed to elucidate the conditions under which the accuracy of retrieved features associated with inaccessible targets would influence the magnitude of FOK judgments. We hypothesized that feature accuracy was likely to matter only in situations where *either* the to-be-remembered stimuli or the retrieved features were conceptual in nature. We found that in the case of word pairs and nameable picture pairs, accuracy of conceptual features influenced FOKs. Accuracy of perceptual attributes did not influence FOK judgments. In addition, when access to conceptual attributes was minimized, perceptual attribute accuracy remained irrelevant to FOK judgments. Assuming that information about a target is represented in memory as a collection of attributes (Underwood, 1969), even when a target becomes inaccessible, information about some of its individual features may survive. The present study demonstrated that some of these features may be evaluated in the attempt to assess the feeling of knowing. Thomas et al. (2011) suggested that the importance of accuracy of features related to a momentarily inaccessible target may be dependent on the degree to which conceptual information can be extracted from that target. The present study supported that hypothesis and demonstrated that semantic feature accuracy, but not perceptual feature accuracy affects the magnitude of the FOK.

5.1. Evaluating semantic attributes

Several aspects to the reported data suggested that participants may evaluate retrieved semantic features when making FOK assessments. In Experiments 1 and 2, FOK judgments were higher when participants answered semantic feature questions correctly as opposed to when perceptual feature questions were answered correctly. In addition, average FOK judgments did not differ when examined as a function of correctly and incorrectly reported perceptual features. They did differ as a function of accuracy of semantic features. When stimuli devoid of semantic features were used, FOK judgment magnitude was influenced only by the quantity of retrieved features, but not by the accuracy of those features.

The essential and defining feature associated with a given target is what gives it meaning. Meaning of a word or nameable picture can be represented as a set of semantic features (Koriat et al., 2003; Smith, Shoben, & Rips, 1974). In the context of both feature comparison models (Smith et al., 1974) and spreading activation models (Collins & Loftus, 1975) of semantic memory, those features are in some way associated with the target. Those associated features may be accessed in the event of a failed retrieval attempt. Similar to Koriat's accessibility model (1993), we have demonstrated that it was access to those features which influenced FOK magnitude. Similar to Thomas et al. (2011) we demonstrated that FOK judgment magnitude was influenced by accuracy of specific features.

Ease of access may also have influenced the difference in FOK judgment magnitude seen when correct and incorrect semantic features were retrieved. Semantic features may be accessed more easily than less inherently meaningful features. When we examined the retrieval latencies associated with semantic feature assessment in Experiment 1, we did find that correct feature assessments were made more quickly than incorrect, whereas correct perceptual information was not accessed more quickly than incorrect information. Regardless of whether FOK magnitude was influenced by the evaluation of semantic features or the increased accessibility to correct semantic features, our results suggest that FOK magnitude was not influenced by the accuracy of perceptual attributes. We had hypothesized that perceptual features could potentially be evaluated as relevant attributes, inherent to the target, if the targets were pictures, or if the targets were encoded with an emphasis on perceptual characteristics. However, when nameable or ambiguous pictures were used, FOK magnitude did not differ as a function of accuracy of the perceptual questions. Similarly, the encouragement of perceptual encoding did not affect whether accuracy of perceptual features would influence FOK magnitude.

5.2. Accessing attributes

Research has demonstrated that memory for pictorial detail was enhanced when images were encoded perceptually as opposed to conceptually (Marks, 1991). Further, Intraub and Nicklos (1985) reported a reverse levels-of-processing effect when participants encoded images. Specifically, 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?"). They hypothesized that the physical encoding questions might have directed participants' attention to aspects of the scene that were not processed automatically, resulting in more elaborate or distinctive memory traces and better retention (Intraub & Nicklos, 1985).

In the present study, after engaging in perceptual encoding processes, participants did have better access to perceptual features. Specifically, in Experiment 1, participants answered the font color (perceptual) question (34% of trials) just as often as they answered the category (semantic) question (36% of trials). However, for items that had been encoded deeply, participants answered the category (semantic) question (66% of trials) more often than they answered the font color (perceptual) question (52% of trials). In Experiment 2, for items that had been encoded perceptually, participants answered the perceptual question (78% of trials) more often than they answered the semantic question (71% of trials). For items that had been encoded semantically, the opposite pattern emerged; participants answered the semantic question (68% of trials) more often than they answered the perceptual question (57% of trials). Conceptually similar to a transfer appropriate processing effect (Morris et al., 1977), these results suggest that encoding orientation affected feature access. However encoding that focused on perceptual characteristics did not influence the ease of access to *correct* perceptual information, or whether those specific features would influence the magnitude of the FOK. Researchers have proposed that when perceptual and conceptual information is in competition, conceptual information "takes precedence" over perceptual information (Booth & Waxman, 2002; Gelman, 2003; Gelman & Markman, 1986). That is, when multiple sources of information are accessible, conceptual information (i.e., linguistic labels) are more often used on categorization and generalization tasks than perceptual information (Nosofsky, 1984). In the first two experiments, we may have unwittingly created a situation where perceptual and conceptual information were in competition.

5.3. Using perceptual attributes

When studied material was devoid of conceptual information, FOK judgments were undiagnostic of later recognition. In fact, recognition performance was inversely related to attribute access. We concluded that, without access to conceptual or meaningful units, participants could not make accurate FOK predictions. Our results suggested that FOK judgments are often times imprecise, because they are based on a variety of attributes associated with an inaccessible target. Some of these attributes will be better predictors of later recognition than others. The present study suggests that perceptual attributes are not good indicators of later retrievability, at least for standard measures of memory.

For FOK judgments to be reasonably accurate, the stimuli must be inherently meaningful to the learner. However, meaningfulness does not necessarily have to be semantic in nature. One could certainly generate examples where perceptual qualities would be important for memory and evaluation. The learner of a new language must focus on auditory/articulatory qualities of the information that they need to process. Similarly, for the art historian, the visibility of brush strokes in a painting may be very meaningful. While our final recognition test required the matching of perceptual information, it may not have required it in a way that was meaningful for our participants.

Alternatively, perceptual attribute accuracy may have failed to affect FOK judgment magnitude because of poor overall encoding of the ambiguous pairs. According to the memory constraint hypothesis, (Hertzog, Dunlosky, & Sinclair, 2010) FOK judgment magnitude and accuracy may be dependent on initial learning. If memory strength is low, then the accuracy of information that is accessible when making the FOK judgments may be reduced. Hertzog et al. (2010) demonstrated that

FOK judgment accuracy improved when stimuli were repeatedly presented at encoding or when participants used interactive imagery at encoding. Similarly, Lupker et al. (1991) demonstrated improved FOK judgment accuracy in association with deep encoding. Further, Sacher, Taconnat, Souchay, and Isingrini (2009) demonstrated that when attention was divided at encoding, FOK prediction accuracy suffered. In Experiment 3 recognition performance was above chance, but low, suggesting that participants may not have effectively encoded these figures. When participants are asked to retrieve an attribute to an unrecallable target, they are being asked to engage in a recollective process. They must draw upon context to make retrieve that information and then make their FOK. Previous research has noted that when encoding is not "deep" enough or otherwise insufficient, recollection is impossible (Jacoby, 1991) Deficits in encoding would not only affect access to the target, but also would affect access to attribute information associated with the target (Jacoby, 1991; Johnson, Hashtroudi, & Lindsay, 1993).

Finally, we cannot rule out the possibility that the attributes we chose to examine were not relevant to FOK judgments. The failed retrieval processes likely left behind idiosyncratic attributes, unique to each individual. While accuracy of these attributes may in fact influence FOK judgment magnitude, it is difficult to experimentally evaluate. That said, all experiments in the present study attempted to influence the accessibility of specific attributes. While we were successful in influencing attribute access for conceptual rich stimuli, we may have failed to successful influence attribute access for perceptual stimuli.
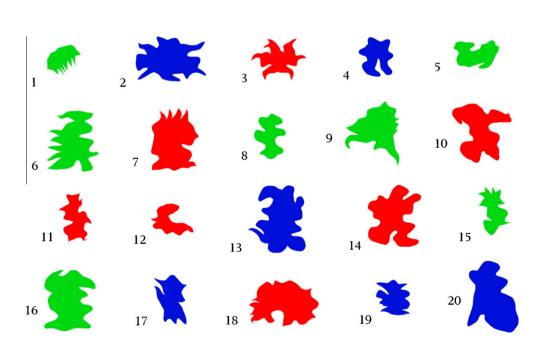
6. Conclusions regarding accessibility and evaluation

Unquestionably, semantic feature accuracy influenced FOK judgment magnitude, and later recognition. These results are consistent with Thomas et al. (2011) who found that accuracy of specific attributes influenced FOK judgment magnitude as well as prediction accuracy. The question remaining is whether semantic attributes are accessed more quickly or are evaluated. Unfortunately, the design of the experiments did not yield response latency data that could be used to thoroughly examine potential accessibility differences. What is clear from these experiments is that accuracy of certain features associated with momentarily inaccessible targets influence FOK judgment magnitude. Those features must be inherently meaningful to the learner. Perceptual elements that are arguably inherent to ambiguous pictorial stimuli seemingly do not influence FOK judgment magnitude.

Acknowledgments

Appendix A

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al (2007). The english lexicon project. *Behavior Research Methods, 39*, 445–459. <http://elexicon.wustl.edu/default.asp> Retrieved.

Begg, I., Duft, S., Lalonde, P., & Melnick, R. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28*(5), 610–632.

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.

Booth, A. E., & Waxman, S. R. (2002). Word learning is 'smart': Evidence that conceptual information effects preschoolers' extension of novel words. *Cognition, 84*, B11–B22.A.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109*, 204–223.

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

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–428.

Connor, L. T., Balota, D. A., & Neely, J. H. (1992). On the relation between feeling of knowing and lexical decision: Persistent subthreshold activation or topic familiarity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 544–554.

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*, 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*, 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*, 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.

Durso, F. T., & Shore, W. J. (1991). Partial knowledge of word meanings. *Journal of Experimental Psychology: General, 120*(2), 190–202.

Eysenck, M. W. (1979). The feeling of knowing a word's meaning. *British Journal of Psychology, 70*(2), 243–251.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183–209.

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

Groninger, L. D. (1979). Predicting recall: The "feeling-that-I-know" phenomenon. *The American Journal of Psychology, 92*, 45–58.

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

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

Hertzog, C., Dunlosky, J., & Sinclair, S. M. (2010). Episodic feeling-of-knowing resolution derives from the quality of original encoding. *Memory and Cognition, 38*, 771–784.

Holmes, J. B., Waters, H. S., & Rajaram, S. (1998). The phenomenology of false memory: Episodic content and confidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1026–1040.

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

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28.

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

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

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 187–194.

Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory and Cognition, 34*, 959–972.

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*, 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*, 1095–1105.

Lovelace, E. (1987). Attributes that come to mind in the TOT state. *Bulletin of the Psychonomic Society, 25*, 370–372.

Lupker, S. J., Harbluk, J. L., & Patrick, A. S. (1991). Memory for things forgotten. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 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*, 116–126.

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

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory and Cognition, 25*, 826–837.

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

Melcher, J. M., & Schooler, J. W. (1996). The misremembrance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory & Language, 35*, 231–245.

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*, 851–861.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16*, 519–533.

Moulin, C. J. A., Perfect, T. J., & Jones, R. W. (2000). The effects of repetition on allocation of study time and judgements of learning in alzheimer's disease. *Neuropsychologia, 38*, 748–756.

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

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

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.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 104–114.

Read, J. D., & Bruce, D. (1982). Longitudinal tracking of difficult memory retrievals. *Cognitive Psychology, 14*, 280–300.

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*, 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.

Sacher, M., Taconnat, L., Souchay, C., & Isingrini, M. (2009). Divided attention at encoding: Effect on feeling-of-knowing. *Consciousness and Cognition: An International Journal, 18*, 754–761.

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

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

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology, 22*, 36–71.

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*, 1074–1083.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A.

Thomas, A. K., Bulevich, J. B., & Dubois, S. J. (2011). Context affects feeling of knowing accuracy in younger and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 96–108.

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology, 21*, 437–484.

Underwood, B. J. (1969). Attributes of memory. *Psychological Review, 76*, 559–573.

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*, 289–335.