

Recognition during recall failure: Semantic feature matching as a mechanism for recognition of semantic cues when recall fails

Anne M. Cleary¹ · Anthony J. Ryals² · Samantha R. Wagner³

Published online: 18 August 2015
© Psychonomic Society, Inc. 2015

Abstract Research suggests that a feature-matching process underlies cue familiarity-detection when cued recall with graphemic cues fails. When a test cue (e.g., *potchbork*) overlaps in graphemic features with multiple unrecalled studied items (e.g., *patchwork*, *pitchfork*, *pocketbook*, *pullcork*), higher cue familiarity ratings are given during recall failure of all of the targets than when the cue overlaps in graphemic features with only one studied target and that target fails to be recalled (e.g., *patchwork*). The present study used semantic feature production norms (McRae et al., Behavior Research Methods, Instruments, & Computers, 37, 547–559, 2005) to examine whether the same holds true when the cues are semantic in nature (e.g., *jaguar* is used to cue *cheetah*). Indeed, test cues (e.g., *cedar*) that overlapped in semantic features (e.g., *a_tree*, *has_bark*, etc.) with four unretrieved studied items (e.g., *birch*, *oak*, *pine*, *willow*) received higher cue familiarity ratings during recall failure than test cues that overlapped in semantic features with only two (also unretrieved) studied items (e.g., *birch*, *oak*), which in turn received higher familiarity ratings during recall failure than cues that did not overlap in semantic features with any studied items. These findings suggest that the feature-matching theory of recognition during recall failure can accommodate recognition of semantic cues during recall failure, providing a potential mechanism for conceptually-based

forms of cue recognition during target retrieval failure. They also provide converging evidence for the existence of the semantic features envisaged in feature-based models of semantic knowledge representation and for those more concretely specified by the production norms of McRae et al. (Behavior Research Methods, Instruments, & Computers, 37, 547–559, 2005).

Keywords Familiarity in recognition memory · Semantic memory · Recognition

What leads something to seem familiar despite a failure to recall any specific prior experience with it? Ryals and Cleary (2012) suggest that at least one basis for this experience may be a feature-matching process, whereby features (i.e., attributes) of the current situation are compared with those of representations stored in memory to produce a sense of familiarity that varies according to the degree of match. Ryals and Cleary (2012) used the example of recognizing the street sign for “Marston” as familiar because of its high degree of feature-match to a recently-seen street sign, “Morton.” Despite failing to trigger recall of having recently passed a street sign for “Morton,” the new sign “Marston” may still seem familiar because of its high degree of feature-overlap with the recently-encountered sign “Morton.” The letter and sound features shared between the cue “Marston” and the memory for “Morton” may participate in the aforementioned feature-matching process such that a detectable familiarity signal emerges upon encountering Marston – a familiarity signal that exceeds some internal criterion for discriminating signals from noise (as in classic signal detection theory) – despite recall failure of the prior experience with the street name “Morton.” Though not specifically concerned with instances of retrieval failure, there are many recognition memory

✉ Anne M. Cleary
Anne.Cleary@colostate.edu

¹ Department of Psychology, Colorado State University, 1876 Campus Delivery, Fort Collins, CO, USA

² Department of Psychology, Northwestern University, Chicago, IL, USA

³ Department of Psychology, Northern Michigan University, Marquette, MI, USA

models that incorporate this feature-matching idea into their mechanisms for test item familiarity-detection (e.g., see Clark & Gronlund, 1996, for a review of global matching models, or Hintzman, 1988, for a specific case).

Ryals and Cleary (2012) found support for feature-matching as a basis for familiarity-detection that occurs during retrieval failure. As a laboratory analog to the “Marston/Morton” example above, Ryals and Cleary (2012) used a variant of the recognition without cued recall (RWCR) paradigm developed by Cleary (2004). Participants viewed words like *distraction* and *forehead* at study and were later presented with non-word test cues that potentially resembled studied words graphemically (e.g., *disfraption* as a cue for the word *distraction*, or *foneheed* for the word *forehead*). Half of the test cues resembled studied words and half did not. Participants were asked to try to use each cue to recall a studied word that resembled it and also to rate the familiarity of the cue itself in relation to the likelihood that a graphemically similar word had appeared at study, regardless of whether or not that target word could be recalled. As had been shown by Cleary (2004), a RWCR effect was found: Among cues for which cued recall did not succeed, familiarity ratings were significantly higher when the cue resembled a studied word than when it did not.

Ryals and Cleary (2012) reasoned that if global feature-matching (e.g., Hintzman, 1988) is the process responsible for RWCR, then the degree of overlap between the features in the cue and those in memory should matter to the effect. They examined this in two ways. First, they reasoned that unless features from study are actually reinstated in the test cue itself, those features should not affect the magnitude of the RWCR effect, even if they are features known to benefit recall itself. This is because in order to increase the familiarity signal that emerges from the feature-matching process, the features must be present in both the cue and in memory. Ryals and Cleary found support for this by examining study word concreteness in one experiment, and study word emotionality in another. Because these features of concreteness and emotionality were not carried within the non-word test cues themselves (e.g., the fact that *forehead* is a concrete word is not itself carried in the non-word cue *foneheed*), these features should not impact RWCR, even though they impact recall itself. Indeed, although concrete words led to a greater likelihood of cued recall than abstract words, when recall failed, concrete words did not lead to a greater RWCR effect. A similar finding was obtained with study word emotionality: Though emotional words were better later recalled in response to the non-word cues than neutral words, the familiarity ratings to the non-word cues that graphemically resembled unrecalled studied words were unaffected by study word emotionality. In short, as is consistent with the feature-matching hypothesis, the magnitude of the RWCR effect for graphemic cues was unaffected by study word concreteness or emotionality.

Second, Ryals and Cleary (2012) reasoned that increasing the overlap between studied features and features in the test cue itself should increase the RWCR effect, since the degree of feature-match between the cue and the memory representations should matter to any global feature-matching process. To examine this, they carried out an experiment in which the non-word test cues (e.g., *potchbork*) graphemically resembled either four studied words (e.g., *pitchfork*, *patchwork*, *pocketbook*, *pullcork*), one studied word (e.g., *pitchfork*) or zero studied words. In support of the feature-matching hypothesis, graphemic cues that resembled four studied words (but for which none of those studied words was recalled in response to the cue) received higher familiarity ratings than graphemic cues for which recall failed but that resembled only one studied word (which in turn received higher ratings than graphemic cues that did not resemble any studied words). In short, Ryals and Cleary (2012) found support for a feature-matching account of RWCR when the cue resemblance to studied items is graphemic in nature.

The present study is concerned with whether the feature-matching approach can plausibly account for RWCR that is found when cue resemblance to studied items is semantic in nature. Cleary (2004) demonstrated that when participants studied words (e.g., *cheetah*) and were tested with semantically related cues (e.g., *jaguar*), participants gave higher ratings to those cues that semantically resembled studied words than to those that did not. If this type of semantic RWCR effect is also based on the type of feature-matching process described by Ryals and Cleary (2012), what exactly are the semantic features that would drive this effect? Before addressing this question, it is important to first address why searching for evidence of semantic feature matching is a theoretically important endeavor.

There were two primary theoretical motivations for the present study. The first was that obtaining empirical support for semantic feature matching would help to provide a theoretical mechanism for a type of empirical phenomenon that currently lacks one – specifically, conceptual forms of recognition without identification. One conceptual form of recognition without identification is the ability to detect that the answer to a question was presented on a recent study list despite being unable to retrieve that answer (Cleary, 2006; Cleary, Staley, & Klein, 2014). When presented with general knowledge questions for which the answers fail to come to mind, participants are able to discriminate questions whose answers were studied from questions whose answers were not, even though they cannot think of the answers themselves. Cleary and colleagues have speculated that the presentation of a study word (e.g., “insomnia”) might create enhanced familiarity with the later test question itself (e.g., “What is the name of the inability to sleep?”) even when the answer cannot be retrieved. However, it is unclear what specific mechanism could enable such increased question familiarity during the answer’s

retrieval failure. Does the answer to the question link with the question itself in the knowledge-base in a way that would allow earlier presentation of the answer to increase the familiarity of the question later on despite a failure to retrieve the very answer that caused the question's familiarization in the first place? Semantic feature matching provides a plausible potential mechanism for this, and the RWCR paradigm used by Ryals and Cleary (2012) provides a useful methodological approach for examining the semantic featuring matching hypothesis.

We hypothesize that semantic features are a type of feature that can participate in a global feature-matching process (such as that specified by global matching models) to produce a familiarity signal whose intensity varies according to the degree of feature match. Support for this semantic feature-matching hypothesis would suggest a potential mechanism for the types of recognition during retrieval failure that are semantic or conceptual in nature, such as the ability to detect that a question's answer was studied despite failing to retrieve that answer (e.g., Cleary, 2006; Cleary et al., 2014) or that a famous scene's name was studied despite failing to retrieve it when viewing the scene (Cleary & Reyes, 2009). By enabling a specific theoretical mechanism to be put forward for such conceptual forms of recognition during retrieval failure, empirical support for the semantic feature-matching hypothesis would represent a significant theoretical advance in understanding how memory takes place during retrieval failure.

The second theoretical motivation was to provide converging evidence on the nature of semantic features themselves. According to the logic of converging operations (e.g., Campbell, 1988), if a given theoretical construct exists (in this case, semantic features), it should be possible to find evidence for it using multiple methodologies. The evidence should not be tied to one specific paradigm or methodological approach. From this perspective, the paradigm used by Ryals and Cleary (2012), when applied to semantic instead of graphemic features, presents a novel methodology from which to seek converging evidence for the existence of a specific type of semantic features.

To seek converging evidence, we turned to the literature that has suggested the plausibility of feature-based semantic knowledge representations (e.g., Plaut, 1995; Griffiths, Steyvers, & Tenenbaum, 2007; Landauer & Dumais, 1997; Seidenberg, 2007; Smith, Shoben, & Rips, 1974; Yee, Overton, & Thompson-Schill, 2009). Although concretely identifying semantic features remains a challenge, McRae and colleagues have made great strides in doing so by creating semantic feature-production norms based on what participants have consistently listed as the features that come to mind when presented with particular words (McRae, Cree, Seidenberg, & McNorgan, 2005). These norms were able to later explain much of the variance in functional magnetic resonance imaging (fMRI) patterns of activation during participant viewing of

objects while thinking about the object's properties in an independent study (Chang, Mitchell, & Just, 2011).

In the present study, we used the semantic feature-production norms of McRae et al. (2005) to examine the plausibility of a feature-matching explanation for RWCR that occurs with semantic cues. The norms published by McRae et al. provide a straightforward means of manipulating the degree of semantic feature overlap between a test cue and the study items in the RWCR paradigm analogously to that in Ryals and Cleary's (2012) Experiment 3. For example, the word "cedar" has high semantic feature overlap with the word "birch," as both words contain the participant-produced features "a_tree", "grows_in_forests", "has_bark", "has_branches", "has_leaves," and "is_tall." We used this database to create targets and cues that overlapped in semantic features.

In order to create varying degrees of feature-overlap between the test cue and the studied items, we created a semantic variant of the method used by Ryals and Cleary (2012). Recall that in their study, the graphemic cue (e.g., *potchbork*) shared graphemic features with four studied words (e.g., *pitchfork*, *patchwork*, *pocketbook*, *pullcork*) only one studied word (e.g., *pitchfork*) or no studied words. In our semantic variant of this method, we analogously varied the number of studied items sharing semantic features (e.g., "a_tree," "has_bark") with the test cue. In the high feature-overlap condition, the test cue "cedar" would have high semantic feature overlap with four studied words (*birch*, *oak*, *pine*, *willow*), whereas in the medium feature-overlap condition, it would have high semantic feature overlap with only two studied words (*birch*, *oak*). As in all studies of RWCR, our primary interest was in cue familiarity ratings given to cues for which none of the corresponding target words were identified.

Method

Participants

Ninety-seven students recruited from the Colorado State University campus participated in exchange for either course credit or payment.

Materials

Ninety-six sets of five related words were chosen for their sharing of semantic features. In each of these sets of five, one word served as the cue and the other four served as the potential study items. Eighty-seven of these sets of five were taken from McRae et al.'s (2005) semantic feature-production norms. On average, each cue shared approximately five semantic features with each of the four potential study items ($M = 4.99$, $SD = 1.95$). For example, the test cue "cedar" shares five features with the target word "oak" ("a_tree," "grows_in_forests," "has_leaves," "is_tall," and

“used_for_making_furniture”). In order to increase the total number of stimulus sets while minimizing inter-set semantic feature overlap, we took an additional nine sets of five from Van Overschelde, Rawson, and Dunlosky (2004) that were chosen to be comparable upon visual inspection in semantic feature overlap to the stimuli from McRae et al. For example, the cue word “rose” from this collection corresponded to the target words “daisy,” “tulip,” “lily,” and “carnation” and this set did not overlap with the sets taken from McRae et al. because there were no flowers from that set.¹

For each participant, stimulus sets were segmented into three conditions: High feature overlap, medium-feature overlap, and no-feature overlap. In the high-feature overlap condition, all four target words for the cue appeared at study. Thus, if the cue was “cedar,” all four target words “birch,” “pine,” “oak,” and “willow” would appear at study. In the medium-feature overlap condition, only two of the target words (e.g., “birch” and “oak”) appeared at study. In the no-feature overlap condition, none of the targets for that cue appeared at study. Counterbalancing was performed such that each cue appeared equally often in each condition (high-feature overlap, medium-feature overlap, no-feature overlap) across participants through random assignment to three different versions of the experiment (one more participant was run than anticipated; we still included this participant for a total of 97 participants).

The stimulus sets were segmented into two separate study-test blocks. Each study list contained 96 study words altogether (16 sets of four feature-sharing words, and 16 sets of two feature-sharing words, all randomly dispersed throughout the study list). Each test list contained 48 cues, 16 of which corresponded to four studied targets, 16 of which corresponded to two studied targets, and 16 of which corresponded to zero studied targets. To further try to minimize unintended overlap across sets, sets that seemed potentially similar to one another were assigned to different blocks. In addition, within blocks, we tried to assign similar-seeming sets to different counterbalancing sets to distribute them across sets as much as possible.

Procedure

The computerized procedure was modeled after Ryals and Cleary (2012) and was carried out using E-prime software. After viewing the initial instructions and pressing “1” to begin, the study list words began appearing, one at a time in a random order, in the upper left-hand corner of the screen at

rate of 1 s per word with no interstimulus interval. After all 96 study words were presented in this manner, the following test instructions appeared on the screen:

“On the test, you will be presented with word cues. Sometimes, these word cues will resemble studied words in their meaning. For example, POPE resembles the word BISHOP. Sometimes, the cues will not resemble any studied words. Your task will be to first rate how familiar the word cue seems to you. Keep in mind that the more a cue resembles a studied word, the more likely it is to seem familiar. You'll be asked to rate the cue on a scale of 0 (extremely unfamiliar) to 10 (extremely familiar). After rating the familiarity of the cue, you'll be asked if you can recall a studied word that resembles the cue. So, if the cue was POPE and it made you recall the word BISHOP from the study list, you would type in BISHOP. If you cannot think of a studied word from the list that resembles the cue, simply press Enter (remember, some of the time the cue will NOT correspond to any studied word).”

Note that because our focus is on residual memory during retrieval failure (where participants presumably have little to rely on in giving a rating other than their sense about the test cue itself) we used resemblance to studied items and level of cue familiarity interchangeably, as has been done in prior studies (e.g., Ryals & Cleary, 2012). This was to clarify for participants the nature of the judgment, given that it might seem awkward to be making a judgment about items that cannot be retrieved from memory and that therefore may have been on the study list but forgotten or may not have been on the study list at all. For example, a participant might ask, “How can I make a judgment about something that I cannot recall from memory?” By mentioning both familiarity with resemblance, we hoped to clarify how such a judgment could still be provided during retrieval failure.

The test list of cues was randomly ordered. Each test cue appeared in the upper left-hand corner for 2 s before a dialog box appeared in the center of the screen prompting the participant for a familiarity rating (see Fig. 1). After typing a rating between 0 and 10, participants were prompted with the second dialog box prompting recall and could type a response before pressing Enter or simply press Enter. The test cue remained on the screen throughout the prompting and receipt of participants' responses and until the next test cue was presented, then the program cycled through the same procedure. After cycling through the procedure for all 48 test cues, participants were prompted to begin the second study list, and the procedure was the same for that study-test block as for the first.

Our primary interest was in the familiarity ratings given to cues for which recall of all of the targets failed. As in Ryals and Cleary (2012), participants only needed to identify one of the

¹ Though care was taken to minimize inter-set semantic feature overlap, it is impossible to eliminate all possible sources of this. While this might increase the baseline ratings that participants give, the ultimate effect that this would likely have on the RWCR effect, however, is diminishment via decreased old-new discriminability (e.g., Cleary, Morris & Langley, 2007; Nomi & Cleary, 2012), as studies have previously shown that increasing the similarity among items in the overall pool of stimuli decreases old-new discriminability and reduces the likelihood of finding an effect.

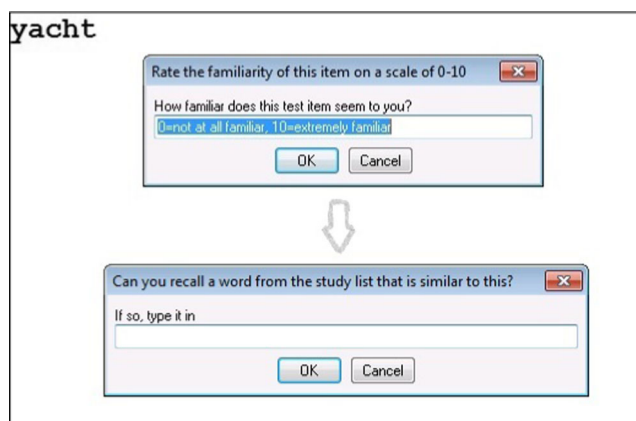


Fig. 1 An illustration of the test procedure. The primary interest was in familiarity ratings given to test cues for which recall failed (i.e., test cues for which no target words were recalled)

four possible target words in response to the cue in order for the trial to be classified as “target identified.” Identification of any of the four target words in any of the three conditions led to that trial being coded as an instance of correct target identification. Data were coded by hand by three different coders.

Results

Because our primary interest was in familiarity-detection that occurs during recall failure, we focus first on the familiarity ratings given to cues for which no targets were correctly identified. A repeated measures ANOVA performed on the familiarity ratings given to cues for which target retrieval failed revealed a significant effect of Feature Match Condition (high-feature overlap vs. medium-feature overlap vs. no-feature overlap), $F(2, 192) = 15.85$, $MSE = 7.73$, $p < .001$, $\eta^2 = .14$. As shown in Fig. 2, ratings were higher in the high-feature overlap condition than in the medium-feature overlap condition, $t(96) = 2.89$, $SE = 0.11$, $p < .01$, $d = .19$, and in turn were higher in the medium-feature overlap condition than in the no-feature overlap condition, $t(96) = 2.84$, $SE = 0.09$, $p < .01$, $d = .17$. Of course, ratings were also higher in the high-feature overlap condition than in the no-feature overlap condition, $t(96) = 5.45$, $SE = 0.10$, $p < .001$, $d = .35$.

We next examined ratings given when recall succeeded (i.e., when at least one target corresponding to the cue was identified). Because recall cannot succeed for cues in the no-feature overlap condition (as none of the corresponding targets would have been studied in this case), we compared ratings given when recall succeeded in the high-feature overlap condition ($M = 7.56$, $SD = 2.04$) with ratings given when recall succeeded in the medium-feature overlap condition ($M = 7.41$, $SD = 2.18$), and found that they did not differ significantly, $t(94) = 1.46$, $SE = .10$, $p = .15$. Though this is consistent with Ryals and Cleary’s (2012) finding that graphemic feature

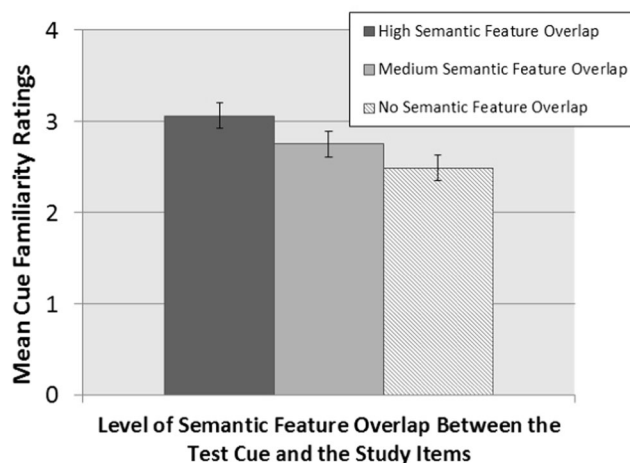


Fig. 2 Cue familiarity ratings given during cued recall failure as a function of the degree of semantic feature overlap between the test cue and the studied words in memory. Error bars represent confidence intervals computed for within-subjects designs (Loftus & Masson, 1994)

overlap affected RWCR to a greater extent than it affected recognition *with* cued recall, a 2×2 Feature Overlap Condition (high-feature overlap vs. medium-feature overlap) \times Recall Status (recall failed vs. recall succeeded) repeated measures ANOVA on ratings revealed no significant interaction, $F(1, 94) = 1.21$, $MSE = .50$, $p = .28$, despite the presence of an overall main effect of Feature Overlap Condition, $F(1, 94) = 8.74$, $MSE = .58$, $p = .004$, $\eta^2 = .085$.

The same 2×2 ANOVA also revealed a significant main effect of Recall Status, $F(1, 94) = 442.28$, $MSE = 4.58$, $p < .001$, $\eta^2 = .825$, such that ratings were higher when recall succeeded than when it failed. It is possible that during successful target retrieval, participants base their ratings, at least in part, on the fact that they can recall at least one target word from the study list and recollect its occurrence on the list. Some support for this idea can be found in the comparison between false recall attempts (when the retrieved word was not a target) and correct recall attempts. A 2×2 Feature Overlap Condition (high-feature overlap vs. medium-feature overlap) \times Correctness (successful target recall vs. false recall) repeated measures ANOVA on ratings revealed a main effect of Correctness, $F(1, 92) = 99.21$, $MSE = 1.80$, $p < .001$, $\eta^2 = .519$, such that ratings were higher overall when recall was correct (high-feature overlap condition $M = 7.51$, $SD = 2.06$; medium-feature overlap condition: $M = 7.34$, $SD = 2.19$) than when recall was false (high-feature overlap condition $M = 6.22$, $SD = 2.30$; medium-feature overlap condition: $M = 5.84$, $SD = 2.40$). (Note that some participants were lost in this analysis due to no responses in one of the false recall categories, which is why the means for successful recall are slightly different from above). This analysis also revealed a main effect of Feature Overlap Condition, $F(1, 92) = 7.41$, $MSE = .92$, $p = .008$, $\eta^2 = .075$. There was no significant interaction, $F(1, 92) = 1.00$, ns .

In terms of the likelihood of recalling at least one target corresponding to the cue, there was a greater likelihood of such recall in the high-feature overlap condition ($M = .42$, $SD = .15$) than in the medium-feature overlap condition ($M = .35$, $SD = .13$), $t(96) = 5.54$, $SE = .01$, $p < .001$, $d = .49$. Using the likelihood of correct target identification in the no-feature overlap condition as an estimate of the correct guessing rates ($M = .10$, $SD = .09$), the likelihood of recall was greater than that expected by guessing in both the high-feature overlap, $t(96) = 18.44$, $SE = .02$, $p < .001$, $d = 2.57$, and the medium-feature overlap, $t(96) = 18.31$, $SE = .01$, $p < .001$, $d = 2.21$, conditions.

Discussion

The present study examined what we refer to as the semantic feature-matching hypothesis for conceptually-based forms of cue recognition that occur during retrieval failure. We sought specifically to determine whether the feature-matching theory of recognition without cued recall (RWCR) can explain the effect that occurs with semantic cues. Prior studies with scenes (Cleary et al., 2012) and graphemic cues (Ryals & Cleary, 2012) have supported a feature-matching theory of RWCR, but these types of RWCR have involved perceptual types of features, which are easier to concretely specify than conceptual features. Given that Cleary (2004) showed that RWCR can also be shown with semantic resemblance, as when a studied word (e.g., *cheetah*) is cued with a semantically similar word (e.g., *jaguar*), the paradigm used by Ryals and Cleary (2012) when applied to these types of cues presents a methodological means of examining the semantic feature-matching hypothesis for conceptually-based forms of recognition during retrieval failure.

Examining the semantic feature-matching hypothesis is a theoretically important endeavor for two primary reasons. First, empirical support for this hypothesis would enable a well-specified theoretical mechanism to be put forward in the literature for how recognition during retrieval failure takes place when the relation between the cues and the targets is conceptual. Examples include participants' ability to detect that the unretrieved answer to a general knowledge question was presented on a recent study list, despite the answer's retrieval failure (Cleary, 2006; Cleary et al., 2014), and participants' ability to detect that the unretrieved name of a famous landmark was presented on a recent study list, despite the name's retrieval failure in response to the picture (Cleary & Reyes, 2009). Currently, there is no specific theoretical mechanism for how such conceptually-based forms of recognition during retrieval failure take place.

Second, support for the semantic feature-matching hypothesis would provide converging evidence, from a converging operations perspective (i.e., the idea that support for the existence of a construct, if the construct exists, should be obtainable

through multiple methods and approaches, also referred to previously as convergent operationalism and methodological triangulation, e.g., Campbell, 1988, p. 28), for the existence of specific kinds of semantic features in the human knowledge-base. What are the features in the cue *jaguar* that can potentially overlap with the studied word *cheetah*? Our paradigm presents a novel methodological means of seeking support for the existence of semantic features of the type investigated in the very different literature on conceptual knowledge representations (e.g., McRae et al., 2005). For this reason, we chose semantic features that were identified in prior semantic feature-production norm research (McRae et al., 2005) and for which support was independently found in fMRI research (Chang et al., 2011). In our converging operations approach, we use these semantic features in a completely novel way by applying them to the RWCR paradigm used by Ryals and Cleary (2012) to examine whether these features can plausibly participate in the type of feature-matching process specified by global matching models (e.g., Clark & Gronlund, 1996) to allow for cue recognition during retrieval failure.

Using the semantic features identified by McRae et al. (2005), we created a semantic feature overlap analog to the graphemic feature overlap experiment reported by Ryals and Cleary (2012). In our study, each test cue (e.g., *cedar*) potentially overlapped in semantic features (e.g., *a_tree*, *has_bark*) with four studied words (e.g. *birch*, *oak*, *pine*, *willow*), two studied words (e.g., *birch*, *oak*), or no studied words. As is consistent with the feature-matching explanation of RWCR, when participants failed to identify any of the potential targets for the semantic test cue, they gave higher cue familiarity ratings when the cue overlapped in semantic features with four studied words than when it overlapped with only two studied words, and also when it overlapped with two studied words than when it did not overlap with any studied words.

This pattern of results has several levels of theoretical importance. At a broad level, the results provide further support for the feature-matching theory of RWCR, with a completely different type of feature than previously investigated, and with a feature-type that is more abstract and difficult to identify than in previous research. This suggests that the feature-matching theory of RWCR is not limited to features that are perceptual in nature, such as with graphemic cues for words (e.g., Ryals & Cleary, 2012), or with spatial layout cues for scenes (e.g., Cleary et al., 2012). The feature-matching theory appears to apply even in cases where the overlapping information from study to test is more abstract or conceptual in nature.

At a more specific theoretical level, this is important because it provides a theoretical mechanism (where none had previously been put forward) for how recognition during retrieval failure can take place when the relationship between the studied item failing to be retrieved and its unstudied but related cue is conceptual. Though lacking in a well-specified mechanism, there have been many manifestations of conceptually-based cue

recognition during target retrieval failure reported in the literature. For example, when one studies the word “insomnia” and later fails to retrieve that studied word in response to the question, “What is the name of the inability to sleep?” the relationship between the unstudied test cue and the unretrieved studied target is conceptual in nature, and participants can detect that the unretrieved answer was studied despite failing to retrieve it from the cue (Cleary, 2006). The present support for the semantic feature-matching hypothesis provides a theoretical mechanism for what might be occurring to enable the detection that the unretrieved word was studied using only the cue itself (the cue in this case being the test question). Specifically, the test question may be used in a global feature-matching process by which the conceptual features in the test question are matched with the conceptual features of encoded items in memory to produce a familiarity signal. Because of the semantic feature overlap between the question and its unretrieved studied answer, that familiarity signal will be greater when the answer to the question was studied than when it was not. This could explain why people can discriminate between questions whose answers were studied and questions whose answers were not studied when the answers fail to be retrieved. The same mechanism could also explain other instances of conceptually-based recognition during retrieval failure, such as the ability to discriminate famous scenes (e.g., Taj Mahal) whose names were studied from famous scenes whose names were not studied despite failing to retrieve those names from the scenes (Cleary & Reyes, 2009).

At a less specific level of theoretical importance, the present findings are important for increasing understanding of the general basis of the RWCR phenomenon. The fact that the feature-matching idea applies even in cases where the mapping between the test cue and the unretrieved target in memory is strictly conceptual in nature suggests that a general theory of the RWCR effect – one that can account for all of its various forms, both perceptual and abstract – may suffice. RWCR may result from cue familiarity-detection brought on by a global feature-matching process like that described by global matching models (e.g., Clark & Gronlund, 1996; Hintzman, 1988).

This is not to say that all feature-types are necessarily weighted equally. Although global matching models assume that all feature-types are weighted equally in the computation of the familiarity signal (in fact, global matching models neither specify the identity of the features nor distinguish between classes of them; see Clark & Gronlund, 1996, for a review), it is possible that in the operation of human cognitive processes, different types of features are weighted differently in the computation of familiarity. In fact, a cross-experiment comparison between the present experiment with semantic features and the analogous experiment reported by Ryals and Cleary (2012) with graphemic features suggests a difference in the level of familiarity elicited by these different feature types (semantic vs. graphemic).

Following from the meta-analytic approach taken by Rhodes and Anastasi (2012), we computed Q_B in order to assess whether the effect sizes differed significantly between the two experiments.² We found that semantic feature overlap led to a smaller RWCR effect than graphemic feature overlap. This was shown by comparing the difference between the high-feature overlap and lower feature overlap condition from the present semantic features experiment (Cohen’s $d = .19$) with that from the graphemic features Experiment 3 of Ryals and Cleary (Cohen’s $d = .80$), $Q_B = 19.13$, $p < .001$. It was also shown by comparing the difference between the lower feature overlap condition and the no-feature overlap condition from the present semantic features experiment (Cohen’s $d = .17$) with that from the graphemic features Experiment 3 from Ryals and Cleary (Cohen’s $d = .46$), $Q_B = 9.39$, $p = .002$.

If different feature types are weighted differently in the computation of the familiarity signal, then models of the process would need to accommodate this. One possibility is to modify existing recognition memory models to accommodate different feature-type weightings (although recognition memory models do not focus specifically on instances of retrieval failure, the basic familiarity-detection processes themselves might be good descriptions of the process that enables cue recognition during target retrieval failure). Another possibility is to perhaps modify existing distributed models of the type used to describe semantic feature overlap in priming situations (e.g., Plaut, 1995) to compute a familiarity signal based on the degree of feature overlap between a cue and the memory representations. This type of model should be able to accommodate different feature types carrying different weightings in the computation of feature-overlap-based familiarity. Other future work might incorporate the notion of semantic richness (e.g., Pexman, Siakaluk, & Yap, 2013; Yap, Pexman, Wellsby, Hargreaves, & Huff, 2012), whereby words that are more semantically rich (such as by having more semantic features as generated in feature lists of the type used by McRae et al., 2005) are more easily recognized and processed. These may all be possible directions for future theory development regarding how different features, and the extent to which they are present in a stimulus, might contribute to the overall computation of the familiarity signal.

At another level of importance to theory development is the following. Because we have made the case that one of the primary theoretical contributions of the present study is in putting forward a theoretical mechanism for a phenomenon currently lacking one (i.e., semantic feature-matching as a basis for cue familiarity in cases of conceptually-based forms of cue recognition during target retrieval failure), alternative

² Ryals and Cleary (2012) had computed Cohen’s d from the t value (Rosenthal, 1991). However, in light of recent work suggesting that this method of computing d is positively biased (e.g., Fritz, Morris & Richler, 2012; Lakens, 2013), we applied Hedge’s correction to our computation of Cohen’s d , both for the present data and for the data of Ryals and Cleary in making our comparison.

theories do not yet exist to be ruled out at this stage. Still, some might wonder if any existing theory of other phenomena would predict that cue familiarity ratings should *not* correspondingly increase with increases in the number of unretrieved studied items sharing semantic features with the cue for which retrieval of those targets failed. Indeed, there is a class of theories – context noise theories – that cannot easily accommodate the pattern reported here.

An example of one such theory is BCDMEM (Bind, Cue, Decide Model of Episodic Memory; Dennis & Humphreys, 2001). In the BCDMEM model, the test cue is not decomposed into features that are then later matched with the features of the memory representations. Instead, rather than being compared to the study list items, the test item is compared to the previous contexts in which that *test item itself* was previously encountered. This theory works well for standard recognition tests (in which the test cue is usually a repeat of an earlier-studied item); however, it has difficulty accommodating the corresponding increase in cue familiarity ratings with increases in number of unretrieved words semantically sharing features with that cue. One would have to assume that the participant generates the particular later test cue upon presentation of every semantically related word at study so that it was as if the test cue had been presented every time, leading to higher ratings when four related words were studied than when only two were studied, and higher ratings when two were studied than when none were studied.

While this idea might be applicable to false memory paradigms, where the critical unstudied item (e.g., *sleep*) deliberately has a strong forward association from each study word (e.g., *bed, rest, awake, dream*, etc.), the idea seems less plausible in the present paradigm. Here, the idea would be that participants would self-generate the test cue (e.g., *cedar*) when presented with the first study word to which it overlaps in semantic features (e.g., *birch*), then again when presented at a later time with another (e.g., *oak*), again at a later time when presented with another (e.g., *pine*), and again at a later time when presented with another (e.g., *willow*). Then, when later presented with the unstudied word *cedar* as a test cue, those four randomly dispersed contexts in which it was earlier self-generated are what lead to higher ratings than when only two related contexts were presented for self-generation (despite failing to recall those contexts). This explanation for our data pattern seems unlikely given the differences between the present paradigm and false memory paradigms, namely: (1) the set of two or four study words are not clustered together in the present study but were instead randomly dispersed throughout the study list, and (2) the test cue was not chosen to critically relate to each study word in strong forward associative fashion but rather was chosen to overlap in semantic features approximately equally with all of them such that any of the set of five feature overlapping words could have served as the cue. That said, future research could be conducted to more definitively rule out context-noise

models as viable explanations for this type of phenomenon, such as by using think aloud/verbal protocol paradigms (e.g., Ericsson & Simon, 1980) in which participants must state aloud what they are thinking of at the time of study presentation.

Along similar lines, some might wonder if the study words each prompt generation of the category to which they belong, and the test cue then does the same so that when participants are presented with the test cue (e.g., *cedar*), they consciously generate the common category name (e.g., *tree*), and then recall having generated that category name in the study list without recalling any of the words that prompted its generation in the list. This is unlikely for several reasons. First, while many of the sets belonged to an obvious common category (e.g., *birch, oak, pine, willow, cedar*), many did not (e.g., *banner, brick, certificate, tray, ruler*). Second, in cases where there was a categorical relationship, the category was typically one of the features shared among the items (e.g., “a tree”) along with others (“e.g., “has leaves”). Third, as mentioned, the words in a set were randomly dispersed throughout the list (not clustered together in a way that would draw attention to their potential common category membership); this would have made it difficult to consistently consciously generate the same word in response to each study word from a set in order to later generate it again in response to the test cue without being able to retrieve any of the words that prompted that word’s generation. Finally, if category generation at study did underlie the enhanced RWCR effect with increasing global feature overlap between the study list and the test cue, we would expect this enhancement effect to be larger in the present study of semantic feature overlap than in the study of graphemic feature overlap by Ryals and Cleary (2012). The reason for this is because it is more obvious what the possible category might be in the case of semantic feature overlap (e.g., *birch, oak, pine, willow*, and *cedar* are all trees) than in the case of graphemic overlap (e.g., it is less obvious what the category name would be for *patchwork, pitchfork, pocketbook, pullcork*, and the cue *potchbork*). As discussed above, the feature overlap effect in the present semantic feature overlap situation is instead significantly *smaller* than the feature overlap effect in the graphemic feature overlap situation used by Ryals and Cleary (2012), making it seem unlikely that category generation underlies these effects.

Some might also wonder if cue familiarity detection during retrieval failure reflects an implicit type of memory. It is an interesting possibility. On the one hand, some evidence points toward it being an explicit, as opposed to implicit, form of memory. Ryals and Cleary (2012) report an experiment (in their Footnote 3) in which they used a remember-know-guess paradigm to examine whether the RWCR effect with graphemic cues would emerge primarily in “know” judgments (which should reflect explicit familiarity) or in “guess” judgments (which might indicate implicit memory, of which participants are unaware). They found that the RWCR effect emerged in “know” judgments and not “guess” judgments, supporting the idea that it

reflects explicit familiarity detection rather than implicit memory. That said, however, Ryals, Yadon, Nomi and Cleary (2011) found evidence that a very similar phenomenon, recognition without identification, may reflect an unconscious form of memory. They found that the recognition without identification effect (whereby unidentified fragments of studied words received higher recognition ratings than unidentified fragments of unstudied words) was related to an ERP signature thought to be reflective of unconscious recognition processes (e.g., Voss & Paller, 2009). This leaves open the possibility that some forms of recognition during retrieval failure reflect unconscious processes while others reflect a more conscious type of familiarity-detection. Future research should further investigate this issue.

If the semantic feature overlap effect shown in the present study does reflect an explicit feeling or sense about the test cue, then another interesting question is how this might relate to the tip-of-the-tongue (TOT) experience, whereby a person senses that a word is in the knowledge-base despite failing to retrieve it. Specifically, might TOT experiences result from semantic feature overlap between a current situation and one or more in memory? Though this idea may be intriguing, some evidence against it comes from the fact that TOT experiences appear to be dissociable from the type of old-new discrimination that characterizes recognition during retrieval failure. As Cleary, Staley and Kein (2014) review, old-new discrimination during retrieval failure and reports of being in a TOT state appear to be driven by different mechanisms. Still, future research should aim to examine the extent to which a semantic feature-matching account can explain TOT experiences, even if to more clearly rule out this potential mechanism as a basis for TOT experiences. For example, do situations of high semantic feature overlap between a cue and memory representations lead to a greater probability of a TOT report during retrieval failure than situations of low semantic feature overlap? Or, does semantic feature overlap have no impact at all on the likelihood of reporting a TOT?

Finally, at a broad level of theoretical importance, from a converging operations perspective, the present results provide independent converging support for feature-overlap approaches to semantic knowledge representation. According to the logic of converging operations, if a theoretical construct exists, it should be possible to arrive at the conclusion of its existence through multiple different methods and approaches; in short, evidence for its existence should not be tied to a particular methodological approach (e.g., Campbell, 1988, p. 28). In the semantic priming literature, some theoretical approaches have emphasized feature overlap (e.g., Plaut, 1995; Seidenberg, 2007; Smith et al., 1974). Our methodological approach points toward the same conclusions, but from a completely different methodological approach than that taken in the priming literature. By suggesting a role of semantic feature-matching in RWCR of semantic cues, our results provide converging support for feature-overlap approaches. In demonstrating that, in cases where cued recall fails, the

semantic features identified by McRae et al. (2005) can participate in the type of global feature-matching process described by Ryals and Cleary (2012) in their study of RWCR with graphemic features, we have obtained independent support, from a completely different methodological approach, for the existence of these semantic features as subcomponents of word or concept representations.

In summary, the present study provides support for the idea that the feature-matching theory of RWCR put forth by Ryals and Cleary (2012) is a plausible explanation for RWCR that occurs with semantic cues. In this way, the present results have allowed us to put forward a potential theoretical mechanism by which conceptual forms of recognition during retrieval failure (e.g., Cleary, 2006; Cleary & Reyes, 2009; Cleary et al., 2014) might occur. At another theoretical level the present results have provided converging evidence, from a converging operations perspective, for the existence of semantic features in the memory base, and also for semantic features of the type specified by McRae et al. (2005). That said, future work should aim to further tease apart feature overlap and associative strength. Although many researchers have found compelling evidence for the greater importance of semantic feature overlap over mere associative strength between items (e.g., Yee et al., 2009), the distinction has presented a historic problem in research on semantic knowledge (e.g., McRae & Jones, 2013), whereby it has proven incredibly difficult for researchers to completely separate the two. The fact that the semantic feature overlap effects reported here are smaller than the graphemic feature overlap effects reported previously (Ryals & Cleary, 2012) points toward feature overlap over associative strength playing a role because associative strength should not exist between a novel non-word test cue (e.g., POTCHBORK) and the studied target words (e.g., PITCHFORK, PATCHWORK, POCKETBOOK, PULLCORK), yet the effects are larger in that situation rather than smaller. If associative strength were giving the present effects a boost over and above semantic feature overlap, we would expect the effects to perhaps be larger than in the graphemic overlap situation. Still, now that we have put forward semantic feature matching as a theoretical mechanism for conceptual forms of recognition during retrieval failure, future work should aim to further investigate the extent to which the strength of associations, in the absence of semantic feature overlap, might contribute.

Acknowledgments This project was supported by the National Science Foundation (NSF) Research Experiences for Undergraduates (REU) and Department of Defense Awards to Stimulate and Support Undergraduate Research Experiences (ASSURE), NSF Grant SES-0552876 to Edward L. DeLosh (Director of the Summer REU program on Mind and Brain at Colorado State University). This material is based upon some work that took place while Anne M. Cleary was serving at the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Appendix

BLOCK 1 Stimuli

Counterbalancing Set 1			Counterbalancing Set 2			Counterbalancing Set 3		
4-Item Study Sets	2-Item Study Sets	Test Cue	4-Item Study Sets	2-Item Study Sets	Test Cue	4-Item Study Sets	2-Item Study Sets	Test Cue
birch	oak	cedar	crossbow	catapult	slingshot	colander	colander	strainer
pine	birch		bow	crossbow		mixer	grater	
oak			catapult			blender		
willow			harpoon			grater		
buzzard	eagle	falcon	cabinet	shelves	bookcase	pipe	pipe	drain
eagle	hawk		closet	cabinet		faucet	faucet	
hawk			shelves			tap		
owl			cupboard			hose		
cabin	cabin	cottage	bag	jar	bottle	bluejay	bluejay	robin
hut	hut		jar	sack		parakeet	finch	
shack			sack			finch		
tent			bin			woodpecker		
bayonet	machete	sword	earmuffs	gloves	mittens	mink	rat	squirrel
knife	bayonet		gloves	earmuffs		rat	skunk	
dagger			scarf			porcupine		
machete			cap			skunk		
shirt	sweater	shawl	book	menu	magazine	rock	rock	stone
vest	shirt		card	book		pearl	emerald	
blouse			menu			emerald		
sweater			envelope			coin		
bison	bison	buffalo	otter	otter	beaver	banner	tray	ruler
cow	bull		walrus	walrus		brick	brick	
ox			seal			certificate		
bull			platypus			tray		
mackerel	catfish	trout	apple	peach	nectarine	hook	pin	tack
cod	cod		plum	plum		pin	hook	
catfish			peach			scissors		
tuna			pear			razor		
goose	turkey	chicken	caribou	fawn	deer	gown	leggings	nylons
partridge	goose		elk	elk		dress	leotard	
pheasant			moose			leggings		
turkey			fawn			leotard		
daisy	tulip	rose	beet	yam	carrot	radio	radio	stereo
tulip	camation		yam	turnip		clock	projector	
lily			cauliflower			projector		
camation			turnip			typewriter		
anchor	ship	yacht	bowl	pan	skillet	tango	tango	ballet
ship	sailboat		kettle	pot		salsa	waltz	
submarine			pan			hip hop		
sailboat			pot			waltz		
butterfly	hornet	wasp	hyena	fox	coyote	hurricane	hurricane	tornado
hornet	housefly		fox	hyena		rain	rain	
moth			wolf			snow		
housefly			jackal			hail		

(continued)

bagpipe	trumpet	trombone	doll	doll	kite	vodka	wine	beer
saxophone	saxophone		drum	rattle		wine	whiskey	
trumpet			sled			rum		
tuba			rattle			whiskey		
sandals	shoes	boots	diesel	diesel	gasoline	chemistry	chemistry	biology
shoes	sandals		oil	oil		physics	physics	
slippers			coal			psychology		
socks			solar			astronomy		
bus	bus	van	chisel	drill	screwdriver	cellar	shed	barn
subway	trolley		sandpaper	chisel		basement	garage	
trolley			pliers			shed		
train			drill			garage		
bike	bike	tricycle	alligator	alligator	crocodile	raisin	raisin	prune
scooter	unicycle		salamander	salamander		grape	grape	
motorcycle			tortoise			avocado		
unicycle			turtle			olive		
bomb	bomb	missile	celery	zucchini	cucumber	guitar	guitar	banjo
bullet	grenade		eggplant	pickle		harp	violin	
grenade			pickle			violin		
spear			zucchini			cello		
BLOCK 2 Stimuli			Counterbalancing			Counterbalancing		
Counterbalancing			Set 2			Set 3		
Set 1			Set 2			Set 3		
4-Item Study	2-Item Study	Test Cue	4-Item Study	2-Item Study	Test Cue	4-Item Study	2-Item Study	Test Cue
Sets	Sets		Sets	Sets		Sets	Sets	
duck	flamingo	pelican	sink	sink	bath tub	curtains	curtains	drapes
swan	swan		toilet	tank		mat	blinds	
seagull			tank			carpet		
flamingo			bucket			blinds		
cheetah	tiger	cougar	bra	bra	camisole	cantaloupe	cantaloupe	honeydew
leopard	cheetah		nightgown	nightgown		onion	pumpkin	
panther			pajamas			tomato		
tiger			robe			pumpkin		
building	chapel	church	oven	oven	stove	giraffe	gorilla	bear
pyramid	cathedral		microwave	microwave		gorilla	chimp	
chapel			toaster			chimp		
cathedral			fridge			elephant		
bazooka	revolver	gun	chandelier	lantern	lamp	ashtray	cigar	cigarette
rifle	rifle		candle	candle		urn	pipe	
revolver			lantern			cigar		
shotgun			light			pipe		
cape	jacket	coat	bracelet	bracelet	necklace	chickadee	chickadee	canary
cloak	cloak		chain	chain		cheese	dandelion	
jacket			ring			beehive		
parka			medal			dandelion		
jeans	trousers	pants	hatchet	hatchet	axe	wheelbarrow	wheelbarrow	crane
trousers	jeans		hammer	hammer		level	level	
belt			sledgehammer			wrench		
skirt			tomahawk			crowbar		
guppy	minnow	perch	paintbrush	pencil	pen	bed	dresser	bureau
minnow	guppy		brush	crayon		pillow	mirror	
sardine			pencil			dresser		
goldfish			crayon			mirror		
emu	emu	ostrich	strawberry	raspberry	blueberry	calf	lamb	sheep
peacock	peacock		cherry	strawberry		pig	goat	
penguin			cranberry			lamb		
rooster			raspberry			goat		
iron	iron	steel	pony	pony	horse	fireman	nurse	doctor
silver	copper		zebra	donkey		teacher	teacher	
copper			camel			lawyer		
gold			donkey			nurse		
raft	raft	canoe	mug	mug	cup	cake	cake	pie
pier	boat		saucer	saucer		biscuit	biscuit	

(continued)

boat			plate			bread		
surfboard			dish			muffin		
accordion	keyboard	piano	parsley	cabbage	lettuce	water	coke	soda
keyboard	accordion		rhubarb	spinach		coke	juice	
clarinet			cabbage			milk		
harpichord			spinach			juice		
lemon	orange	tangerine	football	toy	ball	fork	ladle	spoon
orange	lemon		marble	skateboard		ladle	spatula	
lime			skateboard			spatula		
mandarin			toy			tongs		
chipmunk	mouse	mole	broom	shovel	hoe	bolts	bolts	screws
gopher	chipmunk		rake	spade		clamp	clamp	
mouse			shovel			nails		
groundhog			spade			buckle		
airplane	airplane	jet	ant	cockroach	beetle	clam	crab	lobster
rocket	rocket		grasshopper	grasshopper		crab	squid	
helicopter			cockroach			squid		
balloon			flea			octopus		
taxi	truck	car	chair	couch	sofa	seconds	minutes	hours
limo	limo		couch	bench		minutes	seconds	
ambulance			rocker			years		
truck			bench			days		
bat	stick	baton	apron	shield	thimble	buggy	cart	wagon
whip	wand		armour	armour		trailer	buggy	
stick			helmet			sleigh		
wand			shield			cart		

References

- Campbell, D. T. (1988). *Methodology and Epistemology for Social Science*. (E.S. Overman, Ed. Chicago: University of Chicago Press.
- Chang, K. K., Mitchell, T., & Just, M. A. (2011). Quantitative modeling of the neural representation of objects: How semantic feature norms can account for fMRI activation. *Neuroimage*, *56*, 716–727.
- Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*, 37–60.
- Cleary, A. M. (2004). Orthography, phonology, and meaning: Word features that give rise to feelings of familiarity. *Psychonomic Bulletin & Review*, *11*, 446–451.
- Cleary, A. M. (2006). Relating familiarity-based recognition and the tip-of-the-tongue phenomenon: Detecting a word's recency in the absence of access to the word. *Memory & Cognition*, *34*, 804–816.
- Cleary, A. M., Morris, A. L., & Langley, M. M. (2007). Recognition memory for novel stimuli: The structural regularity hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 379–393.
- Cleary, A. M., Brown, A. S., Sawyer, B. D., Nomi, J. S., Ajoku, A. C., & Ryals, A. J. (2012). Familiarity from the configuration of objects in 3-dimensional space and its relation to déjà vu: A virtual reality investigation. *Consciousness and Cognition*, *21*, 969–975.
- Cleary, A. M., & Reyes, N. L. (2009). Scene recognition without identification. *Acta Psychologica*, *131*, 53–62.
- Cleary, A.M., Staley, S.R., & Klein, K.R. (2014). The effect of tip-of-the-tongue states on other cognitive judgments. In B.L. Schwartz & A.S. Brown's *Tip-of-the-tongue States and Related Phenomena*. 264–280.
- Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452–477.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215–251.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*, 2–18.
- Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*, 211–244.
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528–551.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*(863), 1–12.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476–490.
- McRae, K., & Jones, M.N. (2013). Semantic Memory. In D. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology*. Oxford. 206–219.
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, & Computers*, *37*, 547–559.
- Nomi, J. S., & Cleary, A. M. (2012). Judgments for inaccessible targets: Comparing recognition without identification and the feeling of knowing. *Memory & Cognition*, *40*, 1178–1188.
- Pexman, P. M., Siakaluk, P. D., & Yap, M. J. (2013). Introduction to the research topic meaning in mind: Semantic richness effects in language processing. *Frontiers in Human Neuroscience*, *7*, 723.
- Plaut, D. C. (1995). *Semantic and associative priming in a distributed attractor network*. *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 37–42). Hillsdale: Lawrence Erlbaum Associates.

- Rhodes, M. R., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin*, *138*, 146–174.
- Rosenthal, R. (1991). *Meta-analytic Procedures for Social Research*. Newbury Park: SAGE Publications, Inc.
- Ryals, A. J., & Cleary, A. M. (2012). The Recognition without cued recall phenomenon: Support for a feature-matching theory over a partial recollection account. *Journal of Memory and Language*, *66*, 747–762.
- Ryals, A. J., Yadon, C. A., Nomi, J. S., & Cleary, A. M. (2011). When word identification fails: ERP correlates of recognition without identification and of word identification failure. *Neuropsychologia*, *12*, 3224–3237.
- Seidenberg, M. S. (2007). Connectionist models of reading. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 235–250). New York: Oxford University Press.
- Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A feature model for semantic decisions. *Psychological Review*, *81*, 214–241.
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*, 289–335.
- Voss, J. L., & Paller, K. A. (2009). An electrophysiological signature of unconscious recognition memory. *Nature Neuroscience*, *12*, 349–355.
- Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: Cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, *6*, 72.
- Yee, E., Overton, E., & Thompson-Schill, S. L. (2009). Looking for meaning: Eye movements are sensitive to overlapping semantic features, not association. *Psychonomic Bulletin & Review*, *16*, 869–874.