# An Insight into Language: Investigating Lexical and Morphological Effects in Compound Remote Associate Problem Solving

**Alexander H. Bower (ahbower@uci.edu)**
2275 Social & Behavioral Sciences Gateway Building
University of California
Irvine, CA 92697

**Andrew Burton (ajburton@uci.edu)**
402 Social Science Lab
University of California
Irvine, CA 92617

**Mark Steyvers (mark.steyvers@uci.edu)**
2201 Social & Behavioral Sciences Gateway Building
University of California
Irvine, CA 92697

**William Batchelder**
2129 Social Science Plaza A
University of California
Irvine, CA 92617

**Abstract**

Understanding the processes leading to insight has remained one of psychology's greatest challenges. In this study, we examined how different lexical properties affect cognitive processes involved in a popular class of insight problems: Compound Remote Associates (CRAs). These properties were familiarity, lexeme meaning dominance, and semantic transparency. We found that a higher proportion of problems were solved when they were presented beginning with the most familiar cues, but not when they began with right-headed dominant or the most semantically transparent cues. Further, we found that participants focused their efforts disproportionately on first and last cues, that subjective ratings of insight decreased as trial times elapsed, and that the magnitude of reported insight increased with the number of cues successfully solved. This suggests that participants can monitor their progress in such problems. These results contest longstanding assumptions of requisite periods of impasse and the absence of incremental progress in insightful problem solving.

**Keywords:** compound remote associates; insight; language and thought; problem solving

## Introduction

Insight has sparked some of history's greatest accomplishments – from Einstein's special theory of relativity to Newton's universal law of gravitation. These sudden "aha!" moments also permeate our everyday lives – from practical household problems to puzzles in video games. However, our understanding of the processes underlying insight have remained subject to empirical gaps and theoretical debate (Batchelder & Alexander, 2012). Indeed, a prevailing assumption of the literature has been that insight occurs by merit of one solving an "insight problem" (Topolinski & Reber, 2010). To make meaningful progress toward understanding insight, we must first explore the cognitive mechanisms involved in problems in which it is reported.

One such class of problems are Compound Remote Associates (CRAs) (Bowden & Jung-Beeman, 2003). The CRA task was developed as a modified version of the Remote Associates Test (RAT) (Mednick, 1962), which has been correlated with performance in insight problems. The difference between the original RAT and CRAs is that the latter only uses structural associates based on syntax (Worthen & Clark, 1971). In CRAs, people are presented three cue words and must produce a solution word that is common to all three, forming compound words and phrases. For example, the solution to the triad "COTTAGE, SWISS, CAKE" is CHEESE (forming "COTTAGE CHEESE," "SWISS CHEESE," and "CHEESECAKE," respectively). The task is designed such that a solver must break free of high-frequency associations to access globally satisfactory solutions.

CRAs have many advantages over classic insight problems: 1) they have large, normed databases, 2) many can be completed in single, short experimental sessions, 3) they can be solved with and without insight, 4) people have reliably demonstrated that they can make subjective judgments of insight regarding them, 5) they can be used in neuroimaging studies to identify the neural correlates of insight, and 6) they can be supplemented with time-based measures of solution latencies. As a result, they have been widely used to explore various cognitive domains, such as intuition (Topolinski & Strack, 2008), sleep (Cai et al., 2009), and computational/deep learning (Olteteanu, Gautam, & Famomir, 2015).

Much of the past research on the RAT and CRAs has defaulted to a correlational account that simply assumes insight and ignores the underlying processes that may drive it. This problem was highlighted by Topolinski and Reber (2010), who pointed out that many researchers neglect to explain the *phenomenology* of insight yet rely on it as a sufficient condition.

Recent studies have attempted to mend this by modelling CRA performance. Gupta, Jang, Mednick, and Huber (2013) were among the first to provide a formalized account of individual differences in CRA search behavior. They employed a norm-based model that defined the best guess at solutions based on the average of cues in the Word Association Space (WAS) (Steyvers, Shiffrin, & Nelson, 2005). This was contrasted with a frequency-biased model that assumes people's search is biased by word fluency, based on Griffiths, Steyvers, and Firl's (2007) work with PageRank and associative frequency. As predicted, they found that the probability of a given response is biased toward high-frequency words. Thus, people perform poorly if they're biased in favor of high-frequency incorrect words, precluding access to low-frequency correct responses.

This work was extended by Olteteanu and Falomir (2015), who developed the comRAT-C; a computational model that solves compound RAT queries, based on a cognitive theoretical framework for creative problem solving (CreaCogs) (Olteteanu, 2016). The knowledge base (KB) comprising the CRAs themselves used language data (2-grams pruned for relevance) from the Corpus of Contemporary English (COCA). They found that the comRAT-C used a convergence process similar to that of human solvers, and that the frequency of cues in the KB influences responses. The comRAT-C was able to correctly solve 64 of the 144 items in Bowden and Jung-Beeman's (2003) list of normed CRAs, in addition to suggesting unlisted, yet plausible solutions in more than 20 cases – suggesting its own form of creativity. Overall, their study laid a solid computational framework for formalizing the processes in CRA problem solving.

A promising experimental approach was taken by Smith, Huber, and Vul (2013), who used Latent Semantic Analysis (LSA) to evaluate the similarity between people's guesses, word cues, and answers. They accomplished this by having participants enter every word considered while searching for the answer, regardless of their correctness. By doing this, they focused on the search processes used when generating candidate answers through a probabilistic sampling framework. They found sequential dependencies between responses in a problem, with subjects generating semantically similar chains of responses. Additionally, people seemed to focus primarily on one cue at a time. However, their procedure assumes that guesses accurately reflect the implicit nature of the search, even though the very act of conscious report may alter the search process.

The main body of work on CRAs has focused on associative aspects - not the requisite that responses be syntactic *compounds* (with the notable exception of Olteteanu and Falomir (2015)). Indeed, research has largely ignored the morphological properties of the compounds themselves and how they affect performance and the likelihood of reported insight. We have thus failed to adequately address a critical aspect of their character. This approach has potentially restricted us from discovering how people attain insight in these problems. A look at the nature of compounds and their lexical elements is necessary to better understand the underlying cognitive processes involved in these problems.

## Compound Word Research

Early work on compound words used a lexical decision paradigm (Taft & Forster, 1975), which measures peoples' response times (RT) in classifying words and nonwords. One such study found that only the lexical status of the first constituent word in a compound affects processing, with longer RT for word-word and word-nonword pairs (e.g., DUSTWORTH, FOOTMILGE) than nonword-word and nonword-nonword pairs (e.g., TROWBREAK, MOWDFLISK) (Taft & Forster, 1975). Thus, it appears that morphological decomposition takes place when processing compound words, instead of the words being stored and retrieved as a whole.

There has been considerable work on visual word recognition in recent years facilitated by databases containing lexical characteristics and behavioral data, such as latencies of word naming and lexical decisions for large sets of words (e.g., Balota et al., 2007) and investigations of word length (New, Ferrand, Pallier, & Brysbaert, 2006). Though initially focused on monosyllabic and monomorphemic words, this work has been extended to address processing in multisyllabic words (Yap & Balota, 2009) and English compound words.

Research suggests that English compounds are processed differently from length and frequency-matched monomorphemic words. For instance, both semantically-transparent compounds (e.g., ROSEBUD) and opaque compounds (e.g., HOGWASH) are processed more quickly than their monomorphemic counterparts (e.g., GIRAFFE) (Ji, Gagné, & Spalding, 2011). This sense of morphological complexity has ignited debate in the psycholinguistic literature, with competing perspectives on compound representation and processing (see Fiorentino & Poeppel, 2007).

The current study investigates the roles of three lexical properties involved in compound processing and, by extension, CRAs: word familiarity, semantic transparency, and lexeme meaning dominance. Thus, we investigated if and how they differentially affected CRA performance and the likelihood of insight. To do this, we used Juhasz, Lai, and Woodcock's (2015) database of 629 compound words to construct 21 novel CRA problems. This database, which adapted items from the English Lexicon Project (ELP: Balota et al., 2007), compiled subjective ratings for six properties believed to affect morphological processing. The questionnaires used by these authors are available in their Supplementary Materials. We will now briefly explore each of these selected properties and justify their inclusion in the study.

**Familiarity** Whole word frequencies may be interpreted as analogous to whole word access and have thus been studied in compound word recognition (Juhasz, Lai, & Woodcock, 2015). However, English compound frequencies tend to be low relative to other languages, resulting in experimental challenges and a consequential gap compared to Dutch (Kuperman, Schreuder, Bertram, & Baayen, 2009) and Finnish (Kuperman, Bertram, & Baayen, 2008) counterparts. Rated familiarity can be regarded as a measure of subjective frequency and has been demonstrated to affect word recognition in English monomorphemic words. In particular, familiarity has been shown to influence eye fixation durations, along with word frequency (Juhasz & Rayner, 2003). This was further demonstrated in an experiment by Juhasz, White, Liversedge, and Rayner (2008), which found that familiarity affected gaze duration for both long (ten or more letters) and short (seven or fewer letters) English compound words.

We thus contend that ratings of familiarity can be used as a subjective proxy for word frequency and have a role in affecting morphological processing and CRA performance.

**Semantic Transparency** Semantic transparency also plays an important role in how compounds are processed and represented (Libben, 1998). A fully transparent compound is one in which both constituents contribute to the meaning of the compound word (e.g., SUNLIGHT), while a fully opaque compound is one in which neither constituent contributes to its meaning (e.g., FLAPJACK). There are also partially-opaque compounds, in which only one constituent contributes to the compound's meaning (e.g., JAYWALK, CHEAPSKATE) (Juhasz et al., 2015).

Libben (1998) proposed a model in which semantic transparency is represented in two distinct ways: the semantic relationship between the meaning of a constituent morpheme within a compound, and the meaning of the morpheme independent of it. For example, the opacity of the compound SHOEHORN results from HORN not being transparently related to the compound as a whole, whereas SHOE is fully transparent. Thus, it is classified as a T-O compound (wherein T = transparent, O = opaque). Compounds require some level of semantic transparency to be tied to semantic representations of their lexemes. Using a lexical decision task, Libben, Gibson, Yoon, and Sandra (2003) found that fully opaque and T-O compounds were responded to more slowly than other compound types, though there was a significant priming effect on all four compound types relative to neutral primes.

Research has demonstrated that semantically transparent compounds are especially susceptible to morphological decomposition, and that semantic priming only seems to occur when there is at least one transparent lexeme. Using Dutch compounds, Sandra (1990) used semantic associates of constituents as primes for transparent (e.g., BIRTHDAY primed by DEATH), opaque (e.g., SUNDAY primed by MOON), and pseudo-compounds (e.g., BOYCOTT primed by GIRL). Facilitatory priming effects were only observed for constituents in transparent compounds.

**Lexeme Meaning Dominance** Compared to other languages, English compound words tend to be right-headed (i.e., the second constituent word – or lexeme - is the semantic head of the compound). This lexemic dominance primarily defines the meaning of the compound. In a study by Inhoff, Starr, Solomon, and Placke (2008), location and word frequencies of lexemes were manipulated in lexical decision, naming, and sentence reading tasks. They found an effect for larger word frequency for the dominant lexeme in each task. Lexeme dominance also affected first fixations on compound words. These results suggest the headedness of a compound affects how it is recognized and subsequently processed.

Since all the word cues presented in the CRAs in this experiment are the second lexemes, their contribution to the overall meaning of the compound should affect the speed of access when solving each problem.

## The Current Study

In accordance with the evidence above, we predicted that CRA problems beginning with word cues that 1) are the most familiar, 2) are the most semantically transparent, and 3) have right-headed lexeme dominance would result in the highest levels of performance and reporting of insight.

To test this, we staggered the presentation of word cues on-screen, with cues either increasing or decreasing in ratings for the relevant lexical domain. Thus, we actively constrained and manipulated the search processes used by solvers. As CRA triads are commonly presented at once, this presents an experimental departure that, we hypothesize, differentially affects performance and captures some of the latent features of this process. To our knowledge, this is one of the first studies to actively manipulate cue presentation in CRAs in such a way with precise behavioral predictions.

Another departure is in how problems are scored. CRAs are typically scored according to whether a submission 1) conforms to all three cue words and 2) conforms to the suggested response of the researchers, precluding other "incorrect", yet plausible responses. This does not allow for investigation of partially correct problems, in which fewer than three cues are satisfied by the solution candidate. We address this issue using a lexicon to test whether submitted responses form valid compounds against each individual cue presented, either as a prefix or suffix. This allows for a more comprehensive picture of the processes and strategies employed in such problems.

Together, this study contributes to the CRA literature in three major ways: 1) it uses a staggered presentation of word cues, facilitating semantic activation and lexical search behavior, 2) it investigates the morphological properties of the compounds themselves, and 3) it uses partial scoring for each word cue. The goal of the study is to determine how the aforementioned lexical properties affect solution retrieval and if they influence performance and the probability of reported insight.

## Methods

### Participants

Each experimental condition was composed of two counterbalanced groups, comprising six groups total. All participants ($n$ = 128) were University of California, Irvine undergraduate students, who were awarded course credit through the SONA system for their role in the study. The age distribution was 18-21 ($n$ = 110), 22-25 ($n$ = 11), 26-30 ($n$ = 5), and 31-40 ($n$ = 2). Everyone identified as a native English speaker, with 53 participants identifying as multilingual (though additional languages spoken were not specified).

In the *Familiarity* condition, Group 1 consisted of 23 participants ($n$ = 21 females), and Group 2 consisted of 22 participants ($n$ = 14 females).

In the *Lexeme Meaning Dominance* condition, Group 3 consisted of 21 participants ($n$ = 14 females), and Group 4 consisted of 21 participants ($n$ = 19 females).

Lastly, in the *Semantic Transparency* condition, Group 5 consisted of 21 participants ($n$ = 16 females), and Group 6 consisted of 20 participants ($n$ = 18 females).

Fourteen participants were excluded from the final analysis, as they did not meet the criteria of answering at least two of the three practice problems correctly.

### Materials

We constructed 21 novel CRA problems from compounds that had at least three common stems (thus forming three cues with a common solution). For example, there are 10 compound words in Juhasz et al.'s (2015) database with the shared prefixed stem FOOT. The mean ratings for whatever variable was in question (on a 1-to-7 scale for familiarity and transparency and on a 1-to-10 scale for lexeme meaning dominance) were then sorted in descending order and the words with the highest and lowest values were selected. The mean of these two values was then calculated and the compound word with that value or its closest approximate was selected as the middle term. Using the same example of FOOT for the variable of familiarity: FOOTPRINT has the highest value at 7, FOOTPATH has the closest approximate to the mean with a value of 5.85, and FOOTHILL has the lowest value at 4.71. This forms the CRA problem "PRINT, PATH, HILL," with the solution FOOT. All compounds in this database begin with a prefixed solution stem. Thus, unlike other studies, the solution is always the first lexeme in the compound.

In the event of a tie between two compound word values, the compound with the closest letter length to the other two words was selected. If the competing compound had the same length, the tie was broken by identifying which one more closely matched the mean age of acquisition value of the other two compounds.

Due to the limited number of candidate items, some words were repeated in both problem and solution terms. For example, PORT occurs in the problems "PORT, BASE, SICK" and "FOOD, PORT, BOARD." There was also an instance of a having the same phonetic representation (WASTE and WAIST). Participants were told that words may occur more than once both as cues and as solutions.

Each condition was counterbalanced so that problems were presented in both ascending and descending order across two groups. This was done to control for potential order effects.

### Procedure

Participants were given instructions and a working definition of "insight" (*Insight occurs when the answer suddenly pops into your head, accompanied by a strong burst of positive emotion ("aha!").*). They were then given an example CRA problem ("CREAM, SKATE, WATER," solution = ICE) and were asked to complete three practice problems with feedback. All four of these problems were pulled from the Bowden and Jung-Beeman's (2003) set of CRA norms and were the four easiest problems with uniformly prefixed solution stems.

The experiment was conducted using a MATLAB interface. Word cues were presented sequentially with 5-second delays between each cue. The first word cue appeared in the left-center of the screen, the second appeared in the center, and the third in the right-center. Cues remained on-screen after their presentation for the remainder of the trial. Figure 1 demonstrates the display of a typical problem trial.

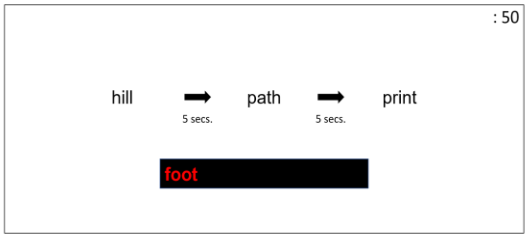


Figure 1: Example of problem trial.

Each trial lasted for one minute. A countdown timer appeared on the top right-hand corner of the screen when 50 seconds remained and turned red when 10 seconds remained. Participants typed their responses in a black box below the cues. They were encouraged to answer as quickly and as accurately as possible. They could submit their response at any time following the presentation of the third cue. Participants were forced to proceed after the minute had expired and whatever was typed into the solution box was accepted as the submitted response.

Following each problem trial, participants were asked to report the level of insight they experienced on a scale of 1 ("no insight") to 7 ("complete insight"). They were also reminded of its operational definition on the bottom of the screen.

At the end of the experiment, participants were asked to provide a brief (150-word max.) description of what strategies they used to solve these problems. We also asked them to describe the difference they felt between solving problems with and without the feeling of insight. This was done to determine individual differences in reporting criteria and as a check for cross-validity with our definition. This data will also be evaluated to inform future, related experiments.

Participants were scored based on how quickly and accurately they responded to each problem.

## Results and Discussion

First, we tested the hypothesis that presentation order of cues according to ratings in each lexical condition would affect performance. These results are shown in Figure 2. Note: "Direction" denotes whether the cue presentation sequentially increased or decreased for the lexical property in question (that is, "Down" indicates that the first cue had the highest rating for the property, while "Up" started with the lowest rating). It appears that the only observed difference was in familiarity, with a higher proportion of problems successfully solved when they began with the most familiar word cue ($M = 0.383$, $SD = 0.126$), rather than the least familiar cue ($M = 0.301$, $SD = 0.161$, $t(226) = 4.304$, $p < .001$, $d = 0.570$). The estimated Bayes factor suggested that the data were .001:1 in favor of the alternative hypothesis, suggesting decisive evidence for a presentation order effect (Jeffreys, 1961). While this finding was not shared by the other properties (lexeme meaning dominance and semantic transparency), there are several other important findings – some of which challenge widely-accepted assumptions regarding the "special process" view (Bowden, Jung-Beeman, Fleck, & Kounios, 2005) of the insight phenomenon.

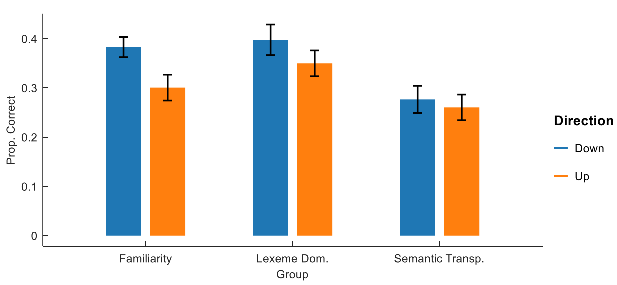


Figure 2: Differences in performance for each lexical property.

For further analysis, we used English compounds derived from the Touchstone Applies Science Associates (TASA) corpus and derived a lexicon of over 122,000 words, including hyphenated compounds. We used this lexicon to test whether a submitted response forms a valid compound against each individual cue presented, either as a prefix or suffix. The results of individual cue matches are shown in Figure 3, which demonstrates that the proportion correct for suffixes is smaller than that of prefixes. Further, submitted responses had a smaller likelihood of being valid prefixes for middle cues ($M = 0.351$, $SD = 0.162$) than for first cues ($M = 0.411$, $SD = 0.182$) and last cues ($M = 0.402$, $SD = 0.174$, $F(2,228) = 8.049$, $p < .001$). The estimated Bayes factor suggested that the data were .032:1 in favor of the alternative hypothesis, or rather, 31.25 times more likely to occur under the model including an effect for cue position than the model without it, providing strong evidence for its effect. This suggests that participants were alternating between cues when attempting to generate a solution, rather than using parallel processing. One possible explanation is that since cue presentation was staggered – and thus their search was guided – there may be primacy and recency effects whereby they were able to test and generate more candidate solutions following the first word cue, then worked backwards once all cues were presented using the third cue.

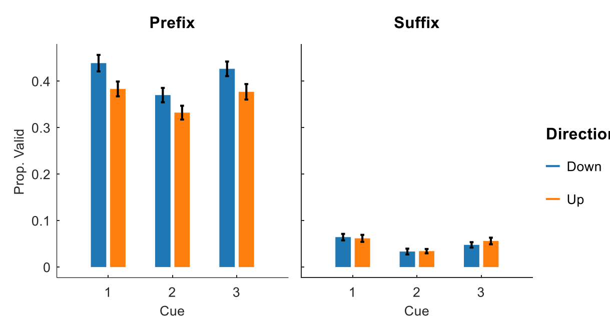


Figure 3: Proportion of valid prefixed (left) and suffixed (right) responses for each word cue position, according to lexicon.

Another interesting finding was that ratings of insight decreased as time elapsed throughout trials, as demonstrated in Figure 4. This finding holds for both correct and incorrect trials. This seemingly challenges the popular assertion that there must be a period of impasse, or mental block, preceding the experience of insight (Ohlsson, 1992). To the contrary, there were higher ratings of insight in the immediate time following the presentation of all three cues (i.e., 10-20 seconds) than in the time before the end of each trial (50-60 seconds). It is possible that participants simply rated solutions that they perceived to be correct as insightful *de facto* (hence, being submitted quickly), and correctly rejected the occurrence of insight for incorrect solutions proffered as a final guess before trials ended.

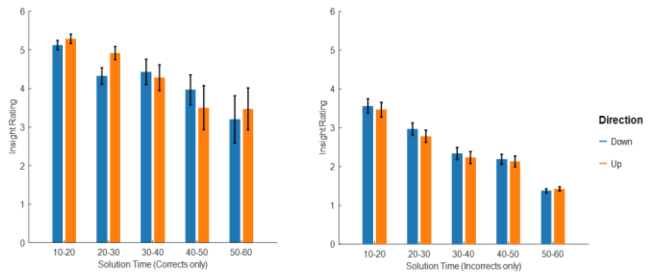


Figure 4: Insight ratings as a function of solution time (in secs.) for correct solutions (left) and incorrect solutions (right).

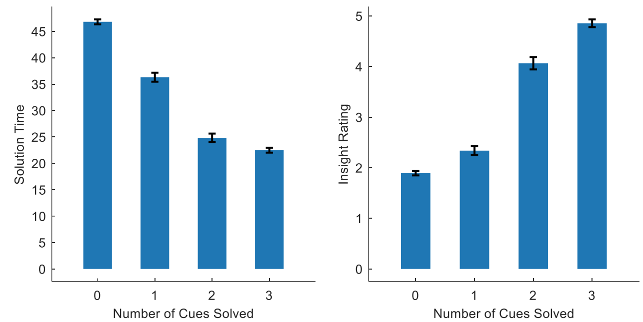


Figure 5: Solution time (in secs.) as a function of cues solved (left); Insight ratings as a function of cues solved (right).

Finally, there is the reporting of insight, itself. As demonstrated in Figure 5, fewer cues were likely to be solved as more time elapses in trials. The magnitude of reported insight also increased along with the number of cues correctly solved. Rather than an all-or-none experience – the "sudden, certain burst" frequently reported and used as a necessary criterion (Chronicle, MacGregor, & Ormerod, 2004) - it appears that participants used ratings of insight to indicate confidence in their answers. Indeed, these ratings increased as a function of the number of cues their proposed solution fit. There is not the presence of absolute insight for totally correct trials (in which all three cues are satisfied by the proposed solution), nor the absence of insight if this is not achieved. Rather, it exists on a continuum. This suggests more of an analytic approach, in which participants reliably monitor their progress in each problem and the likelihood of success using insight as a proxy for said progress. This contrasts previous research which states that incremental feelings of "warmth" do not precede moments of insight and are instead relegated to analytic or non-insightful problem solving strategies (Metcalfe & Wiebe, 1987). It should be noted again that a property of CRAs is that they can be solved with *or* without insight. What we argue here is the usefulness of insight ratings in CRAs to indicate perceived progress.

One limitation to the current work is that it used novel CRAs instead of those with established norms for difficulty and magnitude/frequency of reported insight (such as Bowden & Jung-Beeman, 2003). Applying lexical ratings to such a database for the dimensions present in Juhasz, Lai, and Woodock (2015) would be informative for future studies. Future research could also have subjects generate their own list of compounds given a set of word stems. Through doing this, researchers could collect latency data for how long people take to produce words, indicating their availability in memory. Researchers could also use LSA to analyze these participant-generated sequences of compounds to describe search behavior. These data could be applied across participants to establish cross-reliability and a more naturalistic set of items with norms.

All problems in the current study have suggested solutions that are the first lexeme in the compound. Since stem placement seems to matter in the processing of compounds (Taft & Forster, 1975), it may be beneficial to compile and use compounds with common prefixes and suffixes in future studies.

The self-identified magnitude of insight in our study was still based on subjective report. While this study focuses on the cognitive processes underlying these problems, rather than attempting to formalize insight in a significant manner, similar studies attempting to so may wish to include neural and/or physiological covariates to identify correlates of insight (e.g., EEG, fMRI, skin conductance, eye-tracking) (see Bowden et al., 2005 for suggested neurocognitive approaches). Future studies should also explore participants' differences in reporting thresholds, as one person may be more willing to identify the occurrence of insight than another. These individual differences could be applied to a signal detection theory model.

This study offers modest progress into understanding the linguistic contributors to CRA processing. There are other factors that should be investigated, such as if compounds with noun-noun links and adjective-noun links differentially affect performance. Other variables to investigate are word length effect (New et al., 2006), imageability, age of acquisition, sensory experience, or a combination of the above.

There may also be a reading direction effect present, as cue presentation always proceeded from left-to-right on the screen. To circumvent potential perceptual biases, future studies using a similar design may benefit from counterbalancing the order of reading direction, as well.

Lastly, it is important to be remain cognizant that not all insight problems are the same, and the phenomenology in CRAs may differ from that of other insight problems. It would be premature to make any sweeping statements about modeling insight from discoveries made in one class of problems.

## Conclusion

If we are to solve the problem of insight, we must better understand the cognitive processes underlying the methods we use to study it. Since we've largely neglected to explore these commonly-used procedures, we've defaulted to assumptions that they are "insight problems" simply because they elicit feelings of insight (based on the many and inconsistent criteria of researchers). While there have been both promising empirical and theoretical attempts to address this problem in recent years, much work remains. Better understanding the driving mechanisms, including lexical properties, within CRA problem solving will further inform us about how creativity is exercised and, perhaps, how insight is attained.

## References


Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., …. Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445-459.

Batchelder, W. H., & Alexander, G. E. (2012). Insight problem solving: A critical examination of the possibility of formal theory. *The Journal of Problem Solving*, *5*, 56-100.

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, and Computers*, *35*, 634-639.

Bowden, E., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Science*, *9*, 322-328.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), Language production: Vol., 2. London: Academic Press.

Cai, D. J., Mednick, S. A., Harrison, E. M., Kanady, J. C., & Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *PNAS, 106*, 10130- 10134.444.

Chronicle, E. P., MacGregor, J. N., & Ormerod, T. C. (2004). What makes an insight problem? The roles of heuristics, goal conception, and solution recoding in knowledge-lean problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 14–27.

Corpus of Contemporary American English (COCA): http://corpus.byu.edu/coca

Fiorentino, R., & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive Processes*, *22*, 953-1000.

Griffiths, T. L., Steyvers, M., & Firl, A. (2007). Google and the mind: Predicting fluency with PageRank. *Psychological Science*, *18*, 1069-1076.

Gupta, N., Jang, Y., Mednick, S. C., & Huber, D. E. (2012). The road not taken: Creative solutions require avoidance of high-frequency responses. *Psychological Science*, *23*, 28-284.

Inhoff, A. W., Starr, M. S., Solomon, M., & Placke, L. (2008). Eye movements during the reading of compound words and the influence of lexeme meaning. *Memory and Cognition*, *36*, 675-687.

Jeffreys, H. (1961). *Theory of probability* (3rd Ed.). Oxford, UK: Oxford University Press.

Ji, H., Gagné, C. L., & Spalding, T. L., (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, *65*, 406-430.

Juhasz, B. J., White, S. J., Liversedge, S. P., & Rayner, K. (2008). Eye movements and the use of parafoveal word length information in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1560-1579.

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1312-1318.

Juhasz, B. J., Lai, Y. H., & Woodcock, M. L. (2015). A database of 629 English compound words: ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods*, *47*, 1004-1019.

Kounios, J., & Beeman, M. (2014). The cognitive neuroscience of insight. *Annual Review of Psychology*, *65*, 71-93.

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, *23*, 1089-1132.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 876-895.

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language*, *61*, 30-44.

Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, *84*, 50-64.

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*, 220-232.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight an noninsight problem solving. *Memory and Cognition*, *15*, 238-246.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, *13*, 45-52.

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. T. Keane & K. J. Gilhooly (Eds.), *Advances in the psychology of thinking* (Vol. 1). New York: Harvester Wheatsheaf.

Olteteanu, A. M., Gautam, B., & Falomir, Z. (2015). Towards a visual remote associates test and its computational solver. In *AIC*.

Olteteanu, A. M., & Falomir, Z. (2015). comRAT-C: A computational Remote Associates Test solver based on language data and its comparison to human performance. *Pattern Recognition Letters*, *67*, 81-90.

Olteteanu, A. M. (2016). From Simple Machines to Eureka in Four Not-So-Easy Steps: Towards Creative Visuospatial Intelligence. In V. C. Müller (Ed.), Fundamental Issues of Artificial Intelligence. Cham: Springer International Publishing.

Sandra, D. (1990). In the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *The Quarterly Journal of Experimental Psychology Section A*, *42*, 529-567.

Smith, K. A., Huber, D. E., & Vul, E. (2013). Multiply-constrained semantic search in the Remote Associates Test. *Cognition*, *128*, 64-75.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2005). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Experimental Cognitive Psychology and its Applications*. Washington: American Psychological Association.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*, 638-647.

Topolinski, S., & Reber, R. (2010). Gaining insight into the "Aha" experience. *Current Directions in Psychological Science*, *19*, 402-405.

Topolinski, S., & Strack, F. (2008). Where there's a will – there's no intuition. The unintentional basis of semantic coherence judgements. *Journal of Memory and Language*, *58*, 1032- 1048.

Worthen, B. R., & Clark, P. M. (1971). Toward an improved measure of remote associational ability. *Journal of Educational Measurement*, *8*, 113-123.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*, 502-529.