Wong, E. F., Schauer, G. F., Gordon, P. C., & Holyoak, K. J. (2019). Semantic and visual interference in solving pictorial analogies. In A. Goel, C. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Semantic and Visual Interference in Solving Pictorial Analogies

Emily F. Wong (emilyfwong@ucla.edu)
Department of Psychology, University of California, Los Angeles

Guido F. Schauer (guido.f.schauer@ucla.edu)
Department of Psychology, University of California, Los Angeles
Los Angeles, CA, USA

Peter C. Gordon (pcg@email.unc.edu)
Department of Psychology, University of North Carolina at Chapel Hill

Keith J. Holyoak (holyoak@lifesci.ucla.edu)
Department of Psychology, University of California, Los Angeles

Abstract

Neuropsychological investigations with frontal patients have revealed selective deficits in selecting the relational answer to pictorial analogy problems when the correct option is embedded among foils that exhibit high semantic or visual similarity. In contrast, normal age-matched controls solve the same problems with near-perfect accuracy regardless of whether high-similarity foils are present (in the absence of speed pressure). Using more sensitive measures, the present study sought to determine whether or not normal young adults are subject to such interference. Experiment 1 used eye-tracking while participants answered multiple-choice 4-term pictorial analogies. Total looking time was longer for semantically similar foils relative to an irrelevant foil. Experiment 2 presented the same problems in a true/false format with emphasis on rapid responding and found that reaction time to correctly reject false analogies was greater (and errors rates higher) for those based on semantically or visually similar foils. These findings demonstrate that healthy young adults are sensitive to both semantic and visual similarity when solving pictorial analogy problems. Results are interpreted in relation to neurocomputational models of relational processing.

Keywords: Analogy, semantics, perception, interference, eye-tracking, reaction time

Introduction

Relational reasoning—inferential processes constrained by the relational roles that entities play rather than the specific features of those entities—is a hallmark of human cognition. The basic components of relational processing have been investigated using a wide variety of analogy tasks. The simplest format for analogies involves four terms, expressed as either words or pictures, in the form *A:B::C:?*, where the task is to complete the analogy by selecting the best *D* term from a small set of options. By varying the alternative options, it is possible to assess the degree to which analogical reasoning is influenced by foils that pit semantic and/or visual similarity of individual concepts or objects against relational similarity between pairs of concepts or objects. In the example shown in Figure 1, the task is to select the analogical option (*fish bowl*, based on the relation "lives in" that matches the *A:B* relation) from among a semantic distractor similar to the *C* term (*fish hook*), a visual distractor (*rocket*) and an unrelated option (*camera*).

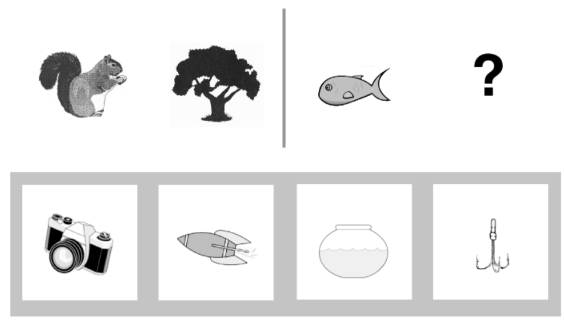

Figure 1. Example of a 4-term pictorial analogy with four alternatives, used in the present experiments (from Krawzcyk et al., 2008).

Krawczyk et al. (2008) administered a set of picture analogies (from which the example shown in Figure 1 is drawn) to neuropsychological patients suffering from frontotemporal lobar degeneration (FTLD) and age-matched neurotypical controls (mean age approximately 60 years). Some of these problems were adapted from an earlier set created by Goranson (2002), and hence are dubbed the Goranson Analogy Test (GAT). In the study by Krawczyk et al., problems were administered one at a time, without speed pressure. In one problem set the options included distractors as in Figure 1; in an alternative set the semantic and perceptual distractors were replaced by two additional unrelated options.

For the set with similar distractors, patients with frontal-variant FTLD were correct on only 49% of the problems, rising to 84% correct for the set without similar distractors. An additional group of patients with temporal-variant FTLD showed a similar level of impairment regardless of whether similar foils were present, suggesting a general semantic deficit (see also Morrison et al., 2004). When similar distractors were present the patients with frontal damage selected similar distractors (mainly semantic, but also visual) more often than control participants. Indeed, the control group achieved near-perfect accuracy (98% correct). Thus, frontal damage appeared to selectively impair the ability to inhibit responding to pictorial analogy problems on the basis of superficial object similarity.

The near-perfect performance of the control participants in solving pictorial analogies even in the presence of similar distractors raises the question of whether and how cognitively unimpaired adults screen out object similarity (both semantic and visual) so as to focus on similarity of relations. Adults sometimes respond on the basis of object similarity when comparing more complex visual scenes (Markman & Gentner, 1993; Walz et al., 2000); however, the simple format of four-term pictorial analogies may allow non-relational information to be filtered out at a very early processing stage, so that choice of the analogical solution is not influenced by the presence of similar but non-relational foils. Alternatively, more sensitive measures may reveal evidence of response competition based on different varieties of similarity.

In two experiments, we investigated this question by administering versions of the GAT analogies used by Krawczyk et al. (2008) to healthy young adults. Eye-tracking methods provide one avenue for investigating online processing that occurs during analogical reasoning prior to making an overt decision (Gordon & Mozer, 2006; Glady, French, & Thibaut, 2016; Hayes, Petrov, & Sederberg, 2011; Vendetti et al., 2017). Accordingly, in Experiment 1 we collected data on gaze durations for the various response options while solving the GAT problems.

Another potentially more sensitive measure is reaction time (RT) to solve analogies under speed pressure. In Experiment 2 we changed the format of the GAT problems from four-alternative forced choice to true/false. For each of the original problems, each of the three foils was used to create a false picture analogy in the form *A:B::C:D*. In addition, participants were instructed to respond as quickly as possible. If semantic and/or visual similarity is screened out easily, then the various types of false analogies should take about the same length of time to reject. However, if college students are unable to avoid processing more superficial types of similarity, then decisions about false analogies in which the *D* term is similar (semantically or visually) to the *C* term may be relatively slow and error-prone.

Experiment 1

If superficial similarity intrudes into analogical reasoning for healthy adults, then they may spend more time looking at semantic and/or visual distractors than at an unrelated option.

Method

Participants. Participants were 32 undergraduates (24 female), mean age 20.4 years (range: 17–34) from the University of California, Los Angeles (UCLA), with normal or corrected to normal vision. They received course credit for participating.

Materials. Picture analogies were based on the 18 GAT problems used by Krawczyk et al. (2008). Two of these served as practice items, and 16 as experimental items. As in the Krawczyk et al. study, two sets of the 16 problems were created, one of which included similar foils and one of which replaced the semantic and visual foils with unrelated options.

Procedure. Pictorial analogies were presented on a computer screen one at a time. The size of each individual image (framed by a gray box) was 128 x 128 pixels (one-tenth of the screen width). A fixation cross was presented for 2 s, followed by the problem. The problem remained on until the participant pressed one of four response keys (corresponding to letters F, G, H, and J) to indicate which of the four alternatives was the correct analogical solution. When a response was made, the screen showed the reverse grayscale image for .25 s, after which the next trial began. Instructions did not emphasize speed of responding. During the experiment eye-tracking data were recorded using an Eyelink II gaze tracker (SR Research Ltd., Mississauga, Ontario, Canada), running under Eyelink Toolbox, PsychToolbox, and MATLAB on dual PCs. No feedback was provided.

For each participant, eight problems were included in the set with similar distractors (Distractor condition), and the other eight in the set with only unrelated foils (No-Distractor condition). Assignment of problems to set was counterbalanced across participants, as was the order of the four response options for each problem. Presentation order of the problems was randomized for each participant.

Results

Data were missing for one participant, who was excluded from analyses. Accuracy overall was 92% correct and did not vary reliably across the Distractor and No-Distractor conditions.

To guide analyses of eye movements, an invisible square of size 192 x 192 pixels around each individual image was defined as the location of that image. Figure 2 presents an example of a pattern of eye movements for an individual analogy problem in the Distractor condition.

To provide evidence of a possible pre-decisional influence of superficial similarity, we focused on dwell time (i.e., total looking times summed across all fixations) for each response option. Figure 3 plots the mean dwell time for each option in both the Distractor and No-Distractor conditions.

Participants' mean total time looking for each of the three foil images, in descending order, was: semantic foil (522 ms, *SE* = 38.2), visual foil (518 ms, *SE* = 54.9), and unrelated foil (404 ms, *SE* = 31.1). Overall, there was significant variation in dwell times depending on the foil condition, $F(2,60) = 3.93$, $p = 0.025$, $\eta^2 = .12$. Individual comparisons between conditions are reported with Bonferroni-corrected *p*-values. Semantic foils had longer dwell times relative to unrelated foils, $t(30) = 3.67$, $p < .001$, $\eta_P^2 = .31$.

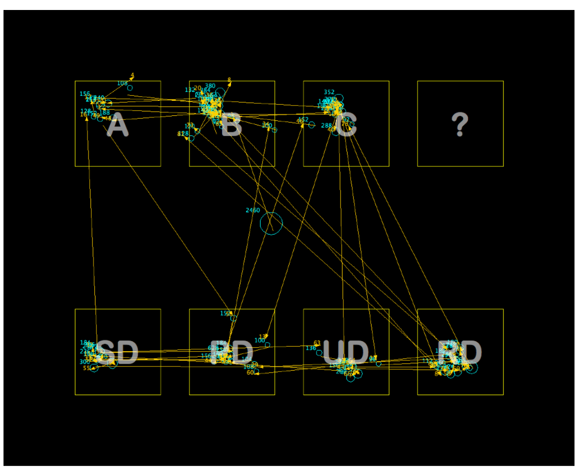

Figure 2. Example of pattern of eye movements during solution of a picture analogy. The above boxes (not visible to participants) indicate regions around the four images in the problem (*A*, *B*, *C*, *?*) and the four response options: semantically similar (S), visually similar (P), unrelated (U), and relational (R, the correct response). The D on each option label indicates this trial is from the Distractor condition.

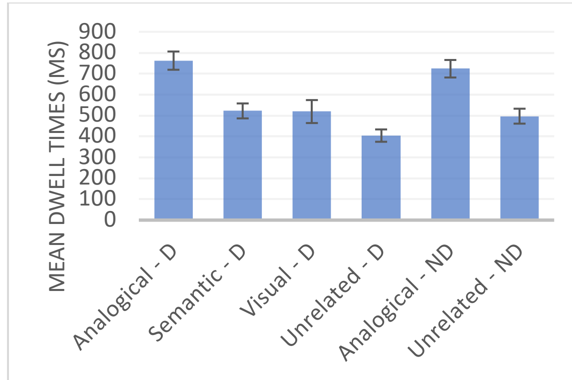

Figure 3. Total dwell time for each type of response image. The dwell time for Unrelated–ND is the mean across the three unrelated options provided in the No-Distractor condition. Error bars indicate +/- 1 standard error of the mean.

Visual foils also tended to have longer dwell times relative to unrelated foils. However, due to greater error variance in the visual foil condition, the latter difference was not reliable after the Bonferroni correction, $t(30) = 2.24$, $p = .098$, $\eta_P^2 = .14$. Dwell times for the two types of similar foils did not differ, n.s.

The eye-tracking data from Experiment 1 provide clear evidence that healthy adults are influenced by the presence of semantic and possibly visual distractors. Although response accuracy was high even in the presence of distractors, participants looked longer at semantically similar foils than at an unrelated option, suggesting that participants were sensitive to superficial similarity prior to making a decision.

Experiment 2

Experiment 2 used the same basic GAT analogies as in Experiment 1 but changed the format from 4-alternative forced choice to true/false. Instead of eye-tracking, the main dependent measure was RT to evaluate the problems under speed pressure.

Method

Participants. A total of 60 UCLA undergraduates (83% female) participated in the experiment. Their mean age was 20.8 years (range: 18–28), with normal or corrected to normal vision. They received course credit for participating.

Materials and Procedure. The experiment was conducted using a computer to display problems and record responses. The materials were based on the GAT problems used by Krawczyk et al. (2008). Each original problem was used to generate four true/false problems, each showing four pictures. As shown in Figure 4, in each problem the *A:B* pair appeared at the top of the display and the *C:D* pair on the bottom. The *D* picture was either the correct analogical response (true), the semantic foil (false), the visual foil (false), or the unrelated option (false). Thus 25% of the problems were true analogies and 75% were false.

A set of four practice problems was created, using two of the GAT problems plus two additional problems taken from other sources. For the actual test trials, 16 analogy sets were created, one from each of the remaining 16 GAT problems. Figure 4 shows one of these sets. This procedure resulted in a total of 64 analogy problems. Each participant solved all 64 problems (i.e., a within-subjects design). To control for order effects the items were counterbalanced in the following way. The 16 sets were randomly assigned in equal numbers to Group A, B, C, or D. Thus, there were a total of 4 sets in each of the groups. Then, four test combinations were formed (I, II, III, IV). Combination I included only the items in Group A that had the analogical option, items in Group B that had the semantic foil option, items in Group C that had the visual foil option, and items in Group D that included the unrelated choice. Combinations II–IV were formed in the same basic manner, completing the counterbalancing of the four problems in each set. The presentation order of combinations I–IV was then counterbalanced across participants. Finally, the order of items within each combination was randomized for each participant.

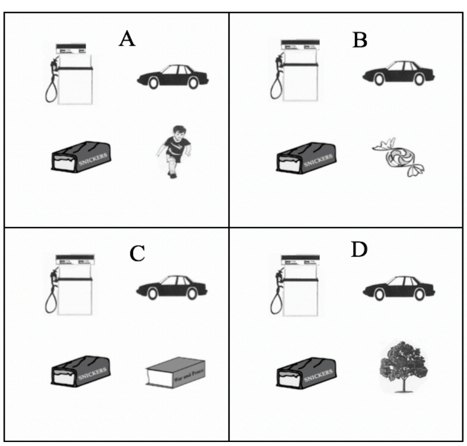

Figure 4. Example set of true/false picture analogies used in Experiment 2, created from the four alternatives of a single GAT problem. In each problem the *A:B* pair appears on top and the *C:D* pair on the bottom. The *D* term varies across problems. Panel A: Analogical (true); Panel B: semantic (false); Panel C: visual (false); Panel D: unrelated (false).

Participants were instructed to respond as quickly as possible while avoiding errors. They were told to press the *v* key to indicate “true” and the *n* key to indicate “false”. Before the actual test trials, participants completed the four practice items (illustrating each of the four basic problem types) and were given feedback after each one. The correct answer was presented for 3,000 ms. No feedback was provided after test trials. On each test trial, a fixation cross appeared in the center of the screen for 1,000 ms before presentation of the analogy problem. The analogy problem remained visible until a response was made. The screen then went blank for 1,000 ms, after which the next fixation cross was presented.

Results

Both error rates and RTs for correct trials were analyzed. In Experiment 1, where the task was a four-alternative forced choice without speed pressure, error rates were low. In Experiment 2, by contrast, the speeded true/false task led to a substantial error rate. The mean error rate was 25% for analogical (true) problems, 48% for the problems with a semantic foil (false), 16% for the problems with a visual foil (false), and 7% for the problems with an unrelated foil (false). For the three types of false problems, a one-way within-subjects ANOVA was highly significant, $F(2, 118) = 150.31$, $p < .001$, $\eta^2 = .72$. Error rates were higher for the semantic foils than the unrelated foils, $t(59)=13.63$, $p < .001$, $\eta_P^2 = .76$. Error rates were also higher for the visual foils than the unrelated foils, $t(59)=5.42$, $p < .001$, $\eta_P^2 = .33$. Finally, semantic foils produced more errors than visual foils, $t(59)=12.66$, $p < .001$, $\eta_P^2 = .73$.

Figure 5 presents the mean correct RTs for each problem type. On average, participants took 3,047 ms to correctly verify problems with the analogical completion, 3,396 ms to correctly reject problems with the semantic foil, 2917 ms to reject those with the visual foil, and 2,518 ms to reject those with the unrelated foil. A within-subjects one-way ANOVA provided strong evidence for variation in RTs among the three types of false analogies, $F(2, 58) = 34.98$, $p < .001$, $\eta^2 = .37$. A Bonferroni correction was again applied to pairwise comparisons between foil conditions. False problems with semantic foils took longer to reject than those with unrelated foils, $t(59) = 6.62$, $p < .001$, $\eta_P^2 = 0.43$. Those with visual foils also yielded longer RTs compared with unrelated foils, $t(59) = 6.27$, $p < .001$, $\eta_P^2 = .40$. Finally, problems with semantic foils produced longer RTs than those with visual foils, $t(59) = 4.46$, $p < .001$, $\eta_P^2 = .77$.

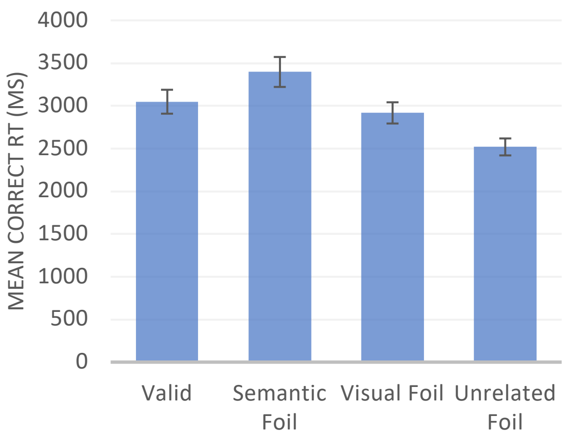

Figure 5. Mean correct RT for each type of picture analogy problem (Experiment 2). Error bars indicate +/- 1 standard error of the mean.

Discussion

The goal of the present study was to assess whether or not healthy young adults are influenced by semantic and/or visual similarity of distractors when solving four-term picture analogy problems. Previous work had indicated that in the absence of speed pressure, healthy older adults show little if any tendency to actually choose similar distractors over the correct, analogical option (Krawczyk et al., 2008). One possibility is that for reasoners with a fully functional frontal cortex, any tendency to select similar distractors is successfully inhibited (Morrison et al., 2004; Knowlton et al., 2012). But an alternative possibility is that healthy adults are able to reason by analogy without evoking a more superficial strategy based on comparing the similarity of *C* and *D* terms, so that superficial similarity simply does not enter into the analogical decision process.

Employing two different methods, the present study found evidence that college students are in fact influenced by both semantic and visual similarity when solving picture analogies. Using a four-alternative forced choice paradigm, in Experiment 1 we tracked eye movements while college students solved picture analogies in the absence of speed pressure. We found that dwell time (total looking time) was elevated for semantic (and possibly visual) foils during the period prior to selection of a response, even though the presence of similar distractors had little impact on the final choice. This finding suggests that similar distractors tended to draw extra attention, even though they were almost always rejected in favor of the analogical solution.

Experiment 2 examined solutions of the same basic picture analogies after they were recast in a true/false format and administered with instructions that emphasized speed of responding. In this situation, the similar distractors (especially the semantic foil) strongly influenced performance by college students. False analogies containing a semantic distractor as the *D* term were often erroneously judged to be true and took longer to correctly reject than any other condition. False analogies based on visual distractors also yielded higher error rates and higher correct RTs than did false analogies based on unrelated *D* terms.

The much more salient impact of similar distractors in Experiment 2 may be related to two ways in which its design differed from that used in both Experiment 1 and in the previous neuropsychological study by Krawczyk et al. (2008). First, speed pressure may be critical. When pressed to respond quickly, as in Experiment 2, there may not be time for inhibitory processes to effectively suppress a tendency to base decisions on superficial similarity.

Second, the true/false format used in Experiment 2 may also have played a role. In the four-alternative forced choice set-up, all options are simultaneously available for comparison, and a common criterion can be applied on an individual trial to determine the "best" alternative (e.g., Lu, Wu, & Holyoak, 2019; Lu, Liu, Ichien, Yuille, & Holyoak, 2019). In the true/false set-up, by contrast, each option has to be evaluated in isolation, and a criterion must be set on each trial to decide whether the analogy is "good enough" to respond "true". Since feedback was never given in our experiments, participants may have been uncertain about the appropriate criterion (especially since the ratio of true and false analogies was unbalanced). Given that the analogies were best solved on the basis of semantic relations, problems including a semantic lure (i.e., those in which the *C* term is semantically related to the *D* term, but not in the same way that *A* is related to *B*) may have often passed the subjective decision criterion, resulting in errors.

Taken together, the present findings seem to rule out the hypothesis that superficial similarity plays no role in analogical reasoning for healthy adults. Depending on test conditions, semantic and visual lures may have relatively subtle effects (a tendency to attract visual attention) or extremely salient effects (generating either errors or slow correct responses).

It would seem, therefore, that our results favor the standard view that analogical reasoning is susceptible to interference from a non-analogical strategy of simply evaluating the similarity of the *C* and *D* terms, without reference to the *A:B* relation. However, another alternative deserves consideration. The analogy "game" bases the correct answer on the most specific possible relation(s) in common across *A:B* and *C:D* (e.g., for the analogy shown in Figure 1, the specific relation "lives in" links squirrel to tree and also fish to fishbowl). But suppose relations emerge in a gradual fashion during the reasoning process, rather than simply being retrieved in an all-or-none fashion. Then the *A:B* and *C:D* relations may at first be vague or incomplete, and only over time reach full specificity. Early in this process of relation encoding, the active relation between *A:B* may be something very general (e.g., a squirrel is somehow related, either semantically or visually, to a tree). At this point, one or both of the foils may match the crude *A:B* relation about equally well as the analogical answer (e.g., a fish is associated with a fishbowl, and similar visually to the pictured rocket). Under this view, speed pressure may force the reasoner to choose the "best" answer before the relations are fully encoded, at a point in time when the analogical answer and the similar foils may be comparable in their degree of match to the partially-encoded *A:B* relation.

This alternative account of interference implies its source may not be a rival non-analogical strategy (e.g., simply comparing *C* and *D* while ignoring *A:B*). Rather, interference may emanate from the analogy process itself, if a fast decision is required when relations are as yet poorly encoded. Future research should attempt to test these alterative accounts of how superficial similarity can infiltrate a process that aims to focus on relations.

Acknowledgements

We thank Robin Gruber and Matthew Weiden for assistance in running Experiment 1, and Brandon Valenica, Justin Shin and Kiran Cherian for assistance with Experiment 2. A preliminary report of Experiment 1 was presented at the Forty-ninth Annual Meeting of the Psychonomic Society (Chicago, November 2008). Preparation of this paper was supported by NSF Grant BCS-1827374 to KJH.

References

Glady, Y., French, R. M., & Thibaut, J. P. (2016). Comparing competing views of analogy making using eye-tracking technology. In *Proceedings of the 38th Annual Meeting of the Cognitive* Science Society (pp. 1349-1354). Austin, TX: Cognitive Science Society.

Goranson, T. E. (2002). On diagnosing Alzheimer's disease: Assessing abstract thinking and reasoning. *Dissertation Abstracts International: Section B: Sciences and Engineering*, *62*, 4785.

Gordon, P. C., & Moser, S. (2007). Insight into analogies: Evidence from eye movements. *Visual Cognition, 15*(1), 20–35.

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's Advanced Progressive Matrices. *Journal of Vision, 11*, 1–11.

Knowlton, B. J., Morrison, R. G., Hummel, J. E., & Holyoak, K. J. (2012). A neurocomputational system for relational reasoning. *Trends in Cognitive Sciences*, *16*, 373-381.

Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., Miller, B. L., & Knowlton, B. J. (2008). Distraction during relational reasoning: The role of prefrontal cortex in interference control. *Neuropsychologia*, *46*(7), 2020–2032.

Lu, H., Liu, Q., Ichien, N., Yuille, A. L., & Holyoak, K. J. (2019). Seeing the meaning: Vision meets semantics in solving visual analogy problems. In A. Goel, C. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Lu, H., Wu, Y. N., & Holyoak, K. (2019). Emergence of analogy from relation learning. *Proceedings of the National Academy of Sciences, USA*, *116*, 4176-4181.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25,* 431-467.

Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *Journal of Cognitive Neuroscience*, *16*(2), 260–271.

Vendetti, M. S., Starr, A., Johnson, E. L., Modavi, K., & Bunge, S. A. (2017). Eye movements reveal optimal strategies for analogical reasoning. *Frontiers in Psychology*, *8*, 932. doi:10.3389/fpsyg.2017.00932

Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition*, *28*, 1205-1212.