



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/211484/>

Version: Published Version

Article:

Sio, U.N., Lortie-Forgues, H. and Marsh, J.E. (2024) Effects of task characteristics and task-switching on remote associates test problem solving. *Journal of Cognitive Psychology*, 36 (5). pp. 595-616. ISSN: 2044-5911

<https://doi.org/10.1080/20445911.2024.2333580>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Effects of task characteristics and task-switching on remote associates test problem solving

Ut Na Sio, Hugues Lortie-Forgues & John E. Marsh

To cite this article: Ut Na Sio, Hugues Lortie-Forgues & John E. Marsh (05 Apr 2024): Effects of task characteristics and task-switching on remote associates test problem solving, Journal of Cognitive Psychology, DOI: [10.1080/20445911.2024.2333580](https://doi.org/10.1080/20445911.2024.2333580)

To link to this article: <https://doi.org/10.1080/20445911.2024.2333580>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 05 Apr 2024.



[Submit your article to this journal](#)



Article views: 72



[View related articles](#)



[View Crossmark data](#)

Effects of task characteristics and task-switching on remote associates test problem solving

Ut Na Sio^a, Hugues Lortie-Forgues^b and John E. Marsh^{c,d}

^aInstitute of Work Psychology, Sheffield University of Management School, The University of Sheffield, Sheffield, UK; ^bCentre for Mathematical Cognition, Loughborough University, Loughborough, UK; ^cSchool of Psychology and Computer Science, University of Central Lancashire, Preston, UK; ^dEngineering Psychology, Humans and Technology, Department of Business Administration, Technology and Social Sciences, Luleå University of Technology, Luleå, Sweden

ABSTRACT

Creative problem solving is often viewed as a search process. However, little is known about the factors that impact this process. To address this question, we conducted two studies to examine whether task characteristics and task-switching influence performance on Remote Associates Test (RAT) problems – problems commonly used to measure creativity and study the creative search process. Consistent with prior research, we found that RAT problem-solving performance was positively associated with the relatedness between the answer and the problem. The association was strongest when the amount of competition within the initial search space was low. Moreover, this interaction was observed irrespective of the methods used to measure the task characteristics. By contrast, we did not replicate the positive effect of task-switching on RAT problem-solving accuracy found in previous studies. However, our findings suggest that task-switching may improve problem-solving speed and facilitate a broader search. Implications for future research are discussed.

ARTICLE HISTORY



Received 5 October 2023
Accepted 15 March 2024


KEYWORDS

Remote associates test; creative problem solving; task characteristics; task switching

Creative problem solving is often viewed as a semantic search process that involves retrieving and combining ideas to form creative solutions (e.g. Beaty & Silvia, 2012; Kenett & Faust, 2019; Mednick, 1962). To better understand this specific search process, one approach is to examine how task characteristics – particularly characteristics which theoretically influence the semantic search process in general – are associated with creative problem-solving performance. By examining the relationships between task characteristics and creative problem-solving performance, insights into the key difficulties associated with the creative search process can be gained. Such insights would help determine the search process critical to successful solution discovery and potentially informs the design of more targeted interventions and manipulations to enhance creative problem solving.

The Remote Associates Test (RAT) problems, developed by Mednick (1962), are widely used to understand the relationships between task characteristics and creative problem solving (e.g. Gupta et al., 2012; Olteţeanu & Schultheis, 2019; Wiley, 1998). RAT problems require individuals to find a word that is associated with three cues (e.g. cues: *stick, maker, point*; answer: *match*). Solving RAT problems involves searching within a problem space for words associated with the cues to retrieve the correct answer. This search process can be viewed as a two-stage process involving: (1) generating a potential answer, and (2) evaluating the correctness of that potential answer. If the potential answer is incorrect, individuals will generate and evaluate another potential answer (Smith et al., 2013; Wu et al., 2020). When solving RAT problems, individuals typically generate guesses (i.e. potential

CONTACT Ut Na Sio  u.n.sio@sheffield.ac.uk  Institute of Work Psychology, Sheffield University of Management School, The University of Sheffield, Conduit Road, S10 1FL, Sheffield, UK

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/20445911.2024.2333580>.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

answers) based on one of the three cues at a time, with subsequent guesses partly based on previous ones (Smith et al., 2013). Given that each cue is associated with multiple words, and these words are in turn linked to others, it is generally assumed that an array of potential guess sequences, i.e. search paths, exists. It is important to note that not all of these search paths will lead to the correct answer. Being able to explore a path efficiently and switch when it seems fruitless are critical skills for successful solution discovery and are considered important for creative thinking (Nijstad et al., 2010). Examining RAT problem solving (e.g. identifying factors that influence RAT problem-solving performance) will facilitate understanding of these creativity-related semantic search processes.

Furthermore, RAT problem-solving performance is often used to measure the impact of interventions and manipulations designed to enhance creativity, such as mindfulness, sleep, and listening to music (see Wu et al., 2020 for a review). Understanding the search process that underlies RAT problem solving could also help clarify the roles of these enhancement methods, allowing for their more effective.

Task characteristics and RAT problem solving

To better understand the characteristics of the search process underlying RAT problem solving, previous studies have examined how various task characteristics influence RAT problem difficulty (e.g. Marko et al., 2019; Oltețeanu & Schultheis, 2019; Valba et al., 2021). A consistent finding from previous studies is that RAT problems in which answers are only remotely related to the cues are generally more challenging to solve. This relationship was observed across various measures of cue-answer relatedness. For example, some studies, e.g. Marko et al. (2019) and Valba et al. (2021), measured relatedness based on associative strength between the cue and answer words – i.e. the probability that the answer word would be generated following the presentation of the cue word in a word association task. Others (e.g. Oltețeanu & Schultheis, 2019) measured relatedness based on the co-occurrence frequency of the cue and answer words in text. Although different estimation methods were used, these studies reported similar findings: RAT problems with low cue-answer relatedness were more difficult to solve, suggesting that the search process underlying RAT problem solving shares commonalities with the semantic retrieval process guided by spreading activation (Collins & Loftus, 1975). When solving a RAT problem, the cues will activate the strongly associated words, from which the activation spreads to other related words. If the answer to the RAT problem

is closely related to the cues (e.g. cues: *cream, skate, water*; answer: *ice*), the answer is more likely to be retrieved quickly. However, if the answer and the cues are only remotely related (e.g. cues: *over, plant, horse*; answer: *power*), individuals will need to broaden the search, which could be done by deliberately generating a longer associative chain or through spreading activation, thus increasing the difficulty of finding the answer.

Apart from cue-answer relatedness, the amount of competition (e.g. the number of strong but irrelevant associates) in the initial problem space is suggested to be another critical factor that impacts performance. Strong but irrelevant associates may trigger many paths that lead to dead ends instead of the answer, reducing the chance of identify the right ones. Past studies have shown that RAT problems with cues associated with strong but irrelevant associates (i.e. competitors) tend to be more difficult (e.g. Schatz et al., 2022; Sio et al., 2022; Vul & Pashler, 2007; Wiley, 1998). There are different ways to examine this effect on RAT problem solving. A common approach involves presenting irrelevant associates alongside the cues and examining whether their presence impacts RAT problem-solving performance. For example, Vul and Pashler (2007) found that RAT problem-solving performance was negatively impacted when irrelevant associates accompanied presentation of the cues (e.g. cues: *tank, hill, secret*; answer: *top*; strong but irrelevant associates: *water, ant, hideout*). Another method is to create or select RAT problems with cues associated with strong but irrelevant associates. For example, Wiley (1998) designed RAT problems in which one of the three cues was related to a baseball term that was not the correct answer (e.g. cues: *stolen, tax, private*; answer: *property*; associated baseball term: *stolen base*). They found that individuals familiar with baseball terms performed worse on these problems compared to those less familiar. Similarly, Sio et al. (2022) found that RAT problems with cues having a large number of strong but irrelevant associates (e.g. cues: *fight, control, machine*; answer: *gun*) were more difficult to solve than those with cues associated with a smaller number of words (e.g. cues: *guy, rain, down*; answer: *fall*). Besides experimental studies, computational studies were also conducted to examine the impact of strong but irrelevant associates on RAT problem solving. A computational study by Schatz et al. (2022) showed that strengthening the connection between the cues and their irrelevant associates in the network led to lower performance in solving RAT problems.

In line with the idea that the presence of strong but irrelevant associates can hinder RAT problem solving,

Storm and Angello (2010) reported that individuals' efficiency in inhibiting competitors was positively associated with their performance on RAT problems. Similarly, in a computational study, Valba et al. (2021) found that suppressing competitors could improve the performance of the computational RAT problem-solving model.

In summary, past studies have identified task characteristics that are associated with RAT problem-solving performance. However, most of these studies have focused on examining the individual influence of these factors, omitting their potential interdependency. Such an approach may not fully capture the RAT problem-solving process if different task characteristics interact when influencing problem difficulty. For example, Becker et al. (2022) and Sio et al. (2022) investigated the interaction between cue-answer relatedness and the amount of competition in the initial problem space when predicting RAT problem-solving performance. Both studies reported an interaction, but interestingly, in opposite directions. Becker et al. (2022) found that cue-answer relatedness was positively associated with RAT problem-solving performance, and this association was strongest when the amount of competition was *high*. Conversely, Sio et al. (2022) found that the positive relationship between cue-answer relatedness and performance was most pronounced when the amount of competition was *low*. Notably, these studies drew different conclusions regarding the relative difficulty of RAT problems. For example, the results of Becker et al. (2022) indicated that RAT problems with high cue-answer relatedness and a high amount of competition were *easiest* to solve. By contrast, Sio et al. (2022) found that RAT problems high on these characteristics were the *hardest* to solve.

Instead of clarifying, these conflicting results further complicate our understanding of the search process involved in RAT problem solving. One potential factor contributing to the differences in results between these two studies is related to the methods they used to measure task characteristics. Specifically, Becker et al. (2022) measured the task characteristics using semantic similarity between words, which was determined by word co-occurrence frequency across a large text corpus. By contrast, Sio et al. (2022) measured the same task characteristics using associative strength between words, which reflects how likely the presence of one word will elicit the other word. Past studies (e.g. Schatz et al., 2022) have suggested that relatedness measures derived from word co-occurrences in text (e.g. semantic similarity) and those built from word associations collected from humans (e.g. associative strength) capture different aspects of word relatedness. For example, although seeing the word *banana* is likely to activate the word *yellow* (i.e. high associative

strength), these two words may not frequently co-occur in text (i.e. low co-occurrence; Schatz et al., 2022). Another key difference between association strength and semantic similarity is that associative strength focuses on the direct stimulus-response relationship between the two words, while semantic similarity takes into account both the co-occurrence of the two words as well as the overall structure of the corpus (De Deyne et al., 2019; Günther et al., 2015). As expected, past studies (e.g. Maki et al., 2004) have found a weak correlation between semantic similarity and associative strength. The difference between these two relatedness measures could be a contributing factor to the discrepancy between the findings of Becker et al. (2022) and Sio et al. (2022).

To further investigate the differences in findings between Becker et al. (2022) and Sio et al. (2022), the current study examined the interaction between cue-answer relatedness and the amount of competition when predicting RAT problem-solving performance, and used both associative strength and semantic similarity to measure these two characteristics.

Task-switching and RAT problem solving

Another aim of the present work was to investigate the effect of task-switching (i.e. alternating between tasks over a short period of time) on RAT problem-solving. We also examined whether the aforementioned task characteristics moderated the task-switching effect.

As discussed before, RAT problem solving involved exploring multiple search paths, with only some of these paths leading to the correct answer. As such, being able to switch paths when the current one proves unfruitful would be critical for successful RAT problem solving. In line with this hypothesis, Davelaar (2015) reported the frequency of individuals shifting their focus between cues (i.e. transitioning to another path based on a different cue) correlated positively with RAT problem-solving performance. Accordingly, task-switching may facilitate RAT problem solving, as interrupting the search process could give individuals an opportunity to start fresh again and explore other paths, facilitating a broader search.

Task-switching may also expand the search scope by taxing inhibition resources. Chrysikou (2019) suggested that task-switching draws on inhibitory resources. When these resources are depleted, it may lead to a less controlled and wider spread of activation, facilitating the retrieval of more remote ideas, and in turn, enhancing RAT problem solving.

In sum, the two proposed explanations both suggest that task-switching can impact the semantic search

process – either by facilitating a shift in the search direction or through widening the spread of activation. These mechanisms may enhance performance on RAT problems and other similar creative tasks that also rely heavily on semantic search. Recent studies have provided evidence supporting the positive role of task-switching on creative problem solving. For instance, previous studies (e.g. George & Wiley, 2019; Lu et al., 2017; Smith et al., 2017) observed a positive effect of task-switching on divergent thinking performance. Furthermore, Smith et al. (2017) found that task-switching facilitated idea generation particularly when generating ideas for flexibly defined topics. Similarly, Lu et al. (2017) and Sio et al. (2017) reported positive effects of task-switching on RAT problem solving and insight problem solving. Sio et al. (2017) found that task-switching was particularly beneficial for solving RAT problems that contain information that could potentially mislead individuals to search in the wrong direction.

While there is evidence supporting the benefit of task-switching on creative problem solving, it is important to note that creative problem-solving tasks are complex and that their difficulty often depends on multiple factors. As previously discussed, research suggests that the difficulty of RAT problems depends on both the remoteness of the answer and the amount of competition in the initial search space. Therefore, to fully grasp the impact of task-switching on creative problem-solving and determine under which conditions the manipulation becomes beneficial, it is essential to examine whether the effect of task-switching is moderated by task characteristics. Although prior studies have explored the relationship between task characteristics and the effect of task-switching, they have typically focused on a single characteristic. Given the complexity of creative problem solving, it is crucial to consider not just one, but multiple, potentially interacting, task characteristics when examining the effect of task-switching – another focus of the present work.

We conducted two studies to address the issues presented above. In both studies, we examined the relationship between RAT problem-solving performance and two task characteristics: (1) cue-answer relatedness and (2) the amount of competition within the initial problem space. We also investigated the impact of task-switching on RAT problem solving and examined whether the impact was moderated by the task characteristics. We broadened our focus beyond performance metrics, exploring whether task-switching could impact the search process. In both studies, we not only measured RAT problem-solving accuracy and solution time but also the relatedness of responses generated during RAT problem solving – an indicator of the

scope of the search undertaken (e.g. Davelaar, 2015; Smith et al., 2013). Previous studies have often used longer inter-response time as another indicator of a broader search scope (e.g. Hass, 2017). However, we did not use this metric in our studies. We believe that task-switching (i.e. presenting problems in multiple segmented short sessions) can create a sense of time pressure. This may lead individuals to respond more quickly, potentially confounding the interpretation of inter-response times.

Study 1

In Study 1, we examined the impact of two task characteristics (cue-answer relatedness and the amount of competition) and task-switching on RAT problem-solving performance. To better understand the effect of task-switching on RAT problem-solving, we also examined whether task-switching impacted the scope of the search conducted during RAT problem solving. All data and syntax files are available through the Open Science Framework: <https://osf.io/uhxq9/>.

Methods

Participants

One hundred English speakers (47 females, 53 males) residing in North America were recruited via Amazon Mechanical Turk (MTurk), with a mean age of 34.17 years ($SD = 12.72$). Only individuals with English as their first language and with an MTurk approval rate of at least 99% were recruited, and 57% of the participants self-reported having a college degree or above. They were paid USD \$2 for completing the study. Participants were asked to read the information sheet and give informed consent at the beginning of the study.

Task

Twenty-four RAT problems were selected from the normative set produced by Bowden and Jung-Beeman (2003) and previous RAT problem-solving studies (e.g. Thompson, 1993). For each RAT problem, two characteristics – (1) cue-answer relatedness and (2) the amount of competition – were measured using associative strength and semantic similarity.

Cue-answer relatedness and the amount of competition measured using associative strength

Cue-answer associative strength. To measure cue-answer relatedness, we calculated the sum of the associative strength from each of the three cues to the answer. Associative strength, as mentioned earlier, reflects the probability that the answer will be generated following

the presentation of the cue, with values ranging from 0 to 1. The task characteristic *cue-answer associative strength*, being the sum of the associative strength from each of the three cues to the answer, ranged from 0 to 3.

The method of estimating cue-answer relatedness based on associative strength has been used in many studies (e.g. Marko et al., 2019). In this study, we used the “Small World of Words” English word association norms (De Deyne et al., 2019) to determine the associative strength between the cue and the answer.

Number of competing associates. For each of the three cues, we counted the number of words that had a stronger association with the cue than the association between the cue and the answer. This value was bounded below by zero and had no upper limit. We summed these numbers and used the total as a measure of the amount of competition within the initial problem space. This method has been used in Sio et al. (2022).

Cue-answer relatedness and the amount of competition measured using semantic similarity

Cue-answer semantic similarity. To measure cue-answer relatedness, we used SemDis (Beaty & Johnson, 2021) to estimate the semantic distance between the answer and the three cues. SemDis calculated a semantic distance score (ranging from 0 to 2) for each cue-answer word pair, based on their co-occurrence frequencies across five semantic spaces: cbowukwacsubtitle, cbowsubtitle, cbowBNCwikiukwac, TASA, and glove. We converted each distance score into a semantic similarity score by subtracting it from 2. Then, we computed the average cue-answer semantic similarity score, which ranged from 0 to 2, with higher values indicating higher similarity, and used this value as our measure of cue-answer relatedness.

Cue-cue semantic similarity. We followed a similar procedure to compute the average semantic similarity (ranging from 0 to 2) score between the three cues and used it as a measure of the amount of competition. A high semantic similarity score suggests that the cues share many close associates. These close associates are

likely to be co-activated by the cues, potentially interfering with the retrieval of the answer (Becker et al., 2022; Davelaar, 2015). Thus, the semantic similarity between cues could be viewed as an indicator of the level of competition in the initial search space and was used as a measure of the amount of competition.

The 24 RAT problems were divided into two sets of 12 RAT problems, matched in terms of these characteristics (see Table 1). Supplemental material Table S1 presents the following values for each of the RAT problems: cue-answer association strength, number of competing associates, cue-answer semantic similarity, and cue-cue semantic similarity. Similar to the findings of previous studies (e.g. Maki et al., 2004), the correlations between the task characteristics measured by associative strength and semantic similarity were small to moderate (cue-answer association strength and cue-answer semantic similarity: $r = .08$; number of competing associates and cue-cue semantic similarity: $r = .37$). This is not unexpected because, as discussed earlier, associative strength and semantic similarity represent different aspects of semantic relatedness (e.g. Schatz et al., 2022).

Procedures

This study used a within-subject design in which participants solved two sets of RAT problems (12 problems in each set): one under a no-switching condition and another under a switching condition. The order of presentation for conditions (switching vs. no-switching) and problem sets was counterbalanced. RAT problems within each problem set were randomised. The experiment was delivered online via Gorilla.sc (Anwyl-Irvine et al., 2020).

In the no-switching condition, each RAT problem was presented once for a maximum 30 s. In the switching condition, the RAT problems were presented repeatedly in three blocks, with each problem presented once for a maximum of 10 s in each block and intermixed randomly with other problems. Correctly solved problems were removed in the next block. The maximum presentation time for each RAT problem was 30 s in both conditions. The entire

Table 1. Mean, standard deviation, and minimum and maximum values of each characteristic in study 1.

| Task characteristics for RAT problems in Study 1 | Set 1 | | | Set 2 | | |
|--|------------------------|------|------|------------------------|------|------|
| | <i>M</i> (<i>SD</i>) | Min | Max | <i>M</i> (<i>SD</i>) | Min | Max |
| Based on associative strength | | | | | | |
| Cue-answer associative strength | 0.05 (0.05) | 0.00 | 0.13 | 0.03 (0.03) | 0.00 | 0.10 |
| Number of competing associates | 93.25 (25.28) | 62 | 127 | 98.67 (33.72) | 24 | 146 |
| Based on semantic similarity | | | | | | |
| Cue-answer semantic similarity | 1.17 (0.04) | 1.11 | 1.22 | 1.17 (0.05) | 1.09 | 1.28 |
| Cue-cue semantic similarity | 1.10 (0.03) | 1.05 | 1.15 | 1.11 (0.04) | 1.06 | 1.17 |

study, which involved solving 24 RAT problems, took a maximum of 12 min to complete.

During each RAT problem presentation, participants could enter their responses at any point. To do this, they first had to key in the response and then press the ENTER key. If they provided the correct answer, they would advance to the next problem; otherwise, they could continue working on the problem within the remaining time limit. Participants were told that they could make as many attempts as they like. All responses were recorded and time-logged.

Before the main experiment, participants were given instructions and three practice problems. They had 30 s to solve each practice problem if the no-switching condition was presented first in the main experiment, or 10 s to solve each practice problem if the switching condition was presented first. In the main experiment, after participants completed one condition, they were informed that they would solve another set of RAT problems under a different condition. They also receive instructions related to the new time limit for the RAT problems before proceeding to the remaining condition.

Results

On average, participants solved 37.25% ($SD = 20.10\%$) of the RAT problems. Participants on average took 11.02 s ($SD = 4.11$ s) to solve a RAT problem. Table 2 presents the means and standard deviations of RAT problem-solving accuracy and solution time for each condition. See supplemental material Table S1 for the solution rate for each RAT problem in our sample. The solution rates were moderately and positively correlated with the normative data reported by Bowden and Jung-Beeman (2003), $r = .66$, $p = .003$.

Model construction and analysis

Task characteristics, task-switching and rat problem-solving performance. One focus of this study was to examine the impact of task characteristics and task-switching on RAT problem-solving performance, measured in terms of accuracy and solution time. For accuracy, we constructed generalised linear mixed-

effects models with a logit link function using whether or not the RAT problem was solved as the outcome variable. The analysis included all trials ($N = 2400$ trials: 100 participants, each solving 24 RAT problems). This sample size gives a power of over 80% of detecting an effect with an effect size of $d = 0.2$ or above. This power analysis was conducted using the R-package “simr” (Green & MacLeod, 2016) based on information from our pilot data.

For solution time, we only included solved trials in the analysis ($N = 894$ trials). We constructed linear mixed-effects models with log-transformed (base 2) solution time as the outcome variable. Solution time data were log-transformed to respect the normality assumption of residuals. Supplemental material Figure S1 presents the normal Q-Q plots of residuals, with and without transforming the outcome variable. For all the residual plots presented in this paper, we conducted only a visual inspection and did not perform any formal test for normality because these tests are likely to be very sensitive to small deviations given our sample size. Our decision aligns with the recommendations of Zuur et al. (2010).

For each performance measure, two sets of models were constructed to evaluate the relationships between task characteristics, task-switching, and RAT problem-solving performance. One set included task characteristics measured using associative strength as the predictor variables while the other included task characteristics measured using semantic similarity.

For all model sets, the predictor variables were entered in the following sequence: We started with a basic model including only random intercepts of participants (i.e. Null model). Next, we included the two task characteristic measures as fixed effects (Model 1), followed by their interaction term (Model 2). We then included Condition (reference group: no-switching) as another fixed effect (Model 3). Lastly, we included interaction terms between the task characteristics and Condition (Model 4). Given that the measures of the task characteristics were on different scales, they were first standardised by subtracting the mean and then dividing the results by the standard deviation. We compared the AIC values of these nested models, selecting the model with the lowest AIC as the best-fitting one.

Task switching and RAT problem-solving search process. We also explored whether task-switching impacted the search process during RAT problem solving. Specifically, we compared the semantic similarity between adjacent responses. High semantic similarity between adjacent responses suggests a narrow

Table 2. Means and Standard Deviations of RAT Problem-Solving Accuracy and Solution Time for Each Condition in Study 1.

| Condition | Accuracy | | Solution Time ^a (in seconds) | |
|--------------|----------|-----------|---|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| No-Switching | 0.40 | 0.27 | 10.90 | 6.96 |
| Switching | 0.35 | 0.20 | 10.64 | 7.53 |

a. Only includes solved trials.

search, where individuals generate new responses largely based on their previous responses. Conversely, low semantic similarity between adjacent responses is expected if individuals expand their search to retrieve more remotely related ideas or switch to a different search direction.

In this analysis, we chose not to use associative strength as a measure of relatedness between adjacent responses. This decision was based on the fact that a majority of adjacent responses (92%) are not listed as word pairs in the “Small World of Words” English word association norms (De Deyne et al., 2019), probably because they are weakly associated. We could only assign them zero associative strength. Doing so resulted in a lack of variability in the data, which in turn would greatly reduce the statistical power to detect an effect.

Prior to the calculation of the semantic similarity between adjacent responses, all responses were spell-checked manually. If participants repeated the same word twice in a row, we treated them as if the participants had entered that word only once and removed the repeated word from the analyses. We then used SemDis (Beaty & Johnson, 2021) to compute the semantic distance between adjacent responses and then convert this to a similarity score by subtracting it from 2.

We constructed a linear mixed-effects model with the semantic similarity score for adjacent responses as the outcome variable. We transformed the semantic similarity scores using a Box–Cox transformation (lambda value: -3) to reduce skewness to avoid violating the assumption of normality in residuals. Supplemental material Figure S2 presents the normal Q-Q plots of residuals, with and without transforming the outcome variable. In the model, we included both tasks and participants as random factors to account for the nested structure of the data – multiple responses collected per task and multiple tasks per participant. We then included Condition (reference group: no-switching) as a fixed effect. We also included trial type (unsolved vs. solved, reference group: unsolved), the time interval between adjacent responses (measured in seconds), and the number of attempts (e.g. 1 for the 1st attempt, 2 for the 2nd attempt, and so on) as fixed effects to statistically control for their potential influence on the semantic similarity between adjacent responses.

For all the statistical tests, we used a significance level of $\alpha = 0.05$. We used the R-package “lme4” (Bates et al., 2015) to construct all mixed-effects models and estimate the 95% CIs for all the coefficients. The p values for the coefficients were computed using the R-package “lmeTest” (Kuznetsova et al., 2017), and interaction

analyses were performed using R-package “interactions” (Long, 2019).

Model results: task characteristics and RAT problem solving

In this section, we focused on the results related to effects of the two task characteristics on RAT problem-solving. For the results concerning the effect of task-switching on RAT problem-solving, see the next section *Model Results: Task-Switching and RAT Problem-Solving*.

RAT problem-solving accuracy. For both model sets, the best-fitting model (i.e. lowest AIC value) included the main effects of the two characteristics and their interaction term, and the main effect of Condition (see Model 3, Tables 3 and 4). Adding the interaction terms between Condition and the two task characteristics did not further improve the model fit (see Model 4, Tables 3 and 4). Tables 5 and 6 present the details of the best-fitting models.

Task Characteristics Measured Using Associative Strength: Our results showed that there was a significant interaction between the cue-answer associative strength and the number of competing associates on RAT problem-solving accuracy ($B = -0.18$, $p = .001$, see Table 5). To explore the interaction, we computed the coefficient of the cue-answer associative strength while holding the number of competing associates at 1 SD below the mean (low), at the mean (medium), and at 1 SD above the mean (high). When the number of competing associates was low, the cue-answer associative strength was positively associated with RAT problem-solving accuracy ($B_{low} = 0.44$, $p < .001$, see Figure 1). However, this association became weaker and eventually statistically non-significant as the number of competing associates increased ($B_{high} = 0.09$, $p = .28$, see Figure 1).

As shown in Figure 1, RAT problems with low cue-answer associative strength (i.e. low cue-answer relatedness) were in general more difficult. Notably, RAT problems with high cue-answer associative strength were also difficult to solve when there was a large number of competing associates (i.e. high amount of competition).

Task Characteristics Measured Using Semantic Similarity: We observed a similar interaction between task characteristics and RAT problem-solving accuracy when these characteristics were measured using semantic similarity ($B = -0.24$, $p = .008$, see Table 6). As a follow-up analysis, we examined the association between cue-answer semantic similarity and accuracy while holding cue-cue semantic similarity at the same cut-off values used previously – i.e. 1 SD below the

Table 3. Summary of the Models on RAT Problem-Solving Accuracy in Study 1 (Task Characteristics Measured Using Associative Strength; Best-Fitting Model in Bold).

| Predictor Variables | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|------------------------|--|-------------|----------|---|-------------|----------|---|--------------------|--------------------|--|-------------|----------|
| | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> |
| ASSOC | 0.29 | 0.06 | <.001*** | 0.26 | 0.06 | <.001*** | 0.26 | 0.06 | <.001*** | 0.26 | 0.09 | .003** |
| COMP | -0.04 | 0.06 | .459 | -0.02 | 0.06 | .797 | -0.01 | 0.06 | .811 | -0.06 | 0.08 | .452 |
| ASSOC*COMP | - | - | - | -0.18 | 0.05 | .001*** | -0.18 | 0.05 | .001*** | -0.07 | 0.07 | .332 |
| Condition | - | - | - | - | - | - | -0.24 | 0.09 | .010** | -0.37 | 0.11 | .001*** |
| ASSOC*Condition | - | - | - | - | - | - | - | - | - | 0.01 | 0.12 | .908 |
| COMP*Condition | - | - | - | - | - | - | - | - | - | 0.10 | 0.12 | .403 |
| ASSOC*COMP*Condition | - | - | - | - | - | - | - | - | - | -0.22 | 0.10 | .038* |
| Model comparison | $\chi(2)^2 = 48.096, p < .001^{***}$ (vs. Null model) | | | $\chi(1)^2 = 11.921, p = .001^{***}$ (vs. Model 1) | | | $\chi(1)^2 = 6.567, p = .010^{**}$ (vs. Model 2) | | | $\chi(2)^2 = 4.668, p = .198$ (vs. Model 3) | | |
| Model ICC ^a | 0.210 | | | 0.212 | | | 0.214 | | | 0.214 | | |
| Model fit: AIC | 2936.8 | | | 2926.9 | | | 2922.3 | | | 2923.7 | | |

Note: 2400 observations, 100 participants, 24 RAT problems per participant; ASSOC: Cue-answer associative strength; COMP: Number of competing associates; For Condition, no-switching was the reference group; * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$; Null model (random effect only): AIC = 2980.9, ICC = 0.202.
a. ICC (Intraclass Correlation) indicates the amount of variance in outcomes that is explained by the random effect.

Table 4. Summary of the Models on RAT Problem-Solving Accuracy in Study 1 (Task Characteristics Measured Using Semantic Similarity; Best-Fitting Model in Bold).

| Predictor Variables | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|--|--|-------------|----------|---|-------------|----------|--|--------------------|-----------------|--|-------------|----------|
| | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> |
| Sim _{cues_answer} | -.20 | 0.05 | <.001 | -0.07 | 0.06 | .240 | -0.06 | 0.06 | .263 | -0.08 | 0.08 | .335 |
| Sim _{cues} | -0.05 | 0.05 | .241 | -0.01 | 0.04 | .774 | -0.01 | 0.05 | .781 | -0.06 | 0.07 | .382 |
| Sim _{cues_answer} *Sim _{cues} | - | - | - | -0.25 | 0.06 | <.001 | -0.26 | 0.06 | <.001 | -0.24 | 0.08 | .003 |
| Condition | - | - | - | - | - | - | -0.24 | 0.09 | .008 | -0.23 | 0.09 | .013 |
| Sim _{cues_answer} *Condition | - | - | - | - | - | - | - | - | - | 0.03 | 0.12 | .788 |
| Sim _{cues} *Condition | - | - | - | - | - | - | - | - | - | 0.09 | 0.10 | .333 |
| Sim _{cues_answer} *Sim _{cues} *Condition | - | - | - | - | - | - | - | - | - | -0.03 | 0.12 | .802 |
| Model comparison | $\chi(2)^2 = 22.875, p < .001$ (vs. Null model) | | | $\chi(1)^2 = 21.596, p < .001$ (vs. Model 1) | | | $\chi(1)^2 = 6.962, p = .008$ (vs. Model 2) | | | $\chi(2)^2 = 1.013, p = .798$ (vs. Model 3) | | |
| Model ICC ^a | 0.206 | | | 0.209 | | | 0.211 | | | 0.211 | | |
| Model fit: AIC | 2962.1 | | | 2942.5 | | | 2937.5 | | | 2942.5 | | |

Note: 2400 observations, 100 participants, 24 RAT problems per participant; Sim_{cues_answer}: Cue-answer semantic similarity; Sim_{cues}: Cue-cue semantic similarity; For Condition, no-switching was the reference group; * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$; Null model (random effect only): AIC = 2980.9, ICC = 0.202.
a. ICC (Intraclass Correlation) indicates the amount of variance in outcomes that is explained by the random effect.

mean (low), the mean (medium), and 1 *SD* above the mean (high). When cue-cue semantic similarity was low (i.e. low amount of competition), the RAT problem-solving accuracy was positively associated with cue-answer semantic similarity ($B_{low} = 0.20, p = .04$, see Figure 2). However, the association reversed and became negative when cue-cue semantic similarity was high ($B_{high} = -0.32, p < .001$, see Figure 2). As shown in Figure 2, RAT problems with low cue-answer semantic similarity (i.e. low cue-answer relatedness) were difficult to solve when

cue-cue semantic similarity was low (i.e. low amount of competition). At the same time, RAT problems with high cue-answer semantic similarity (i.e. high cue-answer relatedness) were also difficult when cue-cue semantic similarity was high (i.e. high amount of competition).

RAT problem-solving solution time. Task Characteristics Measured Using Associative Strength. When predicting solution time, the null model was the best-fitting model, indicated by the lowest AIC value. This implies

Table 5. Details of the Best-Fitting Model on RAT Problem-Solving Accuracy in Study 1 (Task Characteristics Measured Using Associative Strength).

| Predictor Variables | Best-Fitting Model (Model 3, Table 3) | | | | |
|-----------------------------------|---------------------------------------|-------------|----------|----------|---------------------|
| | <i>B</i> (95% CI) | <i>SE B</i> | <i>z</i> | <i>p</i> | Odds Ratio (95% CI) |
| ASSOC | 0.26 (0.15, 0.38) | 0.06 | 4.547 | <.001 | 1.27 (1.16, 1.46) |
| COMP | -0.01 (-0.13, 0.10) | 0.06 | -0.239 | .811 | 0.99 (0.88, 1.05) |
| ASSOC*COMP | -0.18 (-0.28, -0.07) | 0.05 | -3.411 | .001 | 0.84 (0.76, 0.93) |
| Condition | -0.24 (-0.42, -0.06) | 0.09 | -2.570 | .010 | 0.77 (0.66, 0.94) |
| Random Effect Variance | | | | | |
| Participant | 0.895 | | | | |
| Marginal R^2 /Conditional R^2 | 0.032/0.239 | | | | |

Note: 2400 observations, 100 participants, 24 RAT problems per participant; ASSOC: Cue-answer associative strength; COMP: Number of competing associates; For Condition, no-switching was the reference group; The estimated logit coefficients were transformed to odds ratios for easier interpretation.

Table 6. Details of the Best-Fitting Model on RAT Problem-Solving Accuracy in Study 1 (Task Characteristics Measured Using Semantic Similarity).

| Predictor Variables | Best-Fitting Model (Model 3, Table 3) | | | | |
|---|---------------------------------------|-------------|----------|----------|-----------------------------|
| | <i>B</i> (95% <i>CI</i>) | <i>SE B</i> | <i>z</i> | <i>p</i> | Odds Ratio (95% <i>CI</i>) |
| Sim _{cues_answer} | -0.06 (-0.17, 0.05) | 0.06 | -1.119 | .263 | 0.94 (0.84, 1.05) |
| Sim _{cues} | -0.01 (-0.11, 0.08) | 0.05 | -0.279 | .781 | 0.99 (0.90, 1.08) |
| Sim _{cues_answer} *Sim _{cues} | -0.26 (-0.37, -0.15) | 0.06 | -2.646 | <.001 | 0.77 (0.69, 0.86) |
| Condition | -0.24 (-0.42, -0.06) | 0.09 | -4.710 | .008 | 0.77 (0.66, 0.94) |
| Random Effect Variance | | | | | |
| Participant | 0.879 | | | | |
| Marginal <i>R</i> ² /Conditional <i>R</i> ² | 0.027/0.232 | | | | |

Note: 2400 observations, 100 participants, 24 RAT problems per participant; Sim_{cues_answer}: Cue-answer semantic similarity between cues and the answer; Sim_{cues}: Cue-cue semantic similarity; For Condition, no-switching was the reference group; The estimated logit coefficients were transformed to odds ratios for easier interpretation.

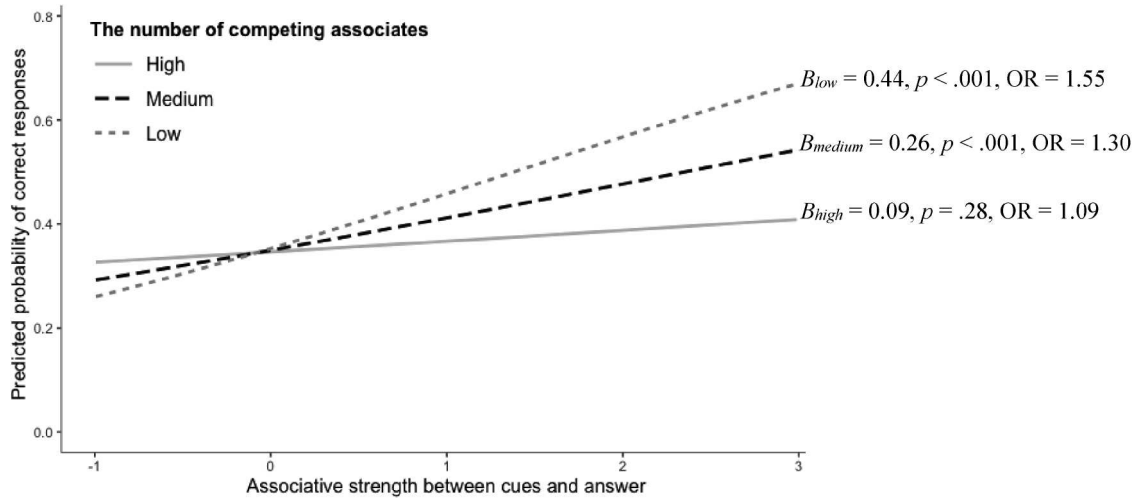


Figure 1. Regression Coefficient of the Cue-Answer Associative Strength When the Number of Competing Associates was Low, Medium, and High (Study 1).

Note. OR = Odds Ratio. The estimated logit coefficients were transformed to odds ratios for easier interpretation.

that none of the predictor variables (i.e. Condition and the two task characteristic measures) predicted RAT problem-solving solution time. See Table 7 for model comparisons.

Task Characteristics Measured Using Semantic Similarity. When task characteristics were measured using semantic similarity, our results showed that the best-fitting model (i.e. lowest AIC value) for predicting RAT

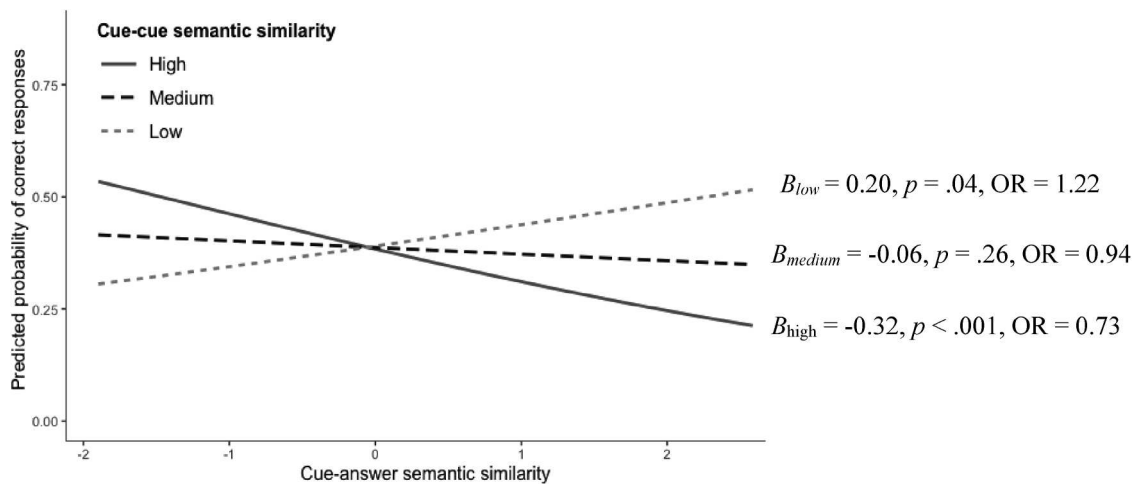


Figure 2. Regression Coefficient of Cue-Answer Semantic Similarity When Cue-Cue Semantic Similarity was Low, Medium and High (Study 1).

Note: OR = Odds Ratio. The estimated logit coefficients were transformed to odds ratios for easier interpretation.

Table 7. Summary of the Models on RAT Problem-Solving Solution Time in Study 1 (Task Characteristics Measured Using Associative Strength).

| Predictor Variables | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|------------------------|---|-------------|----------|--|-------------|----------|--|-------------|----------|--|-------------|----------|
| | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> |
| ASSOC | -0.01 | 0.04 | .810 | -0.01 | 0.04 | .828 | -0.01 | 0.04 | .833 | -0.01 | 0.05 | .861 |
| COMP | 0.01 | 0.04 | .742 | -0.01 | 0.04 | .728 | -0.01 | 0.04 | .738 | -0.06 | 0.05 | .276 |
| ASSOC*COMP | - | - | - | <0.01 | 0.03 | .904 | <0.01 | 0.03 | .921 | <-0.01 | 0.04 | .956 |
| Condition | - | - | - | - | - | - | -0.03 | 0.06 | .641 | -0.01 | 0.08 | .882 |
| ASSOC*Condition | - | - | - | - | - | - | - | - | - | <0.01 | 0.07 | .950 |
| COMP*Condition | - | - | - | - | - | - | - | - | - | 0.09 | 0.08 | .226 |
| ASSOC*COMP*Condition | - | - | - | - | - | - | - | - | - | 0.01 | 0.06 | .895 |
| Model comparison | $\chi(2)^2 = 0.111, p = .946$ (vs. Null model) | | | $\chi(1)^2 = 0.014, p = .903$ (vs. Model 1) | | | $\chi(1)^2 = 0.221, p = .638$ (vs. Model 2) | | | $\chi(2)^2 = 2.599, p = .458$ (vs. Model 3) | | |
| Model ICC ^a | 0.165 | | | 0.164 | | | 0.163 | | | 0.166 | | |
| Model fit: AIC | 2475.8 | | | 2477.8 | | | 2479.5 | | | 2482.9 | | |

Note: 894 observations, 96 participants, 24 RAT problems per participant; ASSOC: Cue-answer associative strength; COMP: Number of competing associates; For Condition, no-switching was the reference group; Null model (random effect only): AIC = 2471.9, ICC = 0.165
a. ICC (Intraclass Correlation) indicates the amount of variance in outcomes that is explained by the random effect.

Table 8. Summary of the Models on RAT Problem-Solving Solution Time in Study 1 (Task Characteristics Measured Using Semantic Similarity; Best-Fitting Model in Bold).

| Predictor variables | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|--|--|-------------|----------|--|--------------------|-----------------|--|-------------|----------|--|-------------|----------|
| | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> |
| Sim _{cues_answer} | 0.15 | 0.03 | <.001 | 0.10 | 0.04 | .007** | 0.10 | 0.04 | .007** | 0.05 | 0.05 | .048* |
| Sim _{cues} | -0.07 | 0.05 | .040* | -0.08 | 0.04 | .027* | -0.08 | 0.04 | .028* | -0.09 | 0.05 | .785 |
| Sim _{cues_answer} *Sim _{cues} | - | - | - | 0.11 | 0.04 | .006** | 0.11 | 0.04 | .006** | 0.11 | 0.06 | .192 |
| Condition | - | - | - | - | - | - | -0.03 | 0.06 | .628 | -0.02 | 0.06 | .056 |
| Sim _{cues_answer} *Condition | - | - | - | - | - | - | - | - | - | 0.10 | 0.07 | .616 |
| Sim _{cues} *Condition | - | - | - | - | - | - | - | - | - | 0.04 | 0.07 | .930 |
| Sim _{cues_answer} *Sim _{cues} *Condition | - | - | - | - | - | - | - | - | - | 0.01 | 0.09 | - |
| Model comparison | $\chi(2)^2 = 25.636, p < .001$ (vs. Null model) | | | $\chi(1)^2 = 7.530, p = .006$ (vs. Model 1) | | | $\chi(1)^2 = 0.238, p = .626$ (vs. Model 2) | | | $\chi(2)^2 = 2.828, p = .419$ (vs. Model 3) | | |
| Model ICC ^a | 0.169 | | | 0.169 | | | 0.167 | | | 0.169 | | |
| Mode fit: AIC | 2450.2 | | | 2444.7 | | | 2446.5 | | | 2449.6 | | |

Note: 894 observations, 96 participants, 24 RAT problems per participant; Sim_{cues_answer}: Cue-answer semantic similarity between cues and the answer; Sim_{cues}: Cue-cue semantic similarity; For Condition, no-switching was the reference group; **p* ≤ .05, ***p* ≤ .01, ****p* ≤ .001; Null model (random effect only): AIC = 2471.9, ICC = 0.165.
a. ICC (Intraclass Correlation) indicates the amount of variance in outcomes that is explained by the random effect.

problem-solving solution time included the main effects of the two task characteristics and their interaction term (Model 2, Table 8). Including the main effect of Condition and its interaction term with the two task characteristics did not further improve model fit (Model 3 and 4, Table 8). Table 9 presents the details of the best-fitting model.

There was a significant interaction between the cue-answer semantic similarity and cue-cue semantic similarity (*B* = 0.11, *p* = .006, see Table 9). As shown in Figure

3, when cue-cue semantic similarity increased (i.e. larger amount of competition), the coefficient of cue-answer semantic similarity turned positive (*B*_{low} = -0.01, *p* = 0.84, *B*_{medium} = 0.10, *p* = .01, *B*_{high} = 0.21, *p* < .01). RAT problems that were high in both cue-answer (i.e. high cue-answer relatedness) and cue-cue semantic similarity (i.e. high amount of competition) took the longest to solve, indicating greater difficulty. This aligns with the findings on accuracy, both suggesting that RAT problems with a

Table 9. Details of the Best-Fitting Model on RAT Problem-Solving Solution Time in Study 1 (Task Characteristics Measured Using Semantic Similarity).

| Predictor Variables | Best-fitting Model (Model 2, Table 5) | | | | | |
|---|---------------------------------------|-------------|----------|-----------|----------|--|
| | <i>B</i> (95% CI) | <i>SE B</i> | <i>t</i> | <i>df</i> | <i>p</i> | |
| Sim _{cues_answer} | 0.10 (0.03, 0.17) | 0.04 | 2.710 | 832.11 | .007 | |
| Sim _{cues} | -0.08 (-0.15, -0.01) | 0.04 | -2.220 | 834.93 | .027 | |
| Sim _{cues_answer} *Sim _{cues} | 0.11 (0.03, 0.19) | 0.04 | 2.740 | 839.760 | .006 | |
| Random Effect Variance | | | | | | |
| Participant | 0.163 | | | | | |
| Marginal R ² /Conditional R ² | 0.032/0.195 | | | | | |

Note: 894 observations, 96 participants, 24 RAT problems per participant; Sim_{cues_answer}: Cue-answer semantic similarity between cues and the answer; Sim_{cues}: Cue-cue semantic similarity.

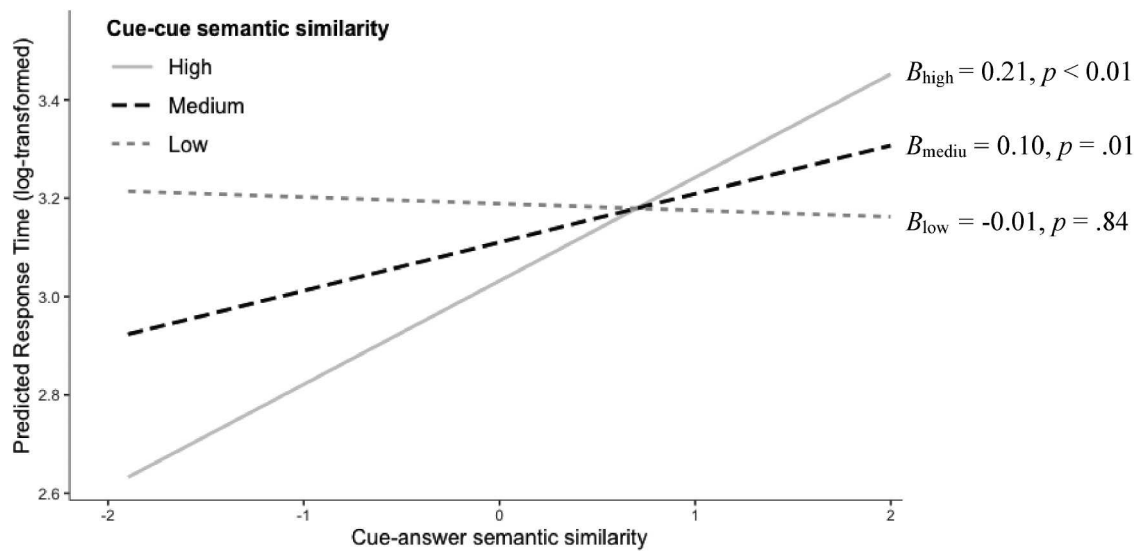


Figure 3. The Regression Coefficient of Cue-Answer Semantic Similarity When Cue-Cue Semantic Similarity was Low, Medium, and High (Study 1).

close answer (i.e. high cue-answer relatedness) can still be challenging when the competition in the initial problem space is high.

Model results: task-switching and RAT problem-solving

RAT problem-solving accuracy. According to the best-fitting models (i.e. models with lowest AIC value; see Tables 4 and 5), the coefficient associated with Condition was negative and statistically significant, suggesting a negative effect of task-switching on RAT problem-solving accuracy (associative strength: $B = -0.24, p = .010$, Table 5; semantic similarity: $B = -0.24, p = .008$, see Table 6). The negative task-switching effect was also not specific to a particular type of RAT problem examined because adding the interaction terms between Condition and the task characteristic measures did not further improve model fit (see Tables 3 and 6 for model comparisons). This pattern of results contrasts with previous studies, which reported a positive effect of task-switching on RAT problem-solving (Lu et al., 2017; Sio et al., 2017).

RAT problem-solving solution time. For both the associative strength and semantic similarity models, neither the inclusion of the main effect of Condition nor its interaction with the task characteristics further improved model fit (see Tables 7 and 8 for model comparisons), suggesting a non-significant impact of task-switching on solution time.

RAT problem-solving search process. We also explored whether task-switching impacted the search process during RAT problem solving. Specifically, we compared

the semantic similarity between adjacent responses between the no-switching and switching conditions. Table 10 presents the mean and standard deviation of the semantic similarity score in each condition.

Table 11 presents the summary of the linear mixed-effects model with the semantic similarity score for adjacent responses as the outcome variable and the variables Condition, trial type, time interval between adjacent responses, and the number of attempts as fixed effects (random effects: participants and tasks). Compared to the model with only random effects, the inclusion of the fixed effects significantly improved model fit, $\chi(4)^2 = 48.65, p < .001$. Both the time interval between adjacent responses and Condition were negatively associated with the semantic similarity between adjacent responses. The negative coefficient of the time interval between responses ($B = -.002, p < 0.001$) implies that adjacent responses with a shorter time interval were more semantically similar to each other. This is consistent with current research on the semantic search process, which suggests that it takes less time to generate semantically similar ideas (e.g. Smith et al., 2013; Troyer et al., 1997). We also found a negative effect of Condition ($B = -0.007, p = .003$), suggesting

Table 10. Mean and Standard Deviation of Semantic Similarity Score for Each Condition in Study 1.

| Condition | Semantic similarity Score for Adjacent Responses | |
|--------------|--|------|
| | M | SD |
| No-Switching | 1.15 | 0.11 |
| Switching | 1.13 | 0.10 |

a. Only includes solved trials.

Table 11. Summary of the Model on Semantic Similarity Scores for Adjacent Responses in Study 1.

| Predictor variables | <i>B</i> (95% <i>CI</i>) | <i>SE B</i> | <i>t</i> | <i>df</i> | <i>p</i> |
|---|-------------------------------|-------------|----------|-----------|----------|
| Solved Trial | −0.002 (−0.011, 0.007) | 0.004 | −0.434 | 2237 | .665 |
| Time Gap Between Responses | −0.002 (−0.002, −0.001) | < 0.001 | −5.850 | 2220 | <.001 |
| Number of Attempt | 0.001 (−0.001, 0.003) | 0.001 | 1.331 | 2207 | .183 |
| Condition | −0.007 (−0.011, −0.002) | 0.002 | −3.015 | 2210 | .003 |
| Randin Effect Variances | | | | | |
| Participant | <.001 | | | | |
| RAT Problem | <.001 | | | | |
| Model ICC ^a | 0.067 | | | | |
| Model AIC | 7041.7 | | | | |
| Marginal <i>R</i> ² / Conditional <i>R</i> ² | 0.022/0.088 | | | | |

Note: 2247 observations, 92 participants, 24 RAT problems per participant; For Condition, no-switching was the reference group; Null model: AIC = −7001.1, ICC = 0.067; Model comparison (vs. Null model: random effects only): $\chi(4)^2 = 48.65, p < .001$.

a. ICC (Intraclass Correlation) indicates the amount of variance in outcomes that is explained by the random effect.

that adjacent responses in the switching condition were less semantically similar to each other than those in the no-switching condition, indicative of a broader scope of search in the switching condition.

Discussion

In Study 1, we found an interaction between cue-answer relatedness and the amount of competition on RAT problem-solving accuracy, regardless of whether these task characteristics were measured using associative strength or semantic similarity. We found a negative effect of task-switching on RAT problem-solving accuracy. However, we observed a positive effect of task-switching on the scope of search conducted during RAT problem solving. The interaction pattern between cue-answer relatedness and the amount of competition on RAT problem-solving accuracy suggests that RAT problems with low cue-answer relatedness were difficult to solve. At the same time, RAT problems with high cue-answer relatedness can still be challenging when the amount of competition was high. The interaction pattern observed is in keeping with the results of Sio et al. (2022). When RAT problem-solving performance was measured in terms of solution time, we observed the same interaction pattern, but this was evident only when these characteristics were measured using semantic similarity. It could be the case that semantic similarity captures a unique aspect of search space that is likely to influence search efficiency. As discussed earlier, unlike associative strength, which focuses solely on the direct relationship between the cue and the target word, semantic similarity takes into

account the rich pattern of relationships spanning the entire network – a factor likely to influence search efficiency (e.g. Marko & Riečanský, 2021). Another possible explanation for the observed interaction could be the idiosyncrasies of our sample of participants or RAT problems. We examine this issue in Study 2.

When examining the effect of task-switching on RAT problem-solving performance, we found a negative effect of task-switching on RAT problem-solving accuracy. These results go against previous findings (Lu et al., 2017; Sio et al., 2017). It is important to note that although we followed the switching schedule used in Sio et al. (2017), which demonstrated a positive effect of task-switching, the study by Sio et al. (2017) was conducted in a controlled lab-based setting, whereas participants in the present study completed their tasks in an uncontrolled online environment. One might suggest that the negative effect observed is due to the online nature of this study. Online participants may face more distractions during the study and, as such, may be less efficient when searching for solutions. Untimely task-switching (i.e. switching occurring too early) may force them to abandon a search before it is completed, negatively impacting RAT problem-solving performance. To explore this possibility, in Study 2, we adjusted the task-switching interval from 10 to 15 s but kept the total presentation time unchanged (30 s). We examined if this change in the switching schedule would affect the direction of the task-switching effect.

Although we did not find a positive effect of task-switching on RAT problem-solving performance, we found that the average semantic similarity score between adjacent responses was lower in the switching condition than in the no-switching condition. This observed difference suggests that participants tended to conduct a broader search in the switching condition compared to the no-switching condition. If the semantic similarity score difference is the result of the interruption induced by task-switching, altering the degree of interruption (i.e. switching frequency) should influence the effect. In Study 2, where participants were instructed to switch tasks every 15 s instead of every 10 s (while the total presentation time remained the same), we expect a weaker effect of task-switching on the search scope.

Study 2

We conducted Study 2 to replicate and extend Study 1's findings. The settings were largely similar to those of Study 1, but with one modification to the switching schedule. While keeping the total problem-solving time at 30 s, we changed the task-switching schedule from every 10 s to every 15 s, reducing the number of

switches from 2 to 1. We examined if reducing switching frequency could influence the magnitude or direction of the task-switching effect on RAT problem solving. All data and syntax files are available through the Open Science Framework: <https://osf.io/uhxq9/>.

Methods

Participants

One hundred and fifty English speakers (77 females, 73 males) residing in the UK were recruited via Prolific, a different recruitment platform than the one used in Study 1. Only individuals with English as their first language and with a Prolific approval rate of at least 99% were recruited. The mean age was 41.31 years ($SD = 13.73$). 59% of the participants self-reported having a college degree or above. They were paid £2 for completing the study. Participants were asked to read the information sheet and give informed consent at the beginning of the study.

Task

Twenty-two RAT problems were selected from the normative set produced by Bowden and Jung-Beeman (2003). Half of these 22 RAT problems were used in Study 1. These 22 RAT problems were divided into two sets of 11 RAT problems, matched in terms of the task characteristics examined. Table 12 presents the mean, standard deviation, and the minimum and maximum values of each characteristic for each set.

Procedures

While the procedures in Study 2 were mostly similar to those of Study 1, there was a difference in the switching condition. Instead of three blocks as in Study 1, the RAT problems in Study 2 were presented in two blocks, with each problem displayed once for 15 s in each block, as opposed to 10 s in Study 1. The experiment was delivered online via Gorilla.sc (Anwyl-Irvine et al., 2020).

Results

On average, participants solved 33.87% ($SD = 16.19\%$) of the RAT problems. Participants on average took 11.85 s

Table 13. Means and Standard Deviations of RAT Problem-Solving Accuracy and Solution Time for Each Condition in Study 2.

| Condition | Accuracy | | Solution Time ^a (in seconds) | |
|--------------|----------|-----------|---|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| No-Switching | 0.34 | 0.21 | 12.12 | 4.90 |
| Switching | 0.34 | 0.19 | 11.74 | 5.20 |

a. Only includes solved trials.

($SD = 3.95$ s) to solve a RAT problem. These results are similar to those in Study 1, where the mean accuracy was 37.25% ($SD = 20.10\%$; Study 1 vs. Study 2: $t(248) = 1.40$, $p = 0.16$) and the mean solution time was 11.02 s ($SD = 4.11$ s, Study 1 vs. Study 2: $t(248) = 1.59$, $p = 0.11$). Table 13 presents the means and standard deviations of RAT problem-solving accuracy and solution time for each condition. Table S2 in supplemental material presents the solution rate for each RAT problem in our sample. Similar to Study 1, the solution rates were moderately and positively correlated with the normative data reported by Bowden and Jung-Beeman (2003), $r = .67$, $p = .001$.

Model construction and analysis

Task characteristics, task-switching, and RAT problem-solving performance. Similar of Study 1, RAT problem-solving performance was measured in terms of accuracy and solution time. RAT problem-solving accuracy and solution time were analysed using the same approach as in Study 1.

For accuracy, we constructed generalised linear mixed-effects models with a logit link function using whether or not the RAT problem was solved as the outcome variable. For solution time, we constructed linear mixed-effects models with log-transformed (base 2) solution time as the outcome variable. Solution time data were log-transformed to respect the normality assumption of residual (see Figure S3 in supplemental material for the normal Q-Q plots of residuals, with and without transforming the outcome variable).

For each performance measure, two sets of models were constructed to evaluate the relationships between task characteristics, task-switching, and RAT problem-solving performance. One set included task characteristics

Table 12. Mean, Standard Deviation, and Minimum and Maximum Values of Each Characteristic in Study 2.

| Task characteristics for RAT problems in Study 2 | Set 1 | | | Set 2 | | |
|--|------------------------|------|------|------------------------|------|------|
| | <i>M</i> (<i>SD</i>) | Min | Max | <i>M</i> (<i>SD</i>) | Min | Max |
| Based on associative strength | | | | | | |
| Associative Strength between the cues and the answer | 0.04 (0.04) | 0.00 | 0.13 | 0.03 (0.05) | 0.00 | 0.16 |
| Number of competing associates | 106.00 (24.51) | 62 | 134 | 108.18 (23.29) | 71 | 147 |
| Based on semantic similarity | | | | | | |
| Semantic similarity between the cues and the answer | 1.19 (0.04) | 1.10 | 1.24 | 1.17 (0.05) | 1.11 | 1.28 |
| Semantic similarity between the cues | 1.10 (0.04) | 1.04 | 1.17 | 1.10 (0.03) | 1.06 | 1.55 |

Table 14. Summary of the Models on RAT Problem-Solving Accuracy in Study 2 (Task Characteristics Measures Estimated Based on Associative Strength; Best-Fitting Model in Bold).

| Predictor Variables | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|------------------------|--|-------------|----------|---|-------------|--------------------|--|-------------|----------|--|-------------|----------|
| | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> |
| ASSOC | 0.30 | 0.05 | <.001*** | 0.15 | 0.06 | .007** | 0.15 | 0.06 | .007** | 0.13 | 0.08 | .099 |
| COMP | -0.11 | 0.05 | .030* | -0.11 | 0.05 | .030* | -0.11 | 0.05 | .030* | -0.18 | 0.07 | .013* |
| ASSOC*COMP | - | - | - | -0.31 | 0.05 | <.001*** | -0.31 | 0.05 | <.001*** | -0.29 | 0.07 | <.001*** |
| Condition | - | - | - | - | - | - | 0.02 | .08 | .803 | -0.01 | 0.10 | .900 |
| ASSOC*Condition | - | - | - | - | - | - | - | - | - | 0.04 | 0.11 | .725 |
| COMP*Condition | - | - | - | - | - | - | - | - | - | 0.14 | 0.10 | .174 |
| ASSOC*COMP*Condition | - | - | - | - | - | - | - | - | - | -0.05 | 0.10 | .603 |
| Model comparison | $\chi(2)^2 = 102.48, p < .001$ (vs. Null model) | | | $\chi(1)^2 = 42.19, p < .001$ (vs. Model 1) | | | $\chi(1)^2 = 0.062, p = .804$ (vs. Model 2) | | | $\chi(2)^2 = 2.302, p = .512$ (vs. Model 3) | | |
| Model ICC ^a | 0.108 | | | 0.112 | | | 0.112 | | | 0.112 | | |
| Model fit: AIC | 4034 | | | 3993.8 | | | 3995.8 | | | 3999.5 | | |

Note: 3300 observations, 150 participants, 22 RAT problems per participant; ASSOC: Cue-answer associative strength; COMP: Number of competing associates; For Condition, no-switching was the reference group; * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$; Null model (random effect only): AIC = 4132.5, ICC = 0.100.
a. ICC (Intraclass Correlation) indicates the amount of variance in outcomes that is explained by the random effect.

measured using associative strength as the predictor variables, while the other included task characteristics measured using semantic similarity. The predictor variables were entered in the sequence same as in Study 1. For each model set, we selected the model with the lowest AIC value as the best-fitting model.

Task-switching and RAT problem-solving search process. When examining the effect of task-switching on semantic similarity scores for adjacent responses (see Table 21 for the mean and standard deviation of the score for each condition), we followed the same procedures as in Study 1: we constructed a generalised linear mixed-effects model using the semantic similarity score as the outcome variable. We included the same random (participants and tasks) and fixed effects (Condition, trial type, time interval between adjacent responses, and the number of attempts) used in Study1. Again, we transformed the semantic similarity scores using a Box-Cox transformation (lambda value: -3) to reduce skewness to avoid violating the

assumption of normality in residuals. Supplemental material Figure S4 present the normal Q-Q plots of residuals, with and without transforming the outcome variable.

We used the R-packages as in Study 1 to conduct all the analyses and used a significance level of $\alpha = 0.05$ for all the statistical tests.

Model results: task characteristics and RAT problem solving

In this section, we focused on the results related to effects of the two task characteristics on RAT problem-solving. The next section *Model Results: Task-Switching and RAT Problem-Solving* focused on the results concerning the effect of task-switching on RAT problem-solving.

RAT problem-solving accuracy. For both model sets (associative strength and semantic similarity models), the best-fitting model (i.e. lowest AIC value) included the main effects of the two task characteristics and their interaction term (see Model 2, Tables 14 and 15).

Table 15. Summary of the Models on RAT Problem-Solving Accuracy in Study 2 (Task Characteristics Measures Estimated Based on Semantic Similarity).

| Predictor Variables | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|--|--|-------------|----------|--|-------------|--------------------|--|-------------|----------|--|-------------|----------|
| | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> |
| Sim _{cues_answer} | -0.08 | 0.04 | .053 | -0.01 | 0.04 | .789 | -0.01 | 0.06 | .790 | 0.02 | 0.06 | .804 |
| Sim _{cues} | -0.22 | 0.04 | <.001*** | -0.29 | 0.04 | <.001*** | -0.29 | 0.05 | <.001*** | -0.28 | 0.06 | <.001*** |
| Sim _{cues_answer} *Sim _{cues} | - | - | - | -0.26 | 0.04 | <.001*** | -0.26 | 0.06 | <.001*** | -0.28 | 0.05 | <.001*** |
| Condition | - | - | - | - | - | - | 0.02 | 0.09 | .819 | 0.01 | 0.08 | .950 |
| Sim _{cues_answer} *Condition | - | - | - | - | - | - | - | - | - | -0.05 | 0.09 | .543 |
| Sim _{cues} *Condition | - | - | - | - | - | - | - | - | - | -0.03 | 0.09 | .756 |
| Sim _{cues_answer} *Sim _{cues} *Condition | - | - | - | - | - | - | - | - | - | 0.02 | 0.08 | .752 |
| Model comparison | $\chi(2)^2 = 46.860, p < .001$ (vs. Null model) | | | $\chi(1)^2 = 52.033, p < .001$ (vs. Model 1) | | | $\chi(1)^2 = 0.052, p = .820$ (vs. Model 2) | | | $\chi(2)^2 = 0.747, p = .862$ (vs. Model 3) | | |
| Model ICC | 0.103 | | | 0.107 | | | 0.107 | | | 0.108 | | |
| Model fit: AIC | 4089.6 | | | 4039.6 | | | 4041.6 | | | 4046.8 | | |

Note: 3300 observations, 150 participants, 22 RAT problems per participant; Sim_{cues_answer}: Semantic similarity between cues and the answer; Sim_{cues}: Semantic similarity between cues; For Condition, no-switching was the reference group; * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$; Null model (random effect only): AIC = 4132.5, ICC = 0.100.

a. ICC (Intraclass Correlation) indicates the amount of variance in outcomes that is explained by the random effect.

Table 16. Details of the Best-Fitting Model on RAT Problem-Solving Accuracy in Study 2 (Task Characteristics Measured Using Associative Strength).

| Predictor Variables | Best-fitting Model (Model 2, Table 8) | | | | |
|---|---------------------------------------|-------------|----------|----------|-----------------------------|
| | <i>B</i> (95% <i>CI</i>) | <i>SE B</i> | <i>z</i> | <i>p</i> | Odds Ratio (95% <i>CI</i>) |
| ASSOC | 0.15 (0.04, 0.26) | 0.06 | 2.697 | .007 | 1.17 (1.04, 1.30) |
| COMP | -0.11 (-0.21, -0.01) | 0.05 | -2.168 | .030 | 0.90 (0.81, 0.99) |
| ASSOC*COMP | -0.31 (-0.41, -0.22) | 0.05 | -6.347 | <.001 | 0.73 (0.66, 0.80) |
| Random effect variance | | | | | |
| Participant | 0.414 | | | | |
| Marginal <i>R</i> ² /conditional <i>R</i> ² | 0.053/0.159 | | | | |

Note: 3300 observations, 150 participants, 22 RAT problem per participant; ASSOC: Cue-answer associative strength; COMP: Number of competing associates; The estimated logit coefficients were transformed to odds ratios for easier interpretation.

Tables 16 and 17 present the details of the best-fitting models. Including the main effect of Condition or the interaction terms between Condition and the two task characteristics did not further improve model fit (see Table 14 and 16 for model comparisons).

We observed the same interaction between cue-answer relatedness and the amount of competition, regardless of whether the task characteristics were measured using associative strength ($B = -0.32, p < .001$, Table 16) or semantic similarity ($B = -0.26, p < .001$, Table 17). When the amount of competition was low (1 *SD* below the

mean), cue-answer relatedness was positively associated with RAT problem-solving accuracy. However, this association became less positive and eventually negative as the amount of competition increased (see Figures 4 and 5). As presented in Figures 4 and 5, our results suggest that RAT problems with low cue-answer relatedness (i.e. remote answers) were in general difficult to solve. RAT problems with high cue-answer relatedness (i.e. close answers) also became difficult when competition with the initial problem space was high. This pattern of results aligns with our findings from Study 1.

Table 17. Details of the Best-Fitting Model on RAT Problem-Solving Accuracy in Study 2 (Task Characteristics Measured Using Semantic Similarity).

| Predictor Variables | Best-fitting Model (Model 2, Table 9) | | | | |
|---|---------------------------------------|-------------|----------|----------|-----------------------------|
| | <i>B</i> (95% <i>CI</i>) | <i>SE B</i> | <i>z</i> | <i>p</i> | Odds Ratio (95% <i>CI</i>) |
| Sim _{cues_answer} | -0.01 (-0.10, 0.07) | 0.04 | -0.267 | 0.789 | 0.99 (0.90, 1.07) |
| Sim _{cues} | -0.29 (-0.38, -0.21) | 0.04 | -6.602 | <.001 | 0.74 (0.68, 0.81) |
| Sim _{cues_answer} *Sim _{cues} | -0.26 (-0.34, -0.19) | 0.04 | -7.031 | <.001 | 0.77 (0.71, 0.83) |
| Random Effect Variance | | | | | |
| Participant | 0.396 | | | | |
| Marginal <i>R</i> ² /Conditional <i>R</i> ² | 0.045/0.147 | | | | |

Note: 3300 observations, 150 participants, 22 RAT problems per participant; Sim_{cues_answer}: Semantic similarity between cues and the answer; Sim_{cues}: Semantic similarity between cues; The estimated logit coefficients were transformed to odds ratios for easier interpretation.

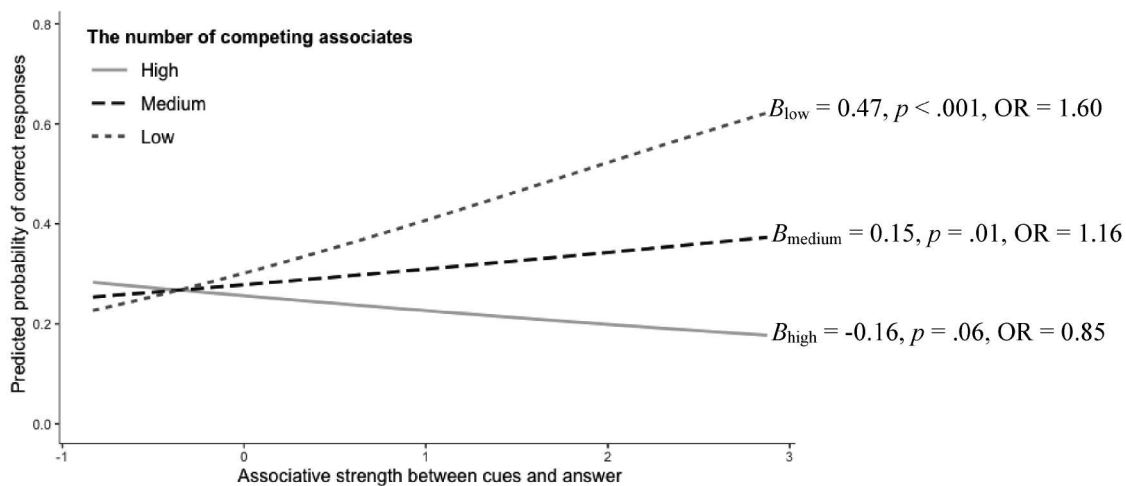


Figure 4. Regression Coefficient of Cue-Answer Associative Strength When the Number of Competing Associates was Low, Medium, and High (Study 2).

Note: OR = Odds Ratio. The estimated logit coefficients were transformed to odds ratios for easier interpretation.

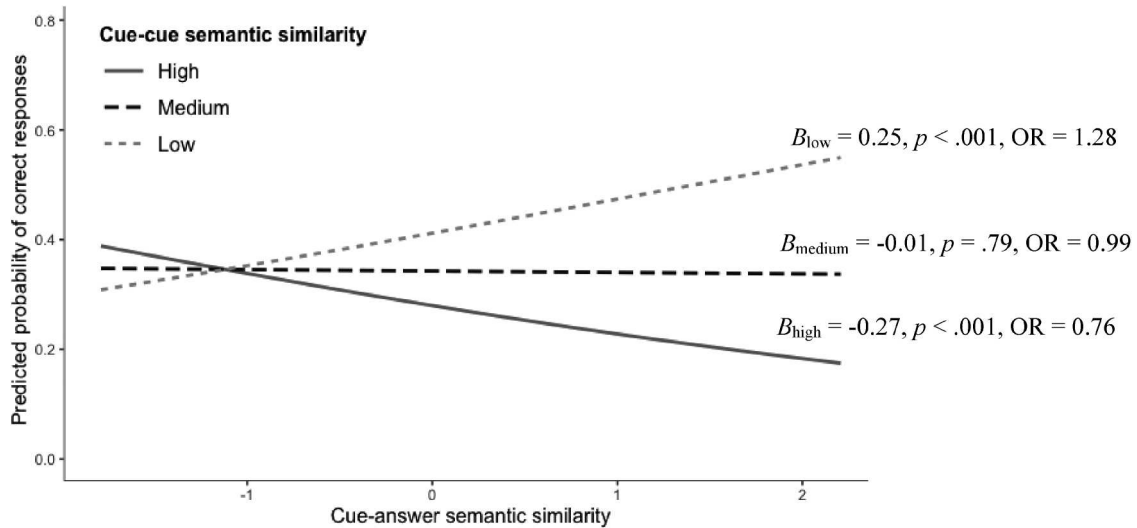


Figure 5. Regression Coefficient of Cue-Answer Semantic Similarity When the Cue-Cue Semantic Similarity was Low, Medium, and High (Study 2).

Note: OR = Odds Ratio. The estimated logit coefficients were transformed to odds ratios for easier interpretation.

Table 18. Summary of the Models on RAT Problem-Solving Solution Time (Task Characteristics Measured Using Associative Strength) in Study 2.

| Predictor Variables | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|----------------------|---|-------------|----------|--|-------------|----------|--|-------------|----------|--|-----------|----------|
| | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE</i> | <i>p</i> |
| ASSOC | -0.01 | 0.03 | .871 | 0.01 | 0.04 | .709 | 0.01 | 0.04 | .685 | -0.01 | 0.05 | .853 |
| COMP | -0.04 | 0.04 | .290 | -0.05 | 0.04 | .189 | -0.05 | 0.04 | .214 | -0.08 | 0.07 | .140 |
| ASSOC*COMP | - | - | - | 0.04 | 0.03 | .162 | 0.04 | 0.03 | .167 | -0.02 | 0.07 | .578 |
| Condition | - | - | - | - | - | - | -0.11 | 0.06 | .055 | -0.01 | 0.10 | .325 |
| ASSOC*Condition | - | - | - | - | - | - | - | - | - | 0.04 | 0.11 | .506 |
| COMP*Condition | - | - | - | - | - | - | - | - | - | 0.14 | 0.10 | .399 |
| ASSOC*COMP*Condition | - | - | - | - | - | - | - | - | - | -0.05 | 0.10 | .570 |
| Model comparison | $\chi(2)^2 = 1.791, p = .408$ (vs. Null model) | | | $\chi(1)^2 = 1.966, p = .161$ (vs. Model 1) | | | $\chi(1)^2 = 3.707, p = .054$ (vs. Model 2) | | | $\chi(2)^2 = 1.374, p = .712$ (vs. Model 3) | | |
| Model ICC | 0.112 | | | 0.112 | | | 0.113 | | | 0.112 | | |
| Model fit: AIC | 3053.6 | | | 3053.6 | | | 3051.9 | | | 3056.5 | | |

Note: 1118 observations, 149 participants, 22 RAT problems per participant; ASSOC: Cue-answer associative strength; COMP: Number of competing associates; For Condition, no-switching was the reference group; Null model (random effect only): AIC = 3051.4, ICC = 0.112.
a. ICC (Intraclass Correlation) indicates the amount of variance in outcomes that is explained by the random effect.

RAT problem-solving solution time. When task characteristics were measured using associative strength, the null model was the best-fitting models (i.e. lowest AIC value; see Table 18 for model comparisons), suggesting none of the predictor variables predicted solution time.

When task characteristics were measured using semantic similarity, the best-fitting model included the main effects of the two characteristics and their interaction term, and the main effect of Condition (see Table 19 for model comparisons and 14b for the details of the best-fitting model). Similar to the findings of Study 1, there was an interaction between the two task characteristics ($B = 0.06, p = .019$). As shown in Figure 6, solution time increased as the cue-answer semantic similarity increased, but this positive association was observed only when the cue-cue

semantic similarity was at medium level or above ($B_{low} = 0.02, p = .24$; $B_{medium} = 0.03, p < .01$; $B_{high} = 0.05, p < .01$). According to the results, RAT problems that were high in both cue-answer and cue-cue semantic similarity took the longest to solve. The interaction pattern is in keeping with what we observed in Study 1.

In sum, Study 2 replicated the interaction of task characteristics on RAT problem-solving performance observed in Study 1.

Model results: task-switching and RAT problem-solving

RAT problem-solving accuracy. Unlike in Study 1, we did not observe a significant effect of task-switching. In both the model sets – the one using associative strength (see Tables 14 and 15) and the one using semantic similarity

Table 19. Summary of the Models on RAT Problem-Solving Solution Time (Task Characteristics Measured Using Semantic Similarity; Besting-Fitting Model in Bold) in Study 2.

| Predictor Variables | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|--|--|-------------|----------|--|-------------|----------|--|--------------------|--------------------|--|-------------|----------|
| | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> | <i>B</i> | <i>SE B</i> | <i>p</i> |
| Sim _{cues_answer} | 0.13 | 0.03 | <.001*** | 0.11 | 0.03 | <.001*** | 0.11 | 0.03 | <.001*** | 0.14 | 0.04 | .001*** |
| Sim _{cues} | 0.05 | 0.03 | .115 | 0.67 | 0.03 | .028** | 0.07 | 0.03 | .028* | 0.04 | 0.04 | .311 |
| Sim _{cues_answer} *Sim _{cues} | – | – | – | 0.59 | 0.03 | .021** | 0.06 | 0.03 | .019* | 0.07 | 0.04 | .049* |
| Condition | – | – | – | – | – | – | –0.11 | 0.05 | .047* | –0.10 | 0.06 | .073 |
| Sim _{cues_answer} *Condition | – | – | – | – | – | – | – | – | – | –0.06 | 0.06 | .314 |
| Sim _{cues} *Condition | – | – | – | – | – | – | – | – | – | 0.05 | 0.06 | .436 |
| Sim _{cues_answer} *Sim _{cues} *Condition | – | – | – | – | – | – | – | – | – | –0.03 | 0.05 | .616 |
| Model comparison | $\chi(2)^2 = 25.288, p < .001$ (vs. Null model) | | | $\chi(1)^2 = 5.381, p = .020$ (vs. Model 1) | | | $\chi(1)^2 = 3.960, p = .046$ (vs. Model 2) | | | $\chi(2)^2 = 2.170, p = .538$ (vs. Model 3) | | |
| Model ICC ^a | 0.109 | | | 0.109 | | | 0.109 | | | 0.109 | | |
| Model fit: AIC | 3053.6 | | | 3026.7 | | | 3024.7 | | | 3028.6 | | |

Note: 1118 observations, 149 participants, 22 RAT problems per participant; Sim_{cues_answer}: Semantic similarity between cues and the answer; Sim_{cues}: Semantic similarity between cues; For Condition, no-switching was the reference group; * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$; Null model (random effect only): AIC = 3051.4, ICC = 0.112.

a. ICC (Intraclass Correlation) indicates the amount of variance in outcomes that is explained by the random effect.

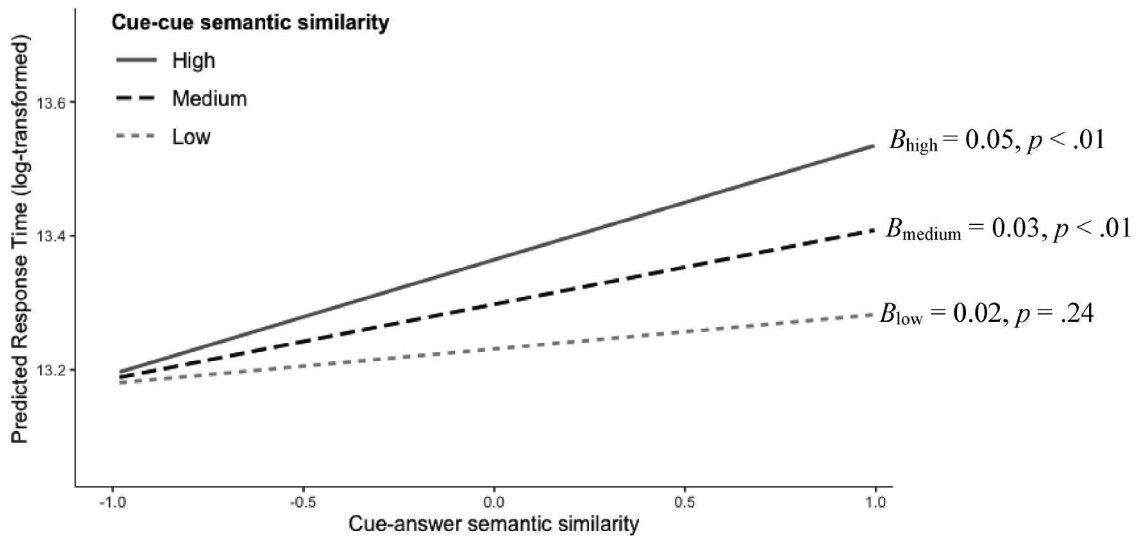


Figure 6. Regression Coefficient of Cue-Answer Semantic Similarity When the Cue-Cue Semantic Similarity was Low, Medium, and High (Study 2).

measures (see Tables 16 and 17), the effect of Condition was not significant in the best-fitting model, all $p > 0.8$.

RAT problem-solving solution time. Different from Study 1, we observed a negative association between task-switching and solution time, indicating that participants solved the RAT problems more quickly when presented in a switching manner than in a no-switching manner. The negative association was significant only for the model with task characteristics measured using semantic similarity ($B = -0.03, p = .047$, see Table 20).

RAT problem-solving search process. As in Study 1, we examined the effect of task-switching on semantic similarity scores for adjacent responses (see Table 21 for the mean and standard deviation of the score for each

condition). Compared to the null model, which only included random effects, the inclusion of the fixed effects – Condition, trial type, time interval between adjacent responses, and the number of attempts – significantly improved model fit, $\chi(4)^2 = 62.94, p < .001$. Table 22 presents the model summary. We observed a negative effect of task-switching on semantic similarity scores. Although the direction of this effect was the same as in Study 1, it was not significant ($B = -0.03, p = .074$; see Table 22).

Discussion

In Study 2, we replicated the interaction effect between the two task characteristics on RAT problem

Table 20. Details of the Best-Fitting Model on RAT Problem-Solving Solution Time (Task Characteristics Measured Using Semantic Similarity) in Study 2.

| Predictor Variables | Best-fitting Model (Model 3, Table 14) | | | | |
|---|--|-------------|----------|-----------|----------|
| | <i>B</i> (95% <i>CI</i>) | <i>SE B</i> | <i>t</i> | <i>df</i> | <i>p</i> |
| Sim _{cues_answer} | 0.11 (0.05, 0.17) | 0.03 | 3.701 | 1058.16 | <.001 |
| Sim _{cues} | 0.07 (0.01, 0.13) | 0.03 | 2.201 | 1046.44 | .028 |
| Sim _{cues_answer} *Sim _{cues} | 0.06 (0.01, 0.11) | 0.03 | −1.987 | 1074.76 | .019 |
| Condition | −0.11 (−0.22, < −0.001) | 0.05 | 2.356 | 1072.76 | .047 |
| Random Effect Variance | | | | | |
| Participant | 0.097 | | | | |
| Marginal R ² /Conditional R ² | 0.029/0.134 | | | | |

Note: 1118 observations, 149 participants, 22 RAT problems per participant; Sim_{cues_answer}: Semantic similarity between cues and the answer; Sim_{cues}: Semantic similarity between cues; For Condition, no-switching was the reference group.

performance observed in Study 1, showing that RAT problems with low cue-answer relatedness (i.e. remote answers) were in general difficult to solve. RAT problems with high cue-answer relatedness (i.e. close answers) also became difficult when competition with the initial problem space was high.

However, we observed differing findings between Study 2 and Study 1 regarding the effect of task-switching on RAT problem-solving. In Study 2, we did not find any negative effect of task-switching on RAT problem-solving accuracy as observed in Study 1. Also, in Study

2, participants solved RAT problems more quickly in the switching than in the no-switching condition, which was not observed in Study 1. In examining the search scope, we observed that, similar to Study 1, the semantic similarity scores for adjacent ideas were lower in the switching condition than in the no-switching condition. However, this difference was not statistically significant ($p = .074$). The differing findings between Study 1 and Study 2 may be due to the difference in switching frequency between the two conduction, a point that will be discussed further in the General Discussion.

Table 21. Mean and Standard Deviation of Semantic Similarity Score for Each Condition in Study 2.

| Condition | Semantic similarity Score for Adjacent Responses | |
|--------------|--|-----------|
| | <i>M</i> | <i>SD</i> |
| No-Switching | 1.14 | 0.11 |
| Switching | 1.14 | 0.10 |

Table 22. Summary of the Model on Semantic Similarity Scores of Adjacent Responses in Study 2.

| Predictor Variables | <i>B</i> (95% <i>CI</i>) | <i>SE B</i> | <i>t</i> | <i>df</i> | <i>p</i> |
|---|-----------------------------|-------------|----------|-----------|----------|
| Solved Trial | −0.003 (−0.010, 0.003) | 0.003 | −0.977 | 3256 | .329 |
| Time Gap Between Responses | −0.001 (−0.002, −0.001) | < 0.001 | −6.887 | 3714 | <.001 |
| Number of Attempt | 0.001 (0.000, 0.002) | 0.001 | 1.931 | 3095 | .053 |
| Condition | −0.003 (−0.006, < 0.001) | 0.002 | −1.785 | 3669 | .074 |
| Randin effect variances | | | | | |
| Participant | <.001 | | | | |
| RAT Problem | <.001 | | | | |
| Model ICC ^a | 0.031 | | | | |
| Model AIC | −11757 | | | | |
| Marginal R ² /Conditional R ² | 0.017/0.046 | | | | |

Note: 3777 observations, 144 participants, 22 RAT problems per participant; For Condition, no-switching was the reference group; Null model: AIC = −11702, ICC = 0.031; Model comparison (vs. Null model: random effects only): $\chi(4)^2 = 62.94$, $p < .001$.

a. ICC (Intraclass Correlation) indicates the amount of variance in outcomes that is explained by the random effect.

General discussion

In our studies, we investigated the association between RAT problem-solving performance and task characteristics – more specifically, (1) cue-answer relatedness and (2) the amount of competition in the initial problem space. These two task characteristics were measured using associative strength and semantic similarity. We also examined the effect of task switching on RAT problem solving and whether this effect was moderated by the two task characteristics examined.

Task characteristics and RAT problem solving

Interaction between task characteristics

Recent studies have identified an interaction between cue-answer relatedness and the amount of competition in the initial problem space when predicting RAT problem-solving performance. However, the interaction pattern was not consistent across studies. For example, Becker et al. (2022) found that cue-answer relatedness was positively associated with RAT problem-solving performance, and this association was the strongest when the amount of competition in the initial space was *high*. By contrast, Sio et al. (2022) reported that the positive association was strongest when the amount of competition in the initial space was *low*. Relatedly, these studies reached different conclusions regarding the relative difficulty of RAT problems. For example, Becker et al. (2022) found

that RAT problems with high cue-answer relatedness and a high amount of competition were easiest to solve. In contrast, Sio et al. (2022) found RAT problems high on both these two characteristics were the hardest.

As mentioned earlier, a potential reason for the discrepancy between the findings of Becker et al. (2022) and Sio et al. (2022) is the different methods used to measure the task characteristics. Becker et al. (2022) used semantic similarity, while Sio et al. (2022) used associative strength. In the present studies, we examined whether the use of different approaches accounts for the difference in their findings by comparing the results obtained with both methods.

We found a significant interaction between cue-answer relatedness and the amount of competition when predicting RAT problem-solving accuracy, and the interaction pattern was consistent with Sio et al. (2022): the positive association between cue-answer relatedness and RAT problem-solving performance was strongest when the amount of competition in the initial space was *low*. We also obtained a similar pattern of findings for RAT problem-solving solution time but only when the cue-answer relatedness and the amount of competition were measured using semantic similarity.

For the RAT problems we examined, our results suggest that RAT problems can be challenging due to low cue-answer relatedness (i.e. remote answers). However, those with high cue-answer relatedness (i.e. close answers) can also be difficult if the amount of competition within the initial problem space is high. The difficulty specific to solving RAT problems having close answers is consistent with a well-known finding in memory research that the likelihood of an item being retrieved in the presence of a cue not only depends on its relation to the cue but also on the interference from other items associated with that cue (e.g. Anderson et al., 1994; Watkins & Watkins, 1975).

Measurement of task characteristics

In addition to observing a similar pattern of interaction across our two studies, we also observed similar interaction patterns regardless of whether the two characteristics (i.e. cue-answer relatedness and amount of competition) were measured using associative strength or semantic similarity. This suggests that the discrepancy in the findings between Becker et al. (2022) and Sio et al. (2022) may be due to factors other than the use of different methods. One reason may be related to the particular RAT problems selected for each study. Specifically, the RAT problems used in Becker et al. (2022) were generally easier than those used in Sio et al. (2022), based on Bowden and Jung-Beeman (2003) norm data (mean solution rate: 0.67 for Becker et al. and 0.36 for Sio et al.).

We also found notable differences in the cue-answer associative strength – a measure of the cue-answer relatedness – between the RAT problems used in the two studies. In Becker et al. (2022), the average cue-answer associative strength was 0.15, ranging from 0.01–0.62, while in Sio et al. (2022), it was 0.04, ranging from 0 to 0.16. It is possible that the two task characteristics interact differently within these distinct ranges, leading to different interaction patterns. Future research could explore this hypothesis by using a broader range of RAT problems.

Irrespective of discrepancies in the interaction patterns observed, the fact that we noted an interaction between task characteristics suggests that studies aiming to manipulate (or control) RAT problem difficulty should consider both task characteristics.

Task switching and RAT problem solving

Studies 1 and 2 also examined the effect of task-switching on RAT problem solving and investigated whether the effect was influenced by the task characteristics examined. The procedures for Studies 1 and 2 were largely similar, except that participants in Study 1 switched between tasks every 10 s, whereas those in Study 2 switched every 15 s.

RAT problem-solving accuracy

Study 1 revealed a negative effect of task-switching on accuracy, while Study 2, with reduced switching frequency, showed no significant effect. In both studies, the effect of task-switching was not moderated by the task characteristics examined.

The absence of a positive effect of task-switching on RAT problem-solving accuracy in our studies suggests that task-switching may not be as beneficial as indicated by prior studies. However, the discrepancy between our findings and previous findings could also be due to methodological differences. Unlike previous task-switching studies that usually used fewer tasks and a between-subject design, participants in our studies tackled a larger number of RAT problems in both the switching and no-switching conditions. This suggests our participants may have retained more unsolved problems in their memory. While many past studies suggested that maintaining unsolved problems can facilitate unconscious problem-solving processes, e.g. spreading activation and incorporating new information (e.g. Dijksterhuis & Aarts, 2010; Yaniv & Meyer, 1987; Zeigarnik, 1927/1938), it is unclear if holding too many unsolved problems simultaneously would disrupt these processes. Future studies could examine whether the quantity of tasks involved in switching influences the effect of task-switching.

RAT problem-solving solution time

When examining the effect of task-switching on RAT problem-solving solution time, we did not find any significant effect of task-switching in Study 1. However, in Study 2, with reduced switching frequency, we found that participants solved the RAT problems more quickly when presented in the switching condition. The effect on solution time observed was not moderated by the task characteristics examined.

In sum, Studies 1 and 2, both using different switching frequencies (2 switches in Study 1 vs. 1 switch in Study 2), showed different effects of task-switching on RAT problem solving. First, Study 1 showed a negative task-switching effect on accuracy, while Study 2 showed no such effect. Second, Study 1 showed no effect of task-switching in enhancing RAT problem-solving speed, while Study 2 showed such effect. These results suggest that switching frequency may moderate the effect of task switching on RAT problem-solving performance. It should be noted, however, that the discrepancy in the findings between Studies 1 and 2 could also be due to the fact that there were systematic differences between participants in the two studies (for example, Study 1 participants were from the US, while Study 2 are from the UK). Future studies could adopt within-subject designs to more rigorously examine the relation between switching frequency and the effect of task switching.

RAT problem-solving search scope

In Study 1, task-switching was associated with a broader search scope. In Study 2, with a reduced switching frequency, although we still found a similar effect of task-switching on search scope, it was not significant. We believe that presenting problems in short, segmented sessions (i.e. task-switching) may have induced a sense of time pressure, prompting individuals to change their search direction more rapidly when their current approach was unfruitful, hence facilitating a broader search. Also, each time individuals return back to the task, it could serve as a fresh start, allowing them to explore new search paths. Following this rationale, it is not surprising that the positive effect of task-switching on search scope was weaker in Study 2 than in Study 1, as participants in Study 2 switched between tasks less frequently. Future research could examine how changes in switching frequency impact on the scope of the semantic search in RAT problem solving.

It is important to highlight that a broader search was not associated with better RAT problem-solving performance in our studies. In Study 1, we found that while task-switching could facilitate a broader search, the manipulation had a negative impact on RAT problem-solving

accuracy. This pattern of results seems to be inconsistent with the suggestion that a broad search is crucial for creative problem solving in general, as many previous studies have proposed (see Nijstad et al., 2010 for a review). It might be the case that a broad search is more beneficial for tasks that emphasise divergent thinking, compared to RAT problem solving, which involve a convergent search process of finding a single correct solution to a problem where there are usually multiple apparent paths (Brophy, 1998; Hommel, 2012; Lee & Theriault, 2013). One direction for future research would be to examine the effect of task-switching on divergent thinking performance.

Limitations

There are several limitations to our studies. First, our analysis of the search scope excluded trials that were solved on the first attempt. As a result, data from individuals proficient in solving RAT problems (i.e. those who solved most of the RAT problems on the first attempt) were underrepresented. This may introduce bias into our results that may reduce the generalizability of our finding to highly creative individuals.

Second, our studies examined whether the task-switching effect was moderated by task characteristics associated with RAT problem-solving performance. We did not include non-task characteristics – e.g. individuals' attentional characteristics and the structure of their semantic networks (Beatty et al., 2014; Cosgrove et al., 2021) – characteristics that are also likely to influence RAT problem-solving performance. Future research could examine whether these individual characteristics moderate the effect of task-switching on RAT problem solving.

Third, our studies did not include a comprehensive range of RAT problems. For example, we only examined problems where the sum of the associative strengths from the three cues to the answer ranged from 0 to 0.16, whereas theoretically, this value can range from 0 to 3, with the associative strength for each cue to the answer varying from 0 to 1. Whether our findings apply beyond the range tested is an open question. Future research could use RAT problems beyond the range of cue-answer associative strength values adopted within our study to provide a more comprehensive understanding of the relationship between task characteristics and RAT problem-solving performance.

Conclusion

Our studies replicated the results of Sio et al. (2022), demonstrating a positive association between cue-answer relatedness and RAT problem-solving

performance, with this association being strongest when competition within the initial search space is low. Moreover, we observed the same pattern of interaction, regardless of whether cue-answer relatedness and the amount of competition were measured using associative strength or semantic similarity. Unlike previous studies, we did not find any positive effect of task-switching on RAT problem-solving accuracy. However, we found evidence suggesting that task-switching could improve problem-solving speed and facilitate a broad search scope.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was funded by the British Academy/Leverhulme Small Research Grant SRG2021\210385 awarded to Ut Na Sio. Hugues Lortie-Forgues is funded by a research grant from Research England via the Centre for Mathematical Cognition. John E. Marsh is funded by a grant received from the Bial Foundation (grant number 201/20) and the German Academic Exchange Service (Deutscher Akademischer Austauschdienst [DAAD, grant number 91875827]).

Ethical approval

Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors. Ethical approval for the studies were granted by the University Research Ethics Committee at the University of Sheffield.

Data availability statement

The data that support the findings of this study and the syntax files are openly available in Open Science Framework at: <https://osf.io/uhxq9/>.

References

- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1063–1087. <https://doi.org/10.1037/0278-7393.20.5.1063>
- Anwyl-Irvine, A. L., Massoné, J., Flitton, A., Kirkham, N. Z., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Bates, D., Mächler, M., Bolker, B., & Walker S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods*, 53(2), 757–780. <https://doi.org/10.3758/s13428-020-01453-w>
- Beaty, R. E., & Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts*, 6(4), 309–319. <https://doi.org/10.1037/a0029171>
- Beaty, R. E., Silvia, P. J., Nusbaum, E. C., Jauk, E., & Benedek, M. (2014). The roles of associative and executive processes in creative cognition. *Memory & Cognition*, 42(7), 1186–1197. <https://doi.org/10.3758/s13421-014-0428-8>
- Becker, M., Davis, S., & Cabeza, R. (2022). Between automatic and control processes: How relationships between problem elements interact to facilitate or impede insight. *Memory & Cognition*, 50(8), 1719–1734. <https://doi.org/10.3758/s13421-022-01277-3>
- Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers*, 35(4), 634–639. <https://doi.org/10.3758/BF03195543>
- Brophy, D. R. (1998). Understanding, measuring, and enhancing individual creative problem-solving efforts. *Creativity Research Journal*, 11(2), 123–150. https://doi.org/10.1207/s15326934crj1102_4
- Chrysikou, E. G. (2019). Creativity in and out of (cognitive) control. *Current Opinion in Behavioral Sciences*, 27, 94–99. <https://doi.org/10.1016/j.cobeha.2018.09.014>
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407–428. <https://doi.org/10.1037/0033-295X.82.6.407>
- Cosgrove, A. L., Kenett, Y. N., Beaty, R. E., & Diaz, M. T. (2021). Quantifying flexibility in thought: The resiliency of semantic networks differs across the lifespan. *Cognition*, 211, 104631. <https://doi.org/10.1016/j.cognition.2021.104631>
- Davelaar, E. J. (2015). Semantic search in the remote associates test. *Topics in Cognitive Science*, 7(3), 494–512. <https://doi.org/10.1111/tops.12146>
- De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The “Small World of Words” English word association norms for over 12,000 cue words. *Behavior Research Methods*, 51(3), 987–1006. <https://doi.org/10.3758/s13428-018-1115-7>
- Dijksterhuis, A., & Aarts, H. (2010). Goals, attention, and (un)consciousness. *Annual Review of Psychology*, 61(1), 467–490. <https://doi.org/10.1146/annurev.psych.093008.100445>
- George, T., & Wiley, J. (2019). Fixation, flexibility, and forgetting during alternate uses tasks. *Psychology of Aesthetics, Creativity, and the Arts*, 13(3), 305–313. <https://doi.org/10.1037/aca0000173>
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. <https://doi.org/10.1111/2041-210X.12504>
- Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun - An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, 47(4), 930–944. <https://doi.org/10.3758/s13428-014-0529-0>
- Gupta, N., Jang, Y., Mednick, S. C., & Huber, D. E. (2012). The road not taken. *Psychological Science*, 23(3), 288–294. <https://doi.org/10.1177/0956797611429710>

- Hass, R. W. (2017). Semantic search during divergent thinking. *Cognition*, 166, 344–357. <https://doi.org/10.1016/j.cognition.2017.05.039>
- Hommel, B. (2012). Convergent and divergent operations in cognitive search. In P. M. Todd, T. T. Hills, & T. W. Robbins (Eds.), *Cognitive search: Evolution, algorithms, and the brain* (pp. 221–235). MIT Press.
- Kenett, Y. N., & Faust, M. (2019). A semantic network cartography of the creative mind. *Trends in Cognitive Sciences*, 23(4), 271–274. <https://doi.org/10.1016/j.tics.2019.01.007>
- Kuznetsova, A., Brockhoff, P. B., & Christensen R. H. B. (2017). lmerTestpackage: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>.
- Lee, C. S., & Theriault, D. J. (2013). The cognitive underpinnings of creative thought: A latent variable analysis exploring the roles of intelligence and working memory in three creative thinking processes. *Intelligence*, 41(5), 306–320. <https://doi.org/10.1016/j.intell.2013.04.008>
- Long, J. A. (2019). *interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions*. R package version 1.1.0, <https://cran.r-project.org/package=interactions>
- Lu, J. G., Akinola, M., & Mason, M. F. (2017). “Switching On” creativity: Task switching can increase creativity by reducing cognitive fixation. *Organizational Behavior and Human Decision Processes*, 139, 63–75. <https://doi.org/10.1016/j.obhdp.2017.01.005>
- Maki, W. S., McKinley, L. N., & Thompson, A. G. (2004). Semantic distance norms computed from an electronic dictionary (WordNet). *Behavior Research Methods, Instruments, & Computers*, 36(3), 421–431. <https://doi.org/10.3758/BF03195590>
- Marko, M., Michalko, D., & Riečanský, I. (2019). Remote associates test: An empirical proof of concept. *Behavior Research Methods*, 51(6), 2700–2711. <https://doi.org/10.3758/s13428-018-1131-7>
- Marko, M., & Riečanský, I. (2021). The structure of semantic representation shapes controlled semantic retrieval. *Memory (Hove, England)*, 29(4), 538–546. <https://doi.org/10.1080/09658211.2021.1906905>
- Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, 69(3), 220–232. <https://doi.org/10.1037/h0048850>
- Nijstad, B. A., De Dreu, C. K. W., Rietzschel, E. F., & Baas, M. (2010). The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European Review of Social Psychology*, 21(1), 34–77. <https://doi.org/10.1080/10463281003765323>
- Oltețeanu, A. M., & Schultheis, H. (2019). What determines creative association? Revealing two factors which separately influence the creative process when solving the remote associates test. *The Journal of Creative Behavior*, 53(3), 389–395. <https://doi.org/10.1002/jocb.177>
- Schatz, J., Jones, S. J., & Laird, J. E. (2022). Modeling the remote associates test as retrievals from semantic memory. *Cognitive Science*, 46. Portico. <https://doi.org/10.1111/cogs.13145>
- Sio, U. N., Kotovsky, K., & Cagan, J. (2017). Interrupted: The roles of distributed effort and incubation in preventing fixation and generating problem solutions. *Memory & Cognition*, 45(4), 553–565. <https://doi.org/10.3758/s13421-016-0684-x>
- Sio, U. N., Kotovsky, K., & Cagan, J. (2022). Determinants of creative thinking: The effect of task characteristics in solving remote associate test problems. *Thinking & Reasoning*, 28(2), 163–192. <https://doi.org/10.1080/13546783.2021.1959400>
- Smith, K. A., Huber, D. E., & Vul, E. (2013). Multiply-constrained semantic search in the Remote Associates Test. *Cognition*, 128(1), 64–75. <https://doi.org/10.1016/j.cognition.2013.03.001>
- Smith, S. M., Gerkens, D. R., & Angello, G. (2017). Alternating incubation effects in the generation of category exemplars. *The Journal of Creative Behavior*, 51(2), 95–106. <https://doi.org/10.1002/jocb.88>.
- Storm, B. C., & Angello, G. (2010). Overcoming fixation. *Psychological Science*, 21(9), 1263–1265. <https://doi.org/10.1177/0956797610379864>
- Thompson, T. (1993). Remote associate problem sets in performance feedback paradigms. *Personality and Individual Differences*, 14(1), 11–14. [https://doi.org/10.1016/0191-8869\(93\)90169-4](https://doi.org/10.1016/0191-8869(93)90169-4)
- Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, 11(1), 138–146. <https://doi.org/10.1037/0894-4105.11.1.138>
- Valba, O., Gorsky, A., Nechaev, S., & Tamm, M. (2021). Analysis of English free association network reveals mechanisms of efficient solution of remote association tests. *PLoS One*, 16(4), e0248986. <https://doi.org/10.1371/journal.pone.0248986>
- Vul, E., & Pashler, H. (2007). Incubation benefits only after people have been misdirected. *Memory & Cognition*, 35(4), 701–710. <https://doi.org/10.3758/BF03193308>
- Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, 1(4), 442–452. <https://doi.org/10.1037/0278-7393.1.4.442>
- Wiley, J. (1998). Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory & Cognition*, 26(4), 716–730. <https://doi.org/10.3758/BF03211392>
- Wu, C.-L., Huang, S.-Y., Chen, P.-Z., & Chen, H.-C. (2020). A systematic review of creativity-related studies applying the remote associates test from 2000 to 2019. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.573432>
- Yaniv, I., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(2), 187–205. <https://doi.org/10.1037/0278-7393.13.2.187>
- Zeigarnik, B. (1938). On finished and unfinished tasks. In W. D. Ellis (Ed. & Trans.), *A source book of gestalt psychology* (pp. 300–314). New York, NY: Harcourt–Brace. (Original work published 1927). <https://doi.org/10.1037/11496-025>
- Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3–14. <https://doi.org/10.1111/j.2041-210X.2009.00001.x>