Title
Exploring mental representation with a memory matching game

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

Authors
Thibodeau, Paul
Levy, Isaac
de Lemos, Mikaela

Publication Date
2021

Exploring Mental Representation with a Memory Matching Game

Paul H. Thibodeau(paul.thibodeau@oberlin.edu)

Mikaela de Lemos (Mikaela.de.Lemos@oberlin.edu)

Isaac Levy (isaac.levy@oberlin.edu)
Department of Psychology, 120 W. Lorain Street
Oberlin, OH 44074 USA

Abstract

Games have been integral to the development of cognitive science. As experimental tasks, games can provide an ideal environment for studying questions about mental representation, memory, and strategic decision making. Here, we explore the potential of memory matching games as an experimental task by using a simple online version to conceptually replicate two classic effects from the cognitive scientific literature—the picture superiority effect (Paivio & Csapo, 1973) and the word length effect (Baddeley, Thomson, & Buchanan, 1975). We manipulate the *Item Format* of the game pieces (pictures vs. words) and their *Label Length* (short vs. long). As expected, we find a picture superiority effect. We do not find the predicted word length effect. We argue that the results of the study, along with several practical properties of the task, support the use of the game for cognitive scientific research.

Keywords: memory; dual coding theory; ecological cognition; imagery; cognitive style; games

Introduction

Imagine playing a simple memory matching game. A grid of 20 cards, consisting of 10 unique pairs, is laid out in front of you. Each card has a cartoon picture on the front, which is face down at the start of the game, and a generic logo on the back. A round of the game consists of turning two cards face up. If they show the same picture, they remain face up. If they show different pictures, they are turned back down, and you choose two new cards to turn over. Your task is to find all of the matching pictures. Figure 1 depicts an intermediate state of such a game.

Here, we explore the cognitive representations and processes involved in playing such a game. Our overarching goal is to explore the suitability of the game as an experimental task for studying questions related to working memory, mental representation, and strategic decision making. Like many other games that have played an important role in the history of cognitive scientific research (Gray, 2017), memory matching games have a number of properties that make them attractive as experimental paradigms. For example, they have well-defined goals and metrics, are relatively simple and engaging to play, yet afford researchers opportunities to manipulate and control factors in a way that can be useful for testing theoretically-motivated research questions.

For this initial exploration of the memory matching game, we used Dual Coding Theory (Paivio, 1991) to speculate on the mental representations and processes involved in playing the game. Dual Coding Theory posits that cognitive processes operate on two separate representational systems. One is sensorimotor and specialized for representing nonverbal information about objects and experiences. The other is symbolic and specialized for representing verbal information. According to the theory, the baseline system for certain cognitive tasks, like memory span tasks, is symbolic (verbal). When trying to store and maintain information in working memory, the default strategy is to covertly rehearse words to oneself. However, since the sensorimotor representational system is independent of the symbolic representational system, imagery can improve memory by creating a secondary trace of the information. That is, engaging both representational systems can have additive effects on memory recall, leading to a picture superiority effect (Paivio & Csapo, 1973).

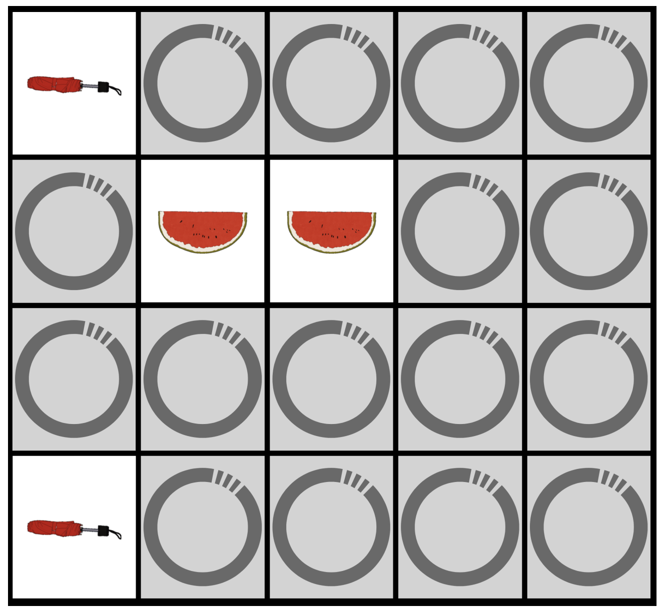

Figure 1: Two pairs have been found in this 10-item version of the memory matching game.

The framework of Dual Coding Theory suggests that the memory matching game may be similarly subserved, primarily, by verbal representations and a memory rehearsal process, and secondarily by modal representations (i.e., mental imagery). As a result, manipulating the *Item Format* of the cards — so that some participants play the game with pictures, while others play with words — should reveal a picture superiority effect. That is, participants should win the game faster (in fewer moves) when they playing with pictures than with words. This is one hypothesis that we test in the current experiment.

In addition, according to Dual Coding Theory, the label length of the items should also affect performance. Participants should be able to rehearse short labels (e.g., bagel, comb) faster, and therefore more accurately, then long labels (e.g., broccoli, calculator). As a result, manipulating the *Label Length* of the items should yield a word length effect (Baddeley et al., 1975), in which participants are faster to win the game when playing with items that have short labels, rather than long. Such a finding would support the view that people encode the information from the cards in a symbolic (verbal) format and use this mental representation to play the game, regardless of whether the the items on the cards are presented as pictures or words (Paivio, 1991).

Finally, in order to better understand how to use the game for research purposes, we also manipulated the *Number of Items* (cards) in the game, and we measured a variety of individual difference variables like cognitive style, age, and gender identity. The hypotheses and analytic plan for the study were pre-registered on the Open Science Framework.[1]

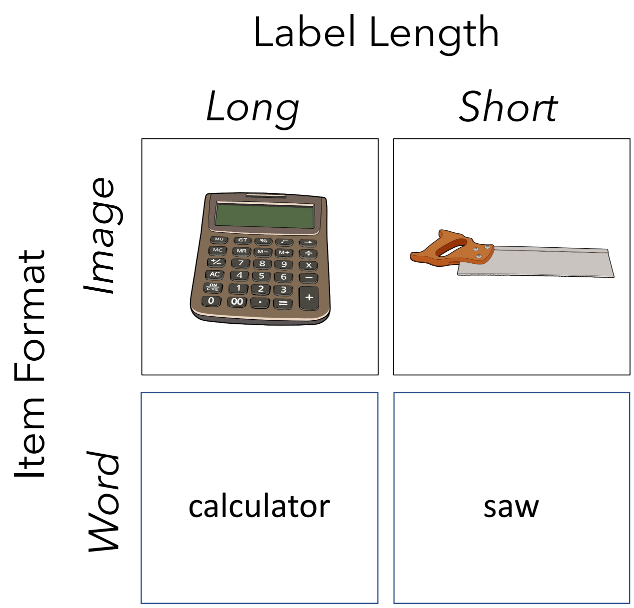

Figure 2: Examples of items used for the memory matching game. The three on the top row have long labels, while the three on the bottom row have short labels. Note that participants either saw images or labels (never both).

Methods

Participants

A total of 416 people participated in the study. Of these, 393 were recruited from Prolific (Palan & Schitter, 2018) and 23 were students. Data were not analyzed from seven participants who submitted incorrect completion codes on Prolific. The average age of participants was 29.2 ($SD = 10.6$). Roughly half (47%) were male and half were female (49%); the remainder of the sample (4%) either identified as non-binary or did not answer this question. Most participants (63%) identified as native English speakers.

We used prior work on the picture superiority effect (Hockley, 2008) and the word length effect (Avons, Wright, & Pammer, 1994) to estimate effect sizes for our manipulations of *Item Format* and *Label Length* (we used $f = .25$ for the picture superiority effect and $f = .5$ for the word length effect). Using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009), we estimated that 200 to 300 participants would be necessary for a suitably powered study ($1-\beta = .8$).

Our initial plan was to collect data from 200 participants on Prolific and from as many as students as we could during the fall semester. However, we made a mistake in our first wave of data collection on Prolific by forgetting to set a geographic restriction on participation. As a result, we collected a second sample of 200 participants on Prolific who lived in the United States (before analyzing the results of the first sample). Our dataset, therefore, includes two waves of data collected on Prolific—one that is international ($n = 194$) and one that is restricted to participants living in the US ($n = 199$)—along with a sample of college students ($n = 23$). We present the results of analyses conducted on the full sample, though we test for and report differences by group.

Materials

The words and images used for the game cards were selected from a stimulus set of clipart images of everyday objects, which were normed along several important dimensions (Saryazdi, Bannon, Rodrigues, Klammer, & Chambers, 2018) (see Figure 2 for examples). We identified 18 items with long names and 18 items with short names from the full set of 225 objects. We were careful to choose items that were similarly familiar, complex, and frequent, and that elicited consistent labels (see Table 1). Independent-samples t-tests revealed no differences by *Label Length* condition along any of the dimensions presented in Table 1. The objects came from a variety of common categories (e.g., office supplies, food, tools).

Participants were asked three questions about their experience playing the game. With regard to the game, participants were asked (a) “Have you played this kind of memory matching game before?” (Yes or No); (b) “How challenging was the memory game for you?” (5-point scale from Very Easy to Very Challenging); and (c) a free response question about why the game was challenging and what matching strategies they used.

Two measures of cognitive style were administered. One was the Object-Spatial Imagery and Verbal Questionnaire

[1] https :// osf .io / jnqx9 / ?view _only = 0084840e673841ccaaa42acaa9bec5e0

Table 1: The mean (and *SD*) of values for the stimulus items by *Label Length* condition along several target dimensions. The images were created and normed by (Saryazdi et al., 2018).

	Short	**Long**
Name agreement	96.4 (4.07)	96.7 (3.98)
Pict-name agreement	4.70 (0.12)	4.74 (0.15)
Familiarity	3.61 (0.66)	3.44 (0.70)
Complexity	3.22 (0.49)	3.35 (0.46)
Image agreement	4.15 (0.50)	4.14 (0.53)
SUBLexUS	11.9 (14.04)	11.0 (17.32)
WordNet	2.28 (3.03)	3.00 (5.67)
AoA	5.83 (1.29)	5.54 (1.45)

(OSIVQ; 10 items) (Blazhenkova & Kozhevnikov, 2009). The OSVIQ measures individual differences in *object imagery*, *spatial imagery*, and *verbal* cognitive styles. Of note, about 10% of participants did not fully complete this measure or the following measure of cognitive style, and, as a result, data from these participants are not included in the analyses of the variables. The internal consistency of the OSVIQ dimensions was acceptable, though somewhat low for the *verbal* dimension (Cronbach's α was .75, .77, and .63 for the *object*, *spatial*, and *verbal* components, respectively). There was a moderate negative correlation between the *object* and *spatial* components, $r(375) = -.205, p < .001$. There was no correlation between the *object* and *verbal* components, $r(373) = .021, p = .679$. There was a moderate positive correlation between the *spatial* and *verbal* components, $r(368) = .249, p < .001$.

The second measure of cognitive style that we included was the Thinking and Working Style Questionnaire (TWSQ; 10 items). The TWSQ is a uni-dimensional measure of how systematically (versus intuitively) people approach their work and creative endeavors (Sagiv, Arieli, Goldenberg, & Goldschmidt, 2010). This measure also had an acceptable level of internal consistency ($\alpha = .72$). It was somewhat correlated with the *spatial* dimension of the OSVIQ, $r(373) = .130, p = .014$. It was not correlated with the *object*, $r(376) = -.013, p = .808$, or *verbal* dimensions, $r(369) = -.08, p = .124$.

Finally, participants were asked about their age, gender identity, education level, English language fluency, and country of residence.

Procedure

The study was conducted online. It consisted of a survey that was implemented in Qualtrics and a memory matching game implemented in Java. After completing the informed consent, participants were randomly assigned to one of the twenty versions of the game, which they played three times. Participants played the same version of the game with the same set of cards each time.

Three aspects of the game were manipulated across participants. First, the items on the game cards were either presented as words or pictures (*Item Format*). Second, the label corresponding to items was either long or short (*Label Length*). Third, the *Number of Items* included in the game varied with five possible values: 8, 10, 12, 15, or 18 (for grid sizes of 16, 20, 24, 30, or 36 cards, respectively).

At the beginning of each game, participants saw a grid of cards arranged on the screen face down. On the face of each card was an item (either a word or image). Each item appeared on two of the cards in the grid. On the back of each card was a generic pattern, which was identical for all of the cards. The cards were shuffled and randomly distributed in the grid at the beginning of each game. On each trial of the game, participants chose two cards to turn over. If the pair showed matching words or images, the two cards remained face up. Otherwise, they were flipped back over. Participants continued playing until every pair was matched and all the cards were face up. We measured the number of moves it took for participants to complete the game and the amount of time it took to for participants to make each move. Our primary dependent measure of interest was the number of moves it took to win the game. We also test for an effect of the manipulation on the amount of time it takes to play the game.

Once participants finished playing the three games, they were asked questions about their experience playing the game and about themselves.

Of note, the study was set up so that participants could only complete it on a computer. If Qualtrics detected that a participant was using a tablet or smartphone, they were asked to use a computer instead and the session was terminated. Nevertheless, to check this design feature, participants were asked if they used "a computer or tablet or smartphone to play the game?" They were also asked, "If you were using a computer, did you use a mouse or trackpad to move the cursor on your screen?"

Results

General Properties of the Game

We first investigated some general properties of the game. We found a significant effect of *Number of Items* on the number of moves per item that it took to win the game, $\chi^2(1) = 529.6, p < .001$. With a smaller *Number of Items*, participants took fewer moves per item to win the game (see Table 2 and Figure 3). For example, participants completed the easiest version of the game (8 items) in 35 moves total, which was 4.4 moves per item; they completed the hardest version of the game (18 items) in 116 moves total, which was 6.4 moves per item.

We also found a significant effect of *Number of Items* on the time that participants took per move, $\chi^2(1) = 6.7, p = .009$. With a smaller *Number of Items*, participants took less time move (see Table 2 and Figure 3). On average, participants in the easiest version (8 items) took about 45 seconds to complete the game, which was 1.3 seconds per move; they

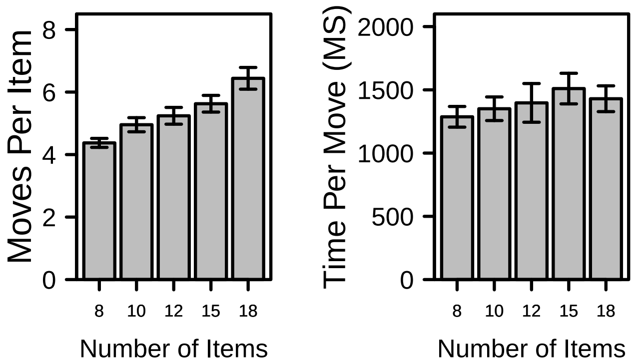

Figure 3: The number of moves (per item) and time (per item) to win the game as a function of the *Number of Items* in the game. Error bars are 95% Confidence Intervals.

Table 2: Means (and standard deviations) for the number of moves, the number of moves per item, the amount of time in seconds, and the amount of time in seconds per move that it took for participants to win the game, depending on the *Number of Items* included in the game.

Items	**Moves**	**Moves/Item**	**Time**	**Time/Move**
8	35.1 (8.0)	4.4 (0.7)	45 (19)	1.29 (.46)
10	49.8 (14.5)	5.0 (1.0)	66 (23)	1.34 (.41)
12	63.2 (17.9)	5.2 (1.1)	86 (36)	1.38 (.48)
15	84.7 (23.9)	5.6 (1.3)	126 (62)	1.52 (.64)
18	116.0 (37.8)	6.4 (1.7)	164 (77)	1.43 (.52)

completed the hardest version of the game (18 items) in two minutes and 44 seconds, which was 1.4 seconds per move.

We investigated the reliability of the task by testing whether the number of moves that it took a participant to win their first game was correlated with the number of moves it took to win their second. We did this separately for each level of *Number of Items* and found that, $r = .41, .35, .61, .50$, and .51, for versions of the game that included, 8, 10, 12, 15, and 18 items, respectively (ps $< .001$). Of note, the correlation between the second and third games (and first and third games) was similar (rs range from .26 to .62). These results suggest that the game reliably measures a relatively stable construct.

Not surprisingly, it took participants more moves to win the game every time they played, $\beta = .23, SE = .02, p < .001$. This is likely the result of proactive interference (Lustig, May, & Hasher, 2001) and suggests that researchers may want to limit the number of times that participants play the game with a given set of items. It may also be worthwhile to test the possibility of having participants play multiple versions of the game with different sets of items.

Finally, participants seemed to enjoy playing the game. Most (87%) reported that it was somewhat, but not especially, challenging ($\leq$ 3 on a 5-point scale), and most participants (87%) reported having played the game before.

Table 3: Results of linear mixed effects model predicting the number of moves it takes to complete the memory matching game as a function of *Label Length*, *Item Format*, and *Number of Items*. Asterisks indicate statistical significance at the $*p < .05$, $**p < .01$, and $***p < .001$ levels.

	Estimate (*SE*)	**t-value**
Intercept	-2.40 (.15)	-16.42***
Label Length: Short	-0.30 (.17)	-1.80
Item Format: Words	-0.31 (.17)	-1.86
Number of Items	0.18 (.01)	16.16***
Length: Short * Number	0.03 (.01)	2.30*
Format: Words * Number	0.04 (.01)	3.16**

Picture Superiority and Word Length

We used mixed effect linear regression models to test our primary hypotheses (Kuznetsova, Brockhoff, Christensen, et al., 2017). As predictors, we included *Item Format* (pictures or words), *Label Length* (short or long), and *Number of Items* (as a continuous variable) as fixed effects. Participants were treated as random, repeated measures, effects. The results of the model are shown in Table 3 and visualized in Figures 4 and 5.

Our first prediction was that we would find a picture superiority effect (Paivio & Csapo, 1973). That is, we expected participants to complete the game in fewer moves when they were playing with pictures, rather than words. Although we did not find the predicted main effect of *Item Format*, we did find an interaction between *Item Format* and the *Number of Items* that supports our hypothesis. As the *Number of Items* increased, participants took fewer moves to play the game with pictures than words (see Figure 4). Post-hoc testing revealed that participants were faster to play the game with pictures, as long as it included 15 or 18 items, $\beta = 0.41, SE = .12, p = .001$. There was no effect of *Item Format* on the game when it was played with less than 15 items (8, 10, or 12), $\beta = 0.04, SE = .11, p = .689$. This finding supports our general prediction that the game would be easier to play with pictures than words, and it suggests that games with smaller numbers of items may be too simple to test certain hypotheses about the representations and processes involved in the game.

Our second prediction was that we would find a word length effect (Baddeley et al., 1975). Using the framework of Dual Coding Theory (Paivio, 1991), we expected participants to store and use verbal codes of the items to play the game, even when the cards showed pictures. In turn, we expected people to be better able to rehearse and recall the locations of items with shorter labels (Baddeley, 2012). The results of the model did not support this prediction. We found no main effect of *Item Format*, and, although we did find an interaction between *Item Format* and the *Number of Cards*, it

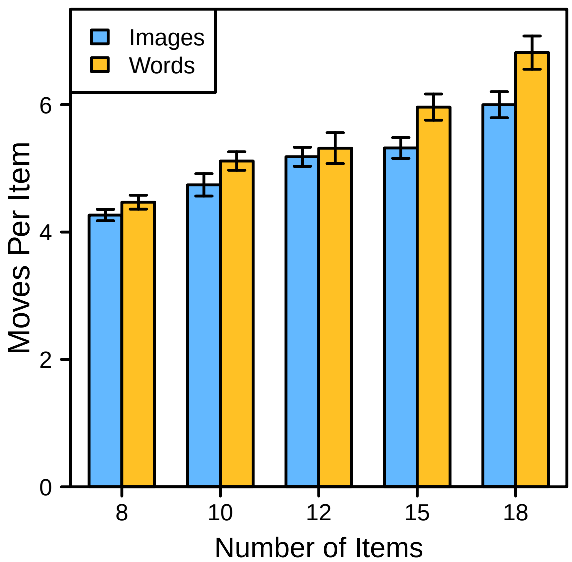

Figure 4: The average number of moves to win the memory matching game as a function of the *Number of Items* in the game and *Item Format* (images vs. words). Error bars reflect standard error.

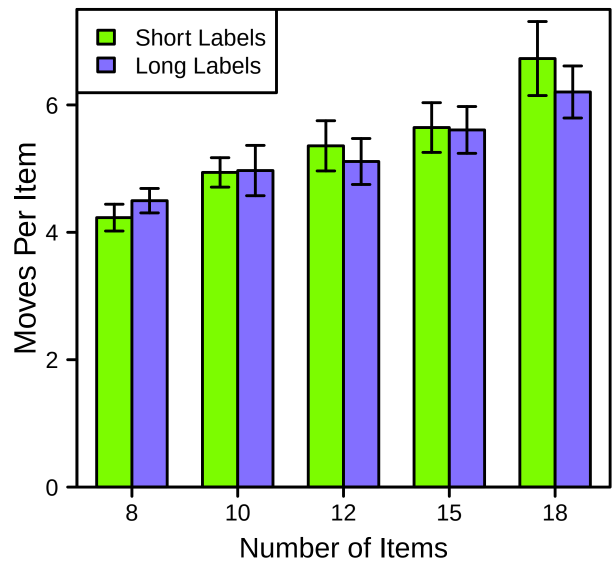

Figure 5: The average number of moves to win the memory matching game as a function of the *Number of Items* in the game and *Label Length* (short vs. long). Error bars reflect standard error.

was inconsistent with our hypothesis. It revealed that, as the *Number of Items* increased, people took *fewer* moves to play the game with items that had long, rather than short, labels (see Figure 5).

We conducted two additional exploratory analyses to further investigate the influence of *Label Length*. First, we tested whether *Label Length* only affected performance on the game when the game was played with words. We did this by adding an interaction term to the model, between *Label Length* and *Item Format*, and found that it was not statistically significant, $\chi^2(1) = 1.77, p = .183$. We also tested for a 3-way interaction between *Label Length*, *Item Format*, and the *Number of Items* to see if this relationship varied depending on the number of cards in the game. This analysis was not statistically significant, $\chi^2(1) = 2.99, p = .084$. These findings suggest that participants likely rely on strategies other then memory rehearsal to remember the locations of the items.

Effects by Sample

Before moving on to test for an influence of cognitive style, we looked at whether participants' geographic location moderated any of the effects of the manipulated variables. For these analyses we grouped the college students with the participants who were recruited from Prolific and lived in the US ($n = 222$), comparing their performance to participants who were recruited from Prolific and lived outside of the US ($n = 194$).

Using a series of linear mixed effects models, we found no differences in the average number of moves it took for participants in these groups to complete the game, $\chi^2(1) = 0.95, p = .330$. We also found no differences in how the manipulated variables affected participants in these two groups: *Number of Items*, $\chi^2(1) = 2.09, p = .149$; *Label Length*, $\chi^2(1) = 0.87, p = .351$; *Item Format*, $\chi^2(1) = 1.69, p = .193$. In sum, the different groups performed similarly on the game.

Cognitive Style

We included two measures of cognitive style in the study. One was the OSVIQ, which assess individual differences in *object imagery*, *spatial imagery*, and *verbal* cognitive styles (Blazhenkova & Kozhevnikov, 2009). We expected that the *object imagery* dimension of the OSVIQ would interact with *Item Format*, but the analysis revealed neither a interaction with *Item Format*, $\chi^2(1) = 0.40, p = .533$, nor a main effect of *object imagery*, $\chi^2(1) = 2.43, p = .119$. For *spatial imagery*, we expected to find a main effect, but the analysis did not support this prediction, $\chi^2(1) = 1.91, p = .167$. Finally, we expected an interaction between the *verbal* component and *Item Format*, but again the analysis failed to support our prediction, $\chi^2(1) = 0.26, p = .609$. In this case, however, we did find a main effect of *verbal* cognitive style, $\chi^2(1) = 14.19, p < .001$. More *verbal* participants tended to require more moves to complete the game, $\beta = .20, SE = .07, p = .007$.

The second measure of cognitive style that we included was the Thinking and Working Style Questionnaire (Sagiv et al., 2010), which measures the degree to which people approach problems systematically, as opposed to intuitively.

We expected that more systematic participants would perform better on the game, but we found no effect of this variable, $\chi^2(1) = 1.56, p = .212$.

Other Individual Differences

Finally, we conducted exploratory tests of other individual difference and background variables, including age, gender, education, previous experience with the game, and whether participants used a mouse or trackpad to play. The only one of these variables that significantly improved the model was gender, $\chi^2(1) = 6.83, p = .009$. Women took fewer moves to complete the game than men, $\beta = -0.12, SE = .05, p = .010$.

Discussion

Games have been central to the study of cognitive science since its inception (Gray, 2017). The goal of the current study was to explore the suitability of a simple memory matching game for empirical research on mental representation and working memory. We were drawn to the idea of using the game as an experimental task for a variety of reasons. Some were practical. For example, the game is easy to set up and run, and it is also simple and engaging to play. In addition, the game seems to both require and reward cognitive effort. When players are able to remember the items and their locations, they can win the game in fewer moves and less time. This means that the game has a built-in attentional check and performance incentive and distinguishes it from many classic tasks in behavioral research (e.g., recognition and recall tasks, memory span tasks, ratings tasks) (Buhrmester, Talaifar, & Gosling, 2018).

To investigate the task, we sought to conceptually replicate two classic findings from the cognitive scientific literature. First, we predicted and found a picture superiority effect. Participants completed the game faster when playing with pictures than words, as long as the game was sufficiently challenging (at least 15 items). Second, we predicted that we would find a word length effect, such that participants would be able to complete the game faster with items that had short, rather than long, labels. We did not find a word length effect.

We also investigated the relationship between the game and two measures of cognitive style. We found that more *verbal* participants, as measured by the OSVIQ, took more moves to win the game. We did not find effects of *object imagery* or *spatial imagery* on performance (as main effects or interaction effects), nor did we find an effect of a systematic working style, as measured by the Thinking and Working Style Questionnaire. More work is clearly need to understand how participants strategic decisions and cognitive tendencies influence how they play the game.

These findings provide mixed support for an account of the game grounded in Dual Coding Theory (Paivio, 1991). On the one hand, finding a picture superiority effect may indicate that two representational systems are at play—one that relies on imagery and one that relies on symbols. On this account, participants encoded both the pictures and words using a symbolic code; they additionally encoded an image when playing with pictures; and, as a result, performed better at the game when playing with pictures. However, not finding a word length effect raises questions about the degree to which playing the game relies on memory rehearsal and the articulatory loop (Baddeley & Hitch, 1974).

On the whole though, we we think the current study represents a promising avenue for future research, and we are excited to explore other research questions with a memory matching game. For example, we are interested in studying how the familiarity of the items and their similarity to each other affect how people play the game. Would it have taken longer for people to complete the game if the items were novel or if they had all come from the same category? Memory matching games provide an ideal environment for exploring a wide range of questions about cognition.

References

Avons, S., Wright, K., & Pammer, K. (1994). The word-length effect in probed and serial recall. *The Quarterly Journal of Experimental Psychology Section A*, *47*(1), 207–231.

Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annual review of psychology*, *63*, 1–29.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Elsevier.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior*, *14*(6), 575–589.

Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(5), 638–663.

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of amazon's mechanical turk, its rapid rise, and its effective use. *Perspectives on Psychological Science*, *13*(2), 149–154.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149–1160.

Gray, W. D. (2017). *Game-xp: Action games as experimental paradigms for cognitive science.* Wiley Online Library.

Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory & cognition*, *36*(7), 1351–1359.

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H., et al. (2017). lmertest package: tests in linear mixed effects models. *Journal of statistical software*, *82*(13), 1–26.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*(2), 199.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *45*(3), 255.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive psychology*, *5*(2), 176–206.

Palan, S., & Schitter, C. (2018). Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27.

Sagiv, L., Arieli, S., Goldenberg, J., & Goldschmidt, A. (2010). Structure and freedom in creativity: The interplay between externally imposed structure and personal cognitive style. *Journal of Organizational Behavior*, *31*(8), 1086–1110.

Saryazdi, R., Bannon, J., Rodrigues, A., Klammer, C., & Chambers, C. G. (2018). Picture perfect: A stimulus set of 225 pairs of matched clipart and photographic images normed by mechanical turk and laboratory participants. *Behavior Research Methods*, *50*(6), 2498–2510.