



# OPEN Working memory capacity estimates moderate value learning for outcome-irrelevant features

Ido Ben-Artzi<sup>1</sup>✉, Roy Luria<sup>1,2</sup> & Nitzan Shaha<sup>1,2</sup>

To establish accurate action-outcome associations in the environment, individuals must refrain from assigning value to outcome-irrelevant features. However, studies have largely ignored the role of attentional control processes on action value updating. In the current study, we examined the extent to which working memory—a system that can filter and block the processing of irrelevant information in one's mind—also filters outcome-irrelevant information during value-based learning. For this aim, 174 individuals completed a well-established working memory capacity measurement and a reinforcement learning task designed to estimate outcome-irrelevant learning. We replicated previous studies showing a group-level tendency to assign value to tasks' response keys, despite clear instructions and practice suggesting they are irrelevant to the prediction of monetary outcomes. Importantly, individuals with higher working memory capacity were less likely to assign value to the outcome-irrelevant response keys, thus suggesting a significant moderation effect of working memory capacity on outcome-irrelevant learning. We discuss the role of working memory processing on value-based learning through the lens of a cognitive control failure.

Throughout their lives, individuals infer causal associations between actions and outcomes as they experience and navigate dynamic and complex environments. Importantly, forming these associations requires cognitive flexibility and resources. Most environments are feature-rich, thus requiring the human agent to filter out irrelevant information when forming action-outcome associations in their minds<sup>1–3</sup>. For example, think of a child considering whether to eat an apple or a pear. The visual and tactile cues of each fruit should predict its taste, whereas its position on a table and the hand with which it is taken should be deemed irrelevant and thus not assigned value nor be considered in the child's choice of fruit.

Reinforcement learning studies have described in depth individuals' tendency to learn action-value associations in a trial-and-error manner<sup>4–6</sup>. However, most of these studies disregarded the influence of outcome-irrelevant information on value-based learning. Recently, new evidence has emerged suggesting individuals do attribute value to outcome-irrelevant features of an action, even when they hold certain and explicit knowledge that those features have no predictive value for the outcome<sup>7–9</sup>. Specifically, in recent studies participants were required to make choices to gain monetary rewards while some aspects of the task had no causal association with the delivery of the reward. Participants were encouraged using both explicit instructions and prolonged training to ignore outcome-irrelevant aspects when making value-based choices. Yet, the evidence clearly suggested individuals engaged in outcome-irrelevant learning, defined as a tendency to assign value to features of the environment that are known to the individual with high certainty as holding no causal association with an adjacent outcome<sup>7–9</sup>. A fundamental question thus remains regarding the cognitive mechanisms that allow the human agent to refrain from assigning value to features of the environment that are known as having no (or little) causal association with an outcome.

Outside the realm of reinforcement learning and value-based learning, attentional control studies have extensively examined the influence of irrelevant information on individuals' choices<sup>10–12</sup>. Specifically, a well-known attentional control system mostly studied outside the context of value-based learning is working memory. Working memory processing is known to reduce attention to irrelevant features including, for example, shape, color, or location of target stimuli<sup>13–15</sup>. Studies have systematically shown that even when observers have explicit and certain knowledge regarding the irrelevance of particular characteristics of a stimulus, attention regulation processes remain imperfect, resulting in a consistent influence of irrelevant information on decision-making<sup>16,17</sup>. A distinct feature of working memory is its limited capacity<sup>18–20</sup>. Previous studies have highlighted individual

<sup>1</sup>School of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel. <sup>2</sup>Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel. ✉email: idobenartzi@mail.tau.ac.il

differences in working memory capacity, demonstrating that lower capacity is associated with a reduced ability to filter out irrelevant information<sup>21–25</sup>.

Since working memory processing has been mostly studied outside the context of value-based decision-making<sup>16,26</sup>, it is unknown whether the same working memory mechanisms that filter task-irrelevant information (e.g., distracting stimuli), are also engaged in the filtering of outcome-irrelevant information during value-based learning<sup>5</sup>. Therefore, in the current study, we examined whether working memory resources are required to refrain from assigning credit to outcome-irrelevant information. Considering findings from the attentional control literature, we further examined whether individuals with low, versus high, working memory capacity will be less capable to withhold credit from being assigned to outcome-irrelevant information, especially when it is relevant to an ongoing task<sup>13,16,27</sup>. It should be noted that some studies addressed working memory in the context of reinforcement learning as a one-shot learning system that perfectly retains recent action-outcome associations<sup>28–31</sup>. However, here we mostly follow the definition of working memory as an attentional system that enables the filtering of irrelevant information<sup>12,22–25</sup>.

In this pre-registered study, individuals performed a card game in which they were asked to choose cards to gain a monetary reward (i.e., multiple armed bandit task). Our task design allowed us to disentangle and estimate value updating to the outcome-irrelevant task features, since only cards, but not their randomly assigned response keys, predicted the monetary outcome. Furthermore, we included visual working memory encoding and retrieval phases with varying degrees of load in between trials of the reinforcement learning card task in order to examine the influence of visual working memory load on outcome-irrelevant learning. Finally, we estimated individuals' working memory capacity using a well-known change-detection task<sup>20,32</sup>. Our results replicated previous studies by demonstrating a robust group-level effect of outcome-irrelevant learning. Moreover, we found significant individual differences such that only 55% of individuals demonstrated outcome-irrelevant learning. Importantly, working memory capacity showed a substantial moderation effect, such that reward had a lower influence on the selection of outcome-irrelevant response keys for high versus low-capacity individuals. However, to our surprise, the within-task working memory load manipulation did not influence outcome-irrelevant learning. We discuss these results by addressing the strengths and limitations of findings from reinforcement learning and attentional control literature, as well as the recent theoretical integration of the two perspectives.

## Methods

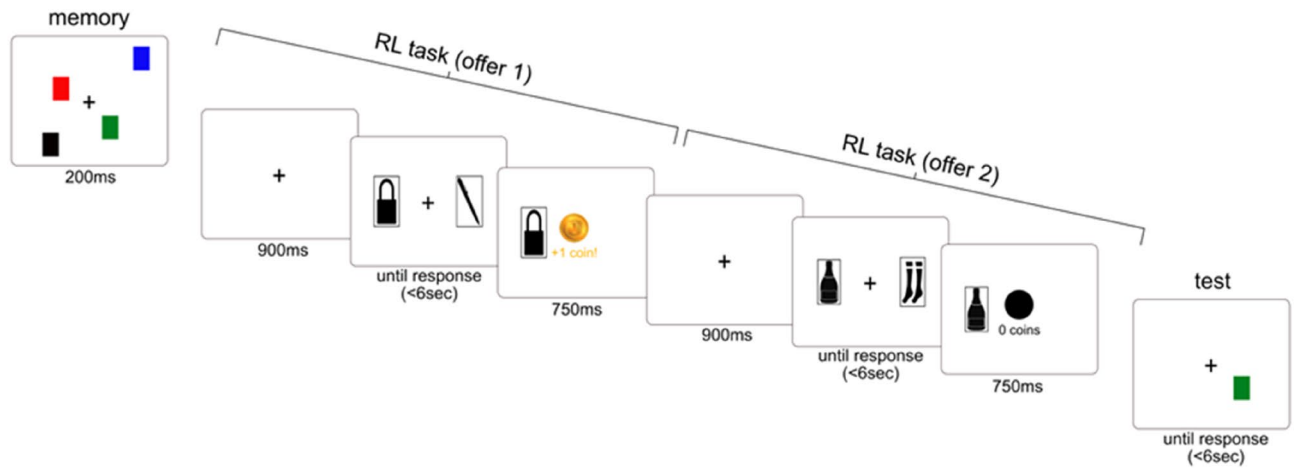
**Participants.** 174 Prolific workers (age mean = 27.1, range 18–49; 80 males, 93 females, 1 other) completed three online sessions across three consecutive days in return for monetary compensation (see SI). All participants reported normal or corrected vision, and no current or past psychiatric or neurological diagnosis (see SI). The study protocol was approved by the Research Ethics Council of Tel-Aviv University and all participants signed informed consent before participating in the study. All methods were carried out in accordance with relevant named guidelines and regulations.

**Procedure.** In the first session, 200 participants performed a working memory capacity measurement. In the second and third sessions 178/174 participants respectively, completed a reinforcement learning task under three working memory load conditions (i.e., no-load, low-load, and high-load).

**Reinforcement learning task.** Participants completed a reinforcement learning bandit task interleaved between the memory array and test array stages of a working memory task (Fig. 1). This design allowed us to examine outcome-irrelevant learning processes under different working memory load conditions. Trials started with a memory array stage where participants had to memorize a visual array (i.e., colored squares). Next, participants made two decisions on two sequential offers of a reinforcement learning bandit task where they had to choose one of two cards to gain monetary rewards. Finally, the trial ended with a test array stage, where participants were asked to report whether a newly presented colored square was part of the initial visual array or not. We will now describe the memory array, multi-armed bandit trials, and test array stages that were included in each trial in more detail:

The memory array stage of the working memory task included the brief presentation (200 ms) of a visual array (i.e., colored squares) which the participants had to remember until the test array stage. To manipulate working memory load, we included three types of visual arrays (see Fig. S1); (a) no-load (a fixed color for the entire block of trials), (b) low load (one colored square, with its color being randomly selected each trial by the computer), and (c) high load (four colored squares, with all colors being randomly selected by the computer each trial without replacement out of nine possible colors). The three conditions (i.e., no load, low and high load) were manipulated between blocks. The locations of the squares was further randomized in each trial by the computer (see SI for further details).

Reinforcement learning bandit task included two offers for each trial (interleaved between memory array and test array stages), which were designed to allow us to estimate credit assignment to the outcome-irrelevant response keys. In each trial, the computer allocated four cards (without return) to the two offers (i.e., first or second). Therefore, each offer allowed individuals to choose one of two cards, and these two cards were further randomly allocated to the right or left sides of the screen (see Fig. 1). Offers in the reinforcement learning bandit task started with a fixation (900 ms), followed by the presentation of two cards in the right/left location. Participants then chose a card freely using a right/left corresponding response-key press ('s' or 'k' keys in a QWERTY keyboard; until-response with a 6 s deadline). After making a choice the unchosen card disappeared and the chosen card remained to allow choice feedback (500 ms). Cards led probabilistically to reward (£0 or £1, play-pound coins; outcome presented for 750 ms) according to a true expected value (which slowly drifted across trials according to a predefined random walk; see SI). Participants were asked to do their best to make choices that



**Figure 1.** Trial sequence in the reinforcement learning task, which was performed under varying working memory loads. Participants were asked to first memorize a visual array (memory array stage), then make two choices across two card offers (reinforcement learning bandit task stage), and finally report whether a target was the same or different compared to the visual array that was memorized at the trial initiation (test array stage). Working memory load was manipulated between blocks by including in the memory stage either four random colored squares (high load), one random colored square (low load), or a fixed color square throughout the entire block, thus not requiring subjects to maintain the color in working memory during the task (no-load).

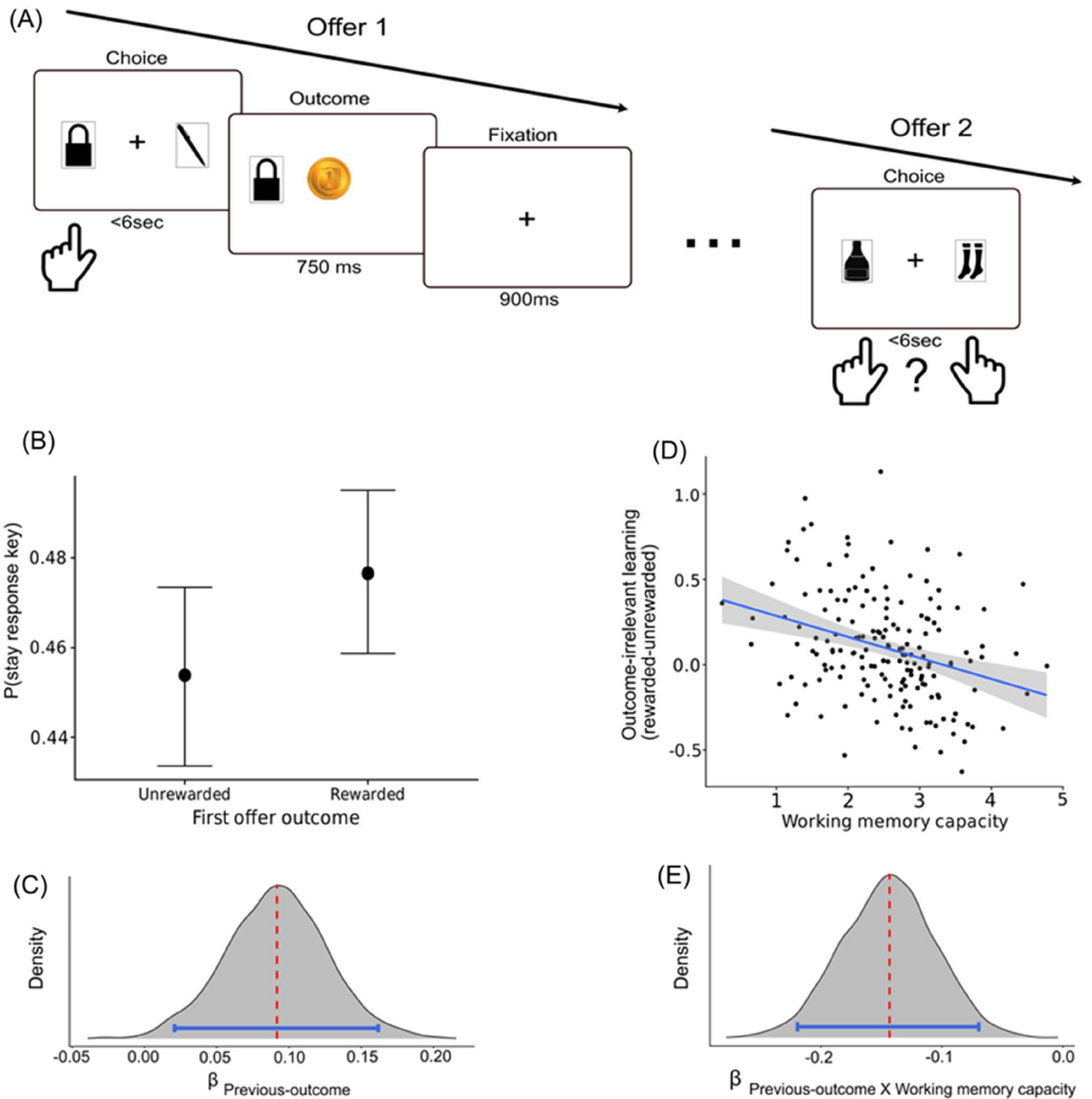
will maximize monetary return. To further ensure participants' motivation, we rewarded them with a monetary payment bonus at the end of the study according to their gains in the task. Importantly, only the cards predicted reward, but not locations or response-keys used to report cards' selection. This fact is important since it renders any credit assignment to the location/response key as outcome-irrelevant learning<sup>8,9</sup>. Participants were told explicitly during the instruction phase that only cards predicted reward and not the response-keys used for their selection (see SI). Before starting the task, participants were asked to complete a short quiz (9 questions long) to ensure they read and understood the instructions (see SI), which also included a specific question regarding which task features predict a monetary outcome. Participants had to show 100% accuracy in the quiz, and if they scored lower, they were prompted to the beginning of the instructions phase and were asked to retake the quiz.

Test array stage included a target screen in which one colored square was randomly selected by the computer in a color that was either the 'same' or 'different' color as the square that appeared in the same location in the memory array stage. Participants were asked to respond 'same' or 'different' by pressing 's' or 'k' response-keys (keyboard mapping was counterbalanced between participants; until-response with a 6 s deadline). Following the target screen, participants were presented with feedback indicating whether their response was correct or incorrect, which was then followed by a fixation screen (inter-trial interval, 500 ms fixation).

Across two sessions participants completed a total of 6 blocks of the reinforcement learning task. Each block included a different set of cards and had 50 trials. Participants received a monetary bonus at the end of the task based on their performance.

**Working memory capacity task.** To measure working memory capacity participants were asked to maintain and retrieve visual information<sup>20,32,33</sup>. On each trial, a visual array of colored squares appeared (set-size of 4 or 8 squares; see SI). Squares in each array had distinct colors and were evenly spread across the screen. Each trial started with the presentation of the squares array (i.e., memory array phase, 200 ms), followed by a fixation cross (i.e., retention phase, 900 ms) and then a target (i.e., test array phase, until-response with a 6 s deadline). The target screen included one colored square randomly selected by the computer in a color that was either the 'same' or 'different' color as the square that appeared in the same location in the memorized array. Participants were asked to respond 'same' or 'different' by pressing 's' or 'k' response-keys (keyboard mapping was counterbalanced between participants). Following the target screen participants saw feedback indicating whether their response was correct or incorrect, followed by a fixation screen (inter-trial interval, 500 ms fixation). Each participant completed 120 trials aimed to identify individual differences in working memory capacity. After data collection, we discovered that due to a technical error one location for target squares on the lower left side of the screen was never probed during the retrieval phase of the change detection task. Further analysis suggested this did not have an impact on our overall conclusion (see SI for further details).

**Estimating outcome-irrelevant learning.** To estimate outcome-irrelevant learning, we examined whether the outcome in the first offer (i.e. £0 vs. £1) affected response key selection in the second offer (see Fig. 2A). Specifically, we reasoned that credit will be assigned not only to the chosen card but also to a response key used for its selection [8,9]. This would mean that after a reward (i.e., £1) was obtained in the first offer, participants should be more likely to stay and choose the second offer with the same response key that was selected in the first one. However, if the choice in the first offer was unrewarded, participants should be more likely to switch their response key selection in the second offer. For example, assume that in the first offer the participant



**Figure 2.** (A) Outcome-irrelevant learning was estimated as a tendency to stay with the previous response key selection as a function of previous-outcome (i.e., £0 vs £1), despite explicit knowledge that response keys did not predict outcomes in the task. (B) Marginal means estimates for response key stay probability show a greater stay tendency after reward vs. unrewarded trial, reflecting group level outcome-irrelevant learning (error bars represent HDI<sub>95%</sub>). (C) Posterior distribution for the regression coefficient of previous-outcome effect on response key stay probability describing group level outcome-irrelevant learning (red/blue lines indicate median/HDI<sub>95%</sub>, respectively). (D) Outcome-irrelevant learning differed substantially between individuals and was moderated by working memory capacity estimates. (E) Posterior distribution for the regression coefficient indicating the moderation effect (i.e., paired interaction) between previous-outcome and working memory capacity (red line showing the posterior median and blue lines the HDI<sub>95%</sub>).

selected card A with a left response key, and observed a reward (£1). The left response key is now assumed to be more valuable in the individual’s mind, making it more likely that the left response key will be used in the second offer, where a different set of cards was offered. Therefore, outcome-irrelevant learning was assessed by estimating a regression parameter coefficient in a hierarchical Bayesian logistic regression, where repetition of response key selection from the first to the second offer (i.e., 0 vs. 1; Stay<sub>response key</sub>) was predicted using the outcome of the

first offer (i.e., previous outcome, £0 vs. £1; Fig. 2B). The positive influence of previous-outcome on  $\text{Stay}_{\text{response key}}$  was considered as evidence for value assignment to the response key (i.e., outcome-irrelevant learning).

**Estimating working memory capacity.** We calculated the detection accuracy in the visual working memory task (where reinforcement learning trials were not included) for each individual in each set-size condition (8 or 4 squares) to estimate working memory capacity. In each condition, capacity was calculated according to  $K = \text{set\_size} \cdot (2 \cdot \text{accuracy} - 1)$ , and the average between the two conditions was used as a measure of overall capacity<sup>34</sup>. In our single-probe task, this is equivalent to a calculation based on hit and false alarm ratios<sup>32</sup>.

**Regression analysis.** Hierarchical Bayesian regression analysis was performed using the ‘brms’ and ‘rstan’ R packages<sup>35,36</sup>. To examine posterior distributions we report 95% high-density interval (HDI) and probability of direction (pd). For all predictors, we used a pre-registered weakly informative prior, performed prior and posterior predictive checks, and verified that the conclusion did not change when using either wider or narrower priors (see SI). Bayes factors were calculated using the ‘bayestestR’ R package for both the null point estimate and the region of practical equivalence (ROPE)<sup>37</sup>. We used a ROPE interval in line with recent guidelines in Bayesian statistics recommending supplementing a null-point Bayes factor estimate, with a more conservative Bayes factor calculating the relative likelihood of the effect being within the ROPE (theoretically negligible) vs. outside the ROPE<sup>38</sup>. We set the ROPE interval to be  $-0.013$  to  $0.013$  based on the rationale that a previous-outcome effect size (Cohen’s  $d$ ) of  $0.007$  or lower on the probability to stay with the same response key should be considered negligible.

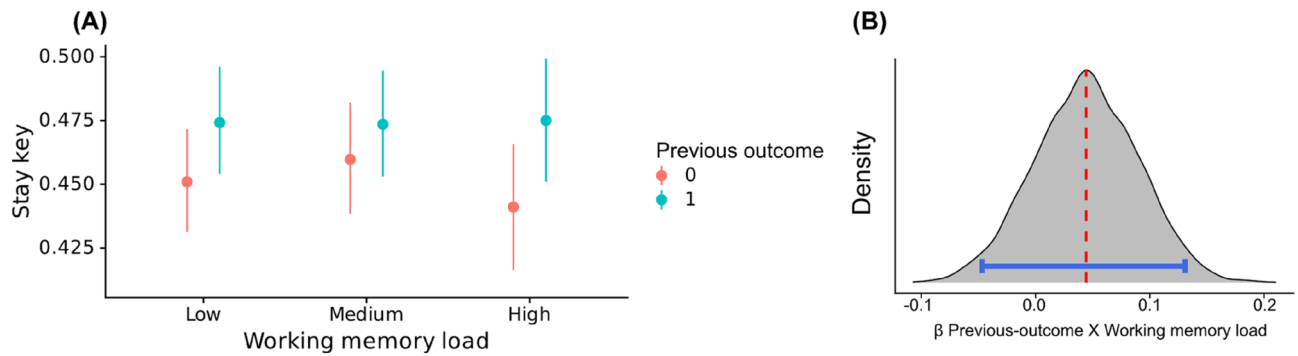
**Data treatment.** The first trial on each block and trials with implausibly quick reaction times ( $< 200$  ms), or exceptionally slow reaction times ( $> 4000$  ms) were omitted (5.52% of all trials). We omitted participants for the following reasons: (a) Participants with a lower than chance accuracy in the working memory capacity task with set size four (1 participant), (b) participants with more than 30% excluded trials in the reinforcement learning task (3 participants), and (c) participants who repeated the same response-key more than 80% of the trials (1 participant), in total 5 participants (age mean = 32, range 21–33; 3 males, 2 females) were excluded altogether.

## Results

Our main aim was to examine whether outcome-irrelevant learning changed as a function of individuals’ working memory capacity, and load manipulation (i.e., no-load, low, and high load). For this aim, we fitted a hierarchical Bayesian logistic regression. The dependent variable was the individuals’ tendency to repeat response key selection from the first to the second offer of the reinforcement learning bandit task (i.e., 0 vs. 1 for different or same, respectively;  $\text{Stay}_{\text{response key}}$ ). We examined three predictors including previous-outcome (i.e. £0 vs. £1, play pounds), working memory load (i.e., no-load, low load, and high load), individual’s working memory capacity, and their paired and triple interactions (see SI). Following recent guidelines on Bayesian regression analysis<sup>39</sup>, we start by performing nested model comparisons, allowing us to drop predictors that have a negligible contribution to the prediction of response key selection, and continue by examination of the posteriors of the best performing model. We assembled four nested models that included the following fixed effects: (a) Model 1 (full) previous-outcome (unrewarded vs. rewarded; describing outcome-irrelevant learning), working memory load manipulation, working memory capacity, and all paired and triple interaction as predictors. (b) Model 2 (working memory load) excluded working memory capacity parameters. (c) Model 3 (capacity) included working memory capacity parameters but excluded working memory load parameters, and finally, (d) Model 4 (null) included only previous-outcome as a predictor. Since we were interested in the population level (fixed effects) parameters, all models included a random effect of subjects on intercept, previous-outcome, working memory load, and previous-outcome  $\times$  working memory load parameters. Leave-one-out cross-validation model comparison<sup>40</sup> indicated that the best explanatory model was Model 3 (with previous-outcome, capacity, and their paired interaction). Specifically, stacking weights with uniform model priors showed 71% support for Model 3, 26% for Model 1, 3% for Model 2, and 0% for Model 4. We then continued to examine the posterior parameter’s distribution for the best fitting model, Model 3.

We found strong evidence for outcome-irrelevant learning, such that participants were more likely to stay with their response key selection after the first offer was rewarded (48%) vs. unrewarded (45%; posterior median = 0.09,  $\text{HDI}_{95\%}$  between 0.03 and 0.16; probability of direction (pd) 99.75%; 0% in ROPE ( $-0.013 - 0.013$ ) and Bayes Factor (BF) of 7.66 against the null and of 7.29 against the modified ROPE; Fig. 2C). This replicates previous findings of outcome-irrelevant learning in human individuals<sup>7–9</sup>. Importantly, we found support for our pre-registered exploratory hypothesis that individuals with low working memory capacity will demonstrate increased outcome-irrelevant learning compared to high capacity individuals such that the interaction between previous-outcome and capacity was negative (posterior median =  $-0.14$ ,  $\text{HDI}_{95\%}$  between  $-0.22$  and  $-0.07$ ; probability of direction (pd) 100%; 0% in ROPE ( $-0.013 - 0.013$ ) and Bayes Factor (BF) of 280 against the null and of 289 against the modified ROPE; Fig. 2E). Figure 2D illustrates a point-estimate of individual coefficients for previous-outcome (posterior mean) as a function of their estimated capacity. To provide the reader with a more intuitive effect-size of the association between working memory capacity and outcome-irrelevant learning, we calculated and found a correlation of  $r = -0.31$  ( $\text{HDI}_{95\%}$  between  $-0.44$  and  $-0.17$ ) between these two estimates. To affirm our results, we estimated working memory capacity using the change detection trials performed within the reinforcement learning task. We found that these estimates were positively correlated with working memory capacity estimates that were calculated using trials from the single change-detection task ( $r = 0.33$ ,  $\text{CI}_{95\%}$  between 0.2 and 0.45,  $\text{BF}_{10} = 4787$ ). Importantly, we found working memory capacity estimates drawn from the reinforcement learning task, to also negatively correlate with outcome-irrelevant learning ( $r = -0.26$ ,  $\text{CI}_{95\%}$  between  $-0.38$  and





**Figure 3.** Outcome-irrelevant learning as a function of working memory load manipulation. (A) Descriptive results show that outcome-irrelevant learning was overall similar under all three load conditions. (B) The posterior distribution for the interaction regression coefficient suggested no concise support for or against the existence of a moderation effect (i.e., paired interaction between previous-outcome and working memory load; red line showing the posterior median and blue lines the HDI<sub>95%</sub>).

–0.11,  $BF_{10} = 47$ ). Thus, overall, we found similar findings regardless of whether working memory capacity was measured using behavior from the single or the dual task.

We did not find support for an effect of our working memory load manipulation on outcome-irrelevant learning as similar outcome-irrelevant learning was found under no load (2.2%) vs. high load (3.4%) (posterior median = 0.04, OR = 1.04). The HDI<sub>95%</sub> was between –0.04 (OR = 0.96) and 0.12 (OR = 1.12) indicating that the most probable values of this interaction effect are small in magnitude<sup>41</sup>. Within these small effects, there is a probability of 84% of the effect being positive, and a probability of 16% of it being negligible (within the ROPE (between 0.013 and -0.013)). A BF analysis indicated similar ambiguity regarding the existence of the effect: there was no support for or against a non-zero effect ( $BF_{01}$  of 1.37) nor in favor of or against a non-negligible effect ( $BF_{01} = 1.42$ ) (see Fig. 3).

To assert participants complied with our instructions and performed both the working memory and reinforcement learning components above chance, we examined accuracy rates. For the working memory task component, we found 94%, 88%, and 67% accuracy rates for no-load, low-load, and high-load, respectively. For the reinforcement learning task component, we found that participants were able to choose the card with the higher true reward probability in 57% of trials (no difference was found between different working memory loads; see SI for further details). Thus, we conclude that performance was well above random in all parts of the task (see SI for Bayesian analysis of accuracy rates). We conducted further analyses to examine whether the moderating effect of working memory capacity is specific to outcome-irrelevant learning or whether it has a more general effect on learning. First, we estimated accuracy rates in cards' choices, with an accurate choice defined as selecting the card with a higher true reward probability. We then estimated and found that working memory capacity was not correlated with participants' accuracy in choosing the more rewarding card in the pair ( $r = 0.08$ ,  $CI_{95\%}$  between –0.07 and 0.22,  $BF_{01} = 3.2$  in favor of the null). Furthermore, we repeated our main regression analysis while including learning accuracy as a covariate. We found that the moderating effect of working memory capacity on outcome-irrelevant learning remained substantial (posterior median = –0.15, HDI<sub>95%</sub> between –0.24 and –0.06; probability of direction (pd) 100%).

To sum up, the results demonstrated support for outcome-irrelevant learning at the group level and supported the moderation of working memory capacity on outcome-irrelevant learning estimates. However, our results did not support our hypothesis for interaction between working memory load and previous-outcome (see SI and Table S1 for further information regarding evidence in favor of the null hypothesis for working memory load manipulation effect).

## Discussion

To engage in goal-directed behavior, human agents must form action-outcome associations within the environment<sup>42</sup>. Reinforcement learning studies have described, in-depth, the mechanisms involved in assigning credit to preceding actions when inferring causal associations<sup>43</sup>. However, little is known regarding how cognitive control processes might prevent individuals from assigning credit to aspects of the environment that are known—with high certainty—to be irrelevant to the observed outcome<sup>44</sup>. To the best of our knowledge, this preregistered study was the first to demonstrate a moderating role for a well-known cognitive control system, namely working memory capacity<sup>45</sup>, on the tendency to assign value to irrelevant features of the environment<sup>7–9</sup>.

Our results replicated, at the group level, a newly described phenomenon coined “outcome-irrelevant learning,” which refers to a tendency to assign credit to outcome-irrelevant elements of a task<sup>7–9</sup>. Specifically, in the current study, participants completed a reinforcement learning task in which they were asked to select cards to gain rewards. Card selection was indicated by a keypress; importantly, only the cards' visual features, but not the response key used for selection, predicted monetary outcome. Outcome-irrelevant learning was illustrated by an increased probability to stay with a response key selection when it was followed by a reward compared to an unrewarded outcome. Note that this result does not reflect a general tendency to repeat one's previous actions, and thus should not be considered as an example of a perseveration tendency<sup>46,47</sup>. Importantly, outcome-irrelevant

learning is different from perseveration as it is a reward-dependent effect, while response-key perseveration effects address a general tendency to repeat one's choices regardless of reward or value. An individual showing a perseveration effect would tend to repeat their response-key choices regardless of the obtained reward. However, an outcome-irrelevant learner would show a lower/greater probability to repeat the previous response-key selection after unrewarded/rewarded trials, respectively. Thus, the main difference between perseveration and outcome-irrelevant learning is that the former is unrelated to reward, while the latter is. Note that both can co-occur in a single individual. In fact, when we predict response-key stay probability our intercept estimates correspond to overall response-key perseveration (tendency to stick/switch to a response key regardless of reward), while the effect of previous-outcome reflects outcome-irrelevant learning (increased/reduced tendency to stick with response key-selection following rewarded/unrewarded outcomes). In that sense, these effects (intercept and slope) are independent by design. Finally, an important aspect of outcome-irrelevant learning is that the effect was demonstrated despite both explicit instructions and practice trials indicating that the response keys were irrelevant to the monetary outcome. Therefore, unlike previous reinforcement learning studies that focused on an ability to learn which features are relevant to an outcome in an unknown environment<sup>2,3,6,48</sup>, outcome-irrelevant learning reflects a fundamental human difficulty to avoid assigning credit to features that are known, with high certainty, as not being causally related to obtaining a certain outcome<sup>7-9</sup>.

An important, and largely unanswered, question regards the underlying cognitive mechanisms of outcome-irrelevant learning. We argue that outcome-irrelevant learning reflects a control failure, whereby human agents find it difficult to avoid assigning credit to representations that are actively held in working memory<sup>49-51</sup>. In the current study, individuals were faced with duality, such that response keys are both task-relevant (thus, must be actively held in mind) and outcome-irrelevant (thus, should be disregarded and not be assigned value regardless of the outcome). One hypothesis for the mechanism allowing individuals to modulate outcome-irrelevant learning might be that control is exerted during value updating, thereby resulting in task-relevant, yet outcome-irrelevant, information to maintain a neutral value. Another possibility might be that response keys are assigned a value, but that cognitive control processes regulate their involvement during the decision-making process. Both explanations lead to the similar conclusion that cognitive resources are required for the mental shielding of outcome-relevant from irrelevant information during value updating. If outcome-irrelevant learning does indeed reflect a cognitive control failure, it should increase when attentional control resources are low.

Our main finding indicated that individuals with low, as compared to high, working memory processing capacity showed increased outcome-irrelevant learning. Working memory capacity is a well-studied phenomenon in the cognitive control literature, and it has been shown to predict individuals' ability to avoid distraction from irrelevant information<sup>20,21,25,32,52</sup>. Individual differences in working memory capacity were previously found to predict interference from irrelevant information across multiple tasks such as the Stroop, dichotic listening, and anti-saccade tasks<sup>16,22,24</sup>. Our main finding extends this literature by demonstrating that working memory capacity also predicts value-based attentional control. This finding, although correlative in nature, is in line with the speculation that lower working memory capacity leads to difficulty in holding distinct representations of the task set, which consequently leads value to be assigned to representations that are activated at the time of selection. An alternative notion of working memory may focus on its role as short-term storage<sup>45</sup>. Accordingly, a potential contrary hypothesis relies on the idea that high working memory individuals are able to hold more information actively in mind, thus allowing them to encode a richer representation of their actions. In such a case, it might be that individuals with higher working memory capacity would show increased, rather than decreased outcome-irrelevant learning. However, our results show the opposite direction. It should be noted that, to date, most reinforcement learning studies have not directly addressed the influence of working memory capacity on credit assignment. Studies that have explored the role of working memory in credit assignment mostly manipulated set-size (e.g., number of arms in a multi-armed bandit task), thus primarily demonstrating an effect of decay and noise on the representations of relevant information<sup>28-31</sup>. Our findings suggest that working memory capacity, as measured by a visual working memory task, can predict the extent of inaccurate credit assignment.

The finding that working memory capacity, but not load manipulation, was associated with outcome-irrelevant learning should be discussed. One possibility might be that we did not include a strong enough load manipulation; however, the no-load (zero squares to remember) and the high-load (four squares to remember) conditions that we used should reflect the extreme ends of the load manipulation<sup>53</sup>. Another explanation might be that participants sacrificed effort on one task for the other, however, overall accuracy rates in both tasks were above chance, suggesting that participants did in fact devote attention and effort to both tasks. We, therefore, suggest a theoretical explanation for our finding, such that the visual stimuli (working memory task component) and the cards' values (reinforcement learning task component) did not tap into the same working memory buffer, thus leading to the observed null effect. In this respect, the discrepancy in modalities might be the reason that loading working memory did not influence outcome-irrelevant learning. Specifically, we loaded working memory with visual information, while outcome-irrelevant information was modulated in the spatial-motor domain. Indeed, although these findings are still in debate, previous studies show that accuracy in a visual working memory task was more strongly disrupted by a visual rather than a spatial interference and vice versa<sup>54-56</sup>. Therefore, future studies are required to shed light on whether outcome-irrelevant learning is also more domain-specific than we assumed here. In a somewhat similar manner, and in line with the activity-silent memory theory<sup>57-59</sup>, information that is not immediately relevant for a present task is maintained with minimal neural activity<sup>60</sup>. Therefore, it could be that visual information from the working memory task component was maintained in activity-silent memory during the performance of the reinforcement learning task component<sup>61</sup>. Another reason for the failure to find an effect of working memory manipulation might be related to the way we defined working memory. Specifically, while we followed the definition of working memory as an active attentional system that filters irrelevant information<sup>24</sup>, several studies have also demonstrated dissociations between attention and working memory<sup>62-65</sup>. Therefore, it might be that the filtering process that is hypothesized to

operate during credit assignment is associated with an attentional filtering process that is not directly controlled by the working memory system. In such a case, it is reasonable that working memory and attentional filtering abilities would correlate between individuals, which would lead to the observed association between working memory and outcome-irrelevant learning. However, this would also mean that working memory manipulation should not show a direct effect on outcome-irrelevant learning, which is also in line with the observed results.

Finally, we note that the outcome-irrelevant learning demonstrated in this study was primarily discussed in regard to the assignment of credit to response keys, but it can also be interpreted in terms of value assignment to spatial location (since the card in the right/left location was always selected using a right/left response key, respectively). The present study was designed to determine the effects of working memory capacity and load manipulation on credit assignment to outcome-irrelevant features of the environment. The results of this investigation showed that individual differences in working memory capacity predict outcome-irrelevant learning. However, it should be noted that we only examined individual differences in visual working memory, and it remains to be investigated whether these differences will also be apparent in different modules. We conclude that resource-demanding cognitive control is exerted during credit assignment, in such a way that credit can be accurately assigned to outcome-relevant actions. These findings contribute to our understanding of how credit assignment is regulated and provide a basis for future cognitive and clinical research.

## Data availability

The data, analysis scripts, and the task can be found at the following address <https://osf.io/rfeqx/>. Preregistration on the open science framework (OSF) website is available at <https://osf.io/6cz29>.

Received: 26 June 2022; Accepted: 4 October 2022

Published online: 16 November 2022

## References

- Collins, A. G. E. & Shenhav, A. Advances in modeling learning and decision-making in neuroscience. *Neuropsychopharmacology* **47**, 104–118 (2022).
- Niv, Y. Learning task-state representations. *Nat. Neurosci.* **22**, 1544–1553 (2019).
- Wilson, R. C. & Niv, Y. Inferring relevance in a changing world. *Front. Hum. Neurosci.* **5**, 189 (2012).
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215 (2011).
- Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80**, 312–325 (2013).
- Niv, Y. *et al.* Reinforcement learning in multidimensional environments relies on attention mechanisms. *J. Neurosci.* **35**, 8145–8157 (2015).
- Rmus, M., Zou, A. & Collins, A. G. E. Plucking a string or playing a G?. *bioRxiv* <https://doi.org/10.1101/2021.08.25.457707>(2021).
- Shahar, N. *et al.* Credit assignment to state-independent task representations and its relationship with model-based decision making. *Proc. Natl. Acad. Sci.* **116**, 15871–15876 (2019).
- Shahar, N. *et al.* Assigning the right credit to the wrong action: compulsivity in the general population is associated with augmented outcome-irrelevant value-based learning. *Transl. Psychiatry* **11**, 1–9 (2021).
- Folk, C. L., Remington, R. W. & Wright, J. H. The structure of attentional control: contingent attentional capture by apparent motion, abrupt onset, and color. *J. Exp. Psychol. Hum. Percept. Perform.* **20**, 317–329 (1994).
- Theeuwes, J. & Burger, R. Attentional control during visual search: The effect of irrelevant singletons. *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 1342–1353 (1998).
- Zanto, T. P. & Gazzaley, A. Neural suppression of irrelevant information underlies optimal working memory performance. *J. Neurosci.* **29**, 3059–3066 (2009).
- Engle, R. W. & Kane, M. J. The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Learn. Motiv. Adv. Res. Theory* **44**, 145–199 (2004).
- Gazzaley, A. & Nobre, A. C. Top-down modulation: Bridging selective attention and working memory. *Trends Cogn. Sci.* **16**, 129–135 (2012).
- Luria, R. & Vogel, E. K. Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia* **49**, 1632–1639 (2011).
- Kane, M. J. & Engle, R. W. Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *J. Exp. Psychol. Gen.* **132**, 47–70 (2003).
- Meier, M. E. & Kane, M. J. Working memory capacity and Stroop interference: Global versus local indices of executive control. *J. Exp. Psychol. Learn. Mem. Cogn.* **39**, 748–759 (2013).
- Conway, A. R. A., Kane, M. J. & Engle, R. W. Working memory capacity and its relation to general intelligence. *Trends Cogn. Sci.* **7**, 547–552 (2003).
- Cowan, N. The magical mystery four: How is working memory capacity limited, and why?. *Curr. Dir. Psychol. Sci.* **19**, 51–57 (2010).
- Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281 (1997).
- Allon, A. S. & Luria, R. Compensation mechanisms that improve distractor filtering are short-lived. *Cognition* **164**, 74–86 (2017).
- Conway, A. R. A., Cowan, N. & Bunting, M. F. The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychon. Bull. Rev.* **8**, 331–335 (2001).
- Engle, R. W. Working memory capacity as executive attention. *Curr. Dir. Psychol. Sci.* **11**, 19–23 (2002).
- Kane, M. J., Bleckley, M. K., Conway, A. R. & Engle, R. W. A controlled-attention view of working-memory capacity. *J. Exp. Psychol. Gen.* **130**, 169–183 (2001).
- Vogel, E. K., McCollough, A. W. & Machizawa, M. G. Neural measures reveal individual differences in controlling access to working memory. *Nature* **438**, 500–503 (2005).
- Oberauer, K. In *Psychology of Learning* **51**, 45–100 (Academic Press, Berlin, 2009).
- de Fockert, J. W., Rees, G., Frith, C. D. & Lavie, N. The role of working memory in visual selective attention. *Science* <https://doi.org/10.1126/science.1056496> (2001).
- Collins, A. G. E., Albrecht, M. A., Waltz, J. A., Gold, J. M. & Frank, M. J. Interactions among working memory, reinforcement learning, and effort in value-based choice: A new paradigm and selective deficits in schizophrenia. *Biol. Psychiatry* **82**, 431–439 (2017).
- Collins, A. G. E., Cuijlo, B., Frank, M. J. & Badre, D. Working memory load strengthens reward prediction errors. *J. Neurosci.* **37**, 4332–4342 (2017).



30. Collins, A. G. E. & Frank, M. J. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *Eur. J. Neurosci.* **35**, 1024–1035 (2012).
31. Yoo, A. & Collins, A. How working memory and reinforcement learning are intertwined: A cognitive, neural, and computational perspective. *J. Cogn. Neurosci.* [https://doi.org/10.1162/jocn\\_a\\_01808](https://doi.org/10.1162/jocn_a_01808) (2021).
32. Balaban, H., Fukuda, K. & Luria, R. What can half a million change detection trials tell us about visual working memory?. *Cognition* **191**, 103984 (2019).
33. Pashler, H. Familiarity and visual change detection. *Percept. Psychophys.* **44**, 369 (1988).
34. Cowan, N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav. Brain Sci.* **24**, 87–114 (2001).
35. Bürkner, P.-C. Advanced Bayesian multilevel modeling with the R package brms. *R J.* **10**, 395 (2018).
36. Carpenter, B. *et al.* Stan: A probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).
37. Makowski, D., Ben-Shachar, M. S. & Lüdtke, D. bayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework. *J. Open Sour. Softw.* **4**, 1541 (2019).
38. Makowski, D., Ben-Shachar, M. S., Chen, S. H. A. & Lüdtke, D. Indices of effect existence and significance in the bayesian framework. *Front. Psychol.* **10**, 2767 (2019).
39. van den Bergh, D. *et al.* A tutorial on Bayesian multi-model linear regression with BAS and JASP. *Behav. Res. Methods* **53**, 2351–2371 (2021).
40. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
41. Chen, H., Cohen, P. & Chen, S. How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Commun. Stat. Simul. Comput.* **39**, 860–864 (2010).
42. Gershman, S. J., Norman, K. A. & Niv, Y. Discovering latent causes in reinforcement learning. *Curr. Opin. Behav. Sci.* **5**, 43–50 (2015).
43. Sutton, R. S. & Barto, A. G. *Introduction to Reinforcement Learning* (MIT Press, Cambridge, 1998).
44. Rmuis, M., McDougale, S. D. & Collins, A. G. The role of executive function in shaping reinforcement learning. *Curr. Opin. Behav. Sci.* **38**, 66–73 (2021).
45. Cowan, N. The many faces of working memory and short-term storage. *Psychon. Bull. Rev.* **24**, 1158–1170 (2017).
46. Miller, K. J., Shenhav, A. & Ludvig, E. A. Habits without values. *Psychol. Rev.* **126**, 292–311 (2019).
47. Wood, W. & Rünger, D. Psychology of habit. *Annu. Rev. Psychol.* **67**, 289–314 (2016).
48. Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V. & Niv, Y. Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron* **93**, 451–463 (2017).
49. Dreisbach, G. & Fröber, K. On how to be flexible (or not): Modulation of the stability-flexibility balance. *Curr. Dir. Psychol. Sci.* **28**, 3–9 (2019).
50. Dreisbach, G. & Haider, H. That's what task sets are for: Shielding against irrelevant information. *Psychol. Res.* **72**, 355–361 (2008).
51. Dreisbach, G. & Haider, H. How task representations guide attention: Further evidence for the shielding function of task sets. *J. Exp. Psychol. Learn. Mem. Cogn.* **35**, 477–486 (2009).
52. Engle, R. W. In *The Nature of Remembering: Essays in Honor of Robert G. Crowder* 297–314 (American Psychological Association, 2001). <https://doi.org/10.1037/10394-016>.
53. Woodman, G. F. & Luck, S. J. Visual search is slowed when visuospatial working memory is occupied. *Psychon. Bull. Rev.* **11**, 269–274 (2004).
54. Darling, S., Della Sala, S. & Logie, R. H. Behavioural evidence for separating components within visuo-spatial working memory. *Cogn. Process.* **8**, 175–181 (2007).
55. Klauer, K. C. & Zhao, Z. Double dissociations in visual and spatial short-term memory. *J. Exp. Psychol. Gen.* **133**, 355–381 (2004).
56. Vergauwe, E., Barrouillet, P. & Camos, V. Visual and spatial working memory are not that dissociated after all: A time-based resource-sharing account. *J. Exp. Psychol. Learn. Mem. Cogn.* **35**, 1012–1028 (2009).
57. Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K. & Postle, B. R. Neural evidence for a distinction between short-term memory and the focus of attention. *J. Cogn. Neurosci.* **24**, 61–79 (2012).
58. Rose, N. S. *et al.* Reactivation of latent working memories with transcranial magnetic stimulation. *Science* <https://doi.org/10.1126/science.aah7011> (2016).
59. Stokes, M. G. 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* **19**, 394–405 (2015).
60. Stokes, M. G. *et al.* Dynamic coding for cognitive control in prefrontal cortex. *Neuron* **78**, 364–375 (2013).
61. Kreither, J., Papaioannou, O. & Luck, S. J. Active working memory and simple cognitive operations. *J. Cogn. Neurosci.* **34**, 313–331 (2022).
62. Bae, G.-Y. & Luck, S. J. Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials. *J. Neurosci.* **38**, 409–422 (2018).
63. Hakim, N., Adam, K. C. S., Gunseli, E., Awh, E. & Vogel, E. K. Dissecting the neural focus of attention reveals distinct processes for spatial attention and object-based storage in visual working memory. *Psychol. Sci.* **30**, 526–540 (2019).
64. Hollingworth, A. *Vis Mem* (Routledge, 2022).
65. Hollingworth, A. & Maxcey-Richard, A. M. Selective maintenance in visual working memory does not require sustained visual attention. *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 1047–1058 (2013).

## Acknowledgements

We thank Mattan Ben-Shachar for his help in Bayesian regression analysis. This work was funded by an Israeli Science Foundation grant for the second author (Grant No. 862/17) and the last author (Grant No. 2536/20).

## Author contributions

I.B.A, R.L and N.S designed the research; I.B.A and N.S analyzed the data; I.B.A, R.L, and N.S interpreted the data and wrote the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-21832-x>.

**Correspondence** and requests for materials should be addressed to I.B.-A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022