Visual Cognition

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pvis20

Relying on the external world: Individuals variably use low- and medium-loading, but rarely high-loading, strategies when engaging visual working memory

S. Böing, B. de Zwart, A. F. Ten Brink, T. C. W. Nijboer & S. Van der Stigchel

To cite this article: S. Böing, B. de Zwart, A. F. Ten Brink, T. C. W. Nijboer & S. Van der Stigchel (02 Apr 2025): Relying on the external world: Individuals variably use low- and medium-loading, but rarely high-loading, strategies when engaging visual working memory, Visual Cognition, DOI: 10.1080/13506285.2025.2484523

To link to this article: https://doi.org/10.1080/13506285.2025.2484523

VISUAL COGNITION
https://doi.org/10.1080/13506285.2025.2484523

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Relying on the external world: Individuals variably use low- and medium-loading, but rarely high-loading, strategies when engaging visual working memory

S. Böing, B. de Zwart, A. F. Ten Brink, T. C. W. Nijboer and S. Van der Stigchel

Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht, the Netherlands

ABSTRACT
In naturalistic environments, people typically rely on external sampling rather than fully using their visual working memory capacity. However, when sampling becomes costly, people memorize more (i.e., loading). To investigate individual differences in sampling versus loading strategy, participants (n=88) performed a copying task under low-cost (immediate accessibility) and high-cost (delayed accessibility) sampling conditions. Participants were categorized as low-loaders (sampling >1 per item), medium-loaders (loading ≥1 per inspection), and high-loaders (loading ≥3 per inspection). Both sampling cost and prior experience (low-cost first versus high-cost first) affected sampling frequency and category. Crucially, low- and medium-loading strategies were common, but individuals seldom exhibited a high-loading strategy that approached working memory capacity limits. Despite individual variation in their preferred strategy, participants flexibly adapted their sampling frequency to task demands without affecting performance. This suggests that while individuals show distinct working memory strategies, they can adjust these flexibly, balancing effort, goals, and prior experience.

ARTICLE HISTORY
Received 8 August 2024
Accepted 21 March 2025

KEYWORDS
Offloading; memory strategies; individual differences; sampling; copy task

Introduction

Research on visual working memory has traditionally been concerned with estimating its maximum capacity (e.g., Luck & Vogel, 2013). To optimize performance on a capacity test, people should try to memorize as much information as possible, which means they should use their memory capacity to the fullest. Although these capacity tasks have provided fundamental insights into the mechanisms of working memory (Luck & Vogel, 2013; Ma et al., 2014) and have been useful in compiling cognitive profiles and subsequently guide diagnoses in clinical care (e.g., Corsi, 1972), they fall short in mimicking actual working memory usage in daily life, and do not grasp the wide variety of working memory strategies that individuals might use to support task performance.

The influence of the environmental context on the use of memory has received increased attention in the fields of cognitive engineering and human-system interactions (e.g., Gray & Fu, 2004; Morgan et al., 2009; Waldron et al., 2007), working memory (Ballard et al., 1995; Draschkow et al., 2021; Droll & Hayhoe, 2008; Grinschgl et al., 2021; Hoogerbrugge et al., 2024; Kvitelashvili & Kessler, 2024; Melnik et al., 2018; Meyerhoff et al., 2021; Risko & Gilbert, 2016; Sahakian et al., 2023; Somai et al., 2020; Van der Stigchel, 2020), and neuropsychology (Böing et al., 2023, 2025). Interestingly, results in these fields indicate that, even when the demands of the environmental context encourage people to shift towards memorization, people still do not employ their full capacity, and often keep relying on the outside world. To not fully tax memory, people may load up less than their maximum capacity or may use a cognitive offloading strategy (e.g., writing things down, creating cues as reminder; Ballard et al., 1995; Böing et al., 2023, 2025; Draschkow et al., 2021; Gray et al., 2006; Meyerhoff et al., 2021; Morrison & Richmond, 2020; Risko & Dunn, 2015; Sahakian et al., 2023; Somai et al., 2020). To illustrate these findings in the context of a natural environment, consider assembling a piece of furniture. The instruction leaflet is available for one to rely on.

CONTACT S. Böing s.boing@uu.nl Experimental Psychology, Helmholtz Institute, Utrecht University, Heidelberglaan 1, 3584CS Utrecht, the Netherlands
Supplemental data for this article can be accessed online at https://doi.org/10.1080/13506285.2025.2484523.

Hence, there is no need to memorize all the screws, their rotation, and their desired location at once because one can inspect and reinspect the building steps depicted on the leaflet as often as desired. In such cases, one can afford to rely on the external world by (re)inspecting information in a just-in-time manner once the need arises (Hoogerbrugge et al., 2023, 2024). Only when this "sampling" is impeded (say, one has to walk a few metres to one's computer screen for a digital leaflet), people load up more information per iteration, and shift from relying on the external world to relying on internal memory (Ballard et al., 1995; Böing et al., 2023, 2025; Draschkow et al., 2021; Fu & Gray, 2000; Sahakian et al., 2023; Somai et al., 2020). This shows the interactive and adaptive nature of engaging working memory in natural environments.

The tendency to rely on the outside world has been observed on a group level, but there may be large individual differences regarding the degree to which one relies on either the external world or internal memory, and how much one switches between those when task demands change. The dynamic allocation of external (sampling) and internal memory resources depends upon many different factors, including memory capacity (Böing et al., 2023, 2025, Meyerhoff et al., 2021; Morrison & Richmond, 2020; Risko & Dunn, 2015), general beliefs about or experienced successes with one's memory functioning (Gilbert, 2015, 2025), the perceived importance of accuracy and/or quick task completion (Risko & Dunn, 2015; Sahakian et al., 2023), the goal of effort minimization (Kvitelashvili & Kessler, 2024; Risko & Dunn, 2015), and, as mentioned before, task setup and information availability (Ballard et al., 1995; Draschkow et al., 2021; Gray et al., 2006; Sahakian et al., 2023; Somai et al., 2020). Task properties impose hard constraints on how to carry out a task, while other (personal) factors make up soft constraints (Gray et al., 2006). The unique combination of factors for a given individual and situation likely determines which strategy is applied. Yet, in both experimental working memory paradigms and clinical assessment of working memory, individual variability in working memory strategies have often been neglected (Logie, 2023). As opposed to looking at aggregate scores of performance, assessing individual variability in a range of working memory tasks can yield theoretical as well as clinical advancements in our understanding of working memory use (Logie, 2023).

The limited body of literature that addressed individual variation in strategy deployment did so *within the context of capacity tasks*. For example, Morrison and colleagues (2016) explored how verbal working memory tasks with different demands prompted inter-task and individual variability in strategy deployment. They found that *internal* strategies (e.g., rehearsing, grouping, or visualization) were not homogenously applied across all participants nor consistent within participants across various tasks. Similarly, the deployment of *external* strategies for memorizing verbal information (Morrison & Richmond, 2020) or future intentions (Meyerhoff et al., 2021; Risko & Dunn, 2015), such as writing something down or placing reminders when given the choice, also varied between individuals. With regards to visuospatial working memory, many different strategies have been described to support task performance, among which chunking, holistic encoding, and visuospatial rehearsal, which usage is likely dependent on the task, the individual, and even on the individual trial (Gonthier, 2021). Although these studies substantiate the idea that there is indeed variation in individual approaches to working memory strategies, these findings are, yet again, based upon paradigms where the goal is to memorize and report as much information as possible, and where it is disregarded that information generally remains available in the outside world. As situations often allow for reliance on the external world, (re)inspecting information in a just-in-time manner might be the strategy of choice for the majority of people, but also in these situations individual differences will likely be present.

To take on a perspective on visual working memory that is not focused on how much information people *can* retain, but rather *how* one uses working memory to interact with external information, we should consider both situational demands and, specifically, individual approaches to these situational demands.

Our primary aim was to characterize individual sampling versus loading strategies and their effects on visual working memory task performance. We used secondary data that was collected during a copying task with changing task demands (Sahakian et al., 2023) that conceptually replicated previous copying task studies. Supplemental to previous studies that aimed to engage working memory in more natural environment, the present study attempts to identify underlying strategies that were adopted by the individual. Participants were instructed to rebuild an example puzzle as fast and accurately as possible in a condition in which the example was easily accessible versus a condition in which it was more time-consuming to access the example. Other than in traditional working memory paradigms, this copy task did not force using one's maximum capacity in order to successfully complete the task (e.g., memorizing as much as possible at once; full-loading), but rather allowed the individual to opt for their preferred load. Additional to the group level analysis that was done before (Sahakian et al., 2023), we here attribute meaning to individual behaviour and describe varying strategies. With the assumption that people differ in their natural inclination to either sample or store information when no strategies are forced, we expected some participants to heavily rely on the external world (i.e., to sample often; low-loaders), and some to rely more on their internal memory (medium to high-loaders). Furthermore, when information was not readily available anymore, we expected that people would adjust their strategy to the new situation in a way specific to the individual. Potentially, some people might stick to sampling (low-loading), while others start memorizing more (medium to high-loading). It is possible that behavioural adjustments to changing situational demands may come at a cost of task performance: we hypothesize that larger changes in sampling behaviour come with switch costs (thus, decreased performance) as one needs to discontinue their current behaviour and adjust to the new situation. To assess this idea, we analyse whether initial sampling frequency influences performance, and whether changes in sampling frequency across conditions influence performance across conditions.

As a secondary aim, we considered how the order of the encountered situational demands (thus, prior experience) plays a role in opting for a specific strategy. When people start with a situation in which information remains continuously available, they might adopt a strategy of relying on the external world without memorizing much. When people are initially confronted with a situation in which information can only be sampled after a waiting time, they might adopt a strategy that is biased towards memorization instead of sampling. The context of the first situation may prime people to keep using their initial strategy even if the context changes: an earlier study showed that using a specific source to support memory (here, the internet) predicted future reliance on the same source (Storm et al., 2017). Vice versa, Patrick and colleagues (2015) found, in a task similar to ours, that exposure to only one trial in which information was hard to retrieve was already sufficient to prime people towards a memorizing strategy. This implies that people's perseverance with previous behaviour may result in carry-over effects that are likely to influence the individual's strategy of choice. However, Risko and Dunn (2015) did not find order effects on the decision to offload. Therefore, we took a closer look at whether and how the order of conditions (i.e., experience) in the current copy task primed individuals' strategy initiation and continuation.

Revealing individual differences in working memory strategies in response to changing situational demands contributes to our knowledge of working memory functioning as a dynamic system rather than a fixed entity that is always put to use in an identical manner.

Materials and methods

Participants

We used an existing open access dataset (Sahakian et al., 2023) for analysis. The authors recruited their participants through the platform Prolific. Participants could reside anywhere in the world, but had to be fluent in English, and were excluded if they participated in an earlier study of the same authors. 88 participants provided viable datasets for analyses. The majority of participants resided in Europe. Participants were only asked about gender and age categories, due to which demographic information is limited to proportions of categories. Of 88 participants, 53.4% were male, 45.5% were female, and 1.14% identified as non-binary. 62.5% of participants were in the age category 18–25 years old, 22.7% were 26–33 years old, 6.8% were 34–41 years old, and 8% were over 42 years old.

All participants gave written informed consent prior to the start of the online experiment. Participation was compensated with (the equivalent of) £6.25 upon completion of the task. The study was in accordance with the Declaration of Helsinki, and was approved by the Ethics Committee of the Faculty of Social and Behavioural Sciences of Utrecht University (21-0297).

Task

A previously developed copying task (Somai et al., 2020) was adapted such that it was suitable for online task administration (Sahakian et al., 2023). Results obtained from the online paradigm showed to yield qualitatively similar results as earlier conducted lab-based studies (Draschkow et al., 2021; Sahakian et al., 2023; Somai et al., 2020), confirming the reliability of the task. Different from the previously published group-level analyses and results, we used data derived from the online study to obtain information about individual differences in the allocation of external or internal memory resources across different situational demands. An extensive overview of the experiment, procedure and pre-processing of the data can be found in the original paper (Sahakian et al., 2023). Here, we provide an overview of relevant task features and variables.

The layout of the Copy Task is depicted in Figure 1(a). Participants were instructed to copy 6 items from the 4×4 example grid (model; shown on the left side of the screen) to a 4×4 grid (workspace; located on the inner right side of the screen) as fast and accurately as possible. Items were dragged from the 4×4 resources grid that was located to the far right of the screen. Stimuli consisted of 20 polygons (Figure 1(c); Arnoult, 1956), and 20 colours were added that were selected from the HSLuv (www.hsluv.org) colour space with 90% saturation and 65% luminance, resulting in 400 unique stimuli. For each trial in the experiment, a random selection without replacement of four shapes and four colours was used to create 16 unique stimuli. From these 16 stimuli, 6 were randomly selected with replacement, and randomly positioned in the model for each trial (Figure 1(b); Sahakian et al., 2023). Whenever participants dragged an item over the workspace, the closest cell in the grid was highlighted in yellow. If the item was released at the correct cell, it would automatically align to the centre of that cell. Conversely, if the item was released at an incorrect location, it would return to its initial location in the resources. Note that stimuli remained available in the resource grid, even if they had already been correctly placed. As the experiment was conducted online, a cursor-directed aperture was incorporated in the experimental trials to extract sampling behaviour (Anwyl-Irvine et al., 2022). This technique involves covering the display with an opaque black overlay whilst leaving only a circular area around the cursor transparent, being just large enough to allow full visibility of the model at once (Figure 1(b) and Video S2 – osf.io/w7zag). The transparency of the aperture followed a Gaussian function: it was fully transparent at the centre and gradually less transparent towards the edges.

The Copy Task consisted of two experimental conditions, each consisting of 24 trials (48 trials in total). In the low-cost condition, the moving speed of the aperture was smoothly aligned with the pace of the mouse movement. In the high-cost condition, the aperture moved with a reduced speed (approximately 1.67 s) when crossing the midline from the workspace to the model (across the dark grey border) making sampling more time-consuming. There was no delay when participants moved the cursor from model to workspace (see Video S3 – osf.io/3z8xn). The conditions were blocked, as this allowed individuals to engage in a consistent strategy (Janssen & Gray, 2012; Patrick et al., 2015). The order of blocks was counterbalanced, which allowed dividing the groups based on the order in which they encountered the experimental conditions (i.e., situational demands). When the low-cost condition was followed by the high-cost condition, we referred to this order as "low-cost first". When the high-cost condition was completed first, and the low-cost condition thereafter, we referred to this order as "high-cost first".

In part of the trials (59.2%), participants were interrupted during a trial, and had to answer a two-alternative forced choice (2-AFC) question ("probe questions"). This study does not focus on 2-AFC data (the reader is referred to the original paper of Sahakian et al. (2023) for further details), but we consider the trials valid to be included in the current analysis.

Outcome measures

To describe individual differences in strategy adoption, we derived several outcome measures from the task. First, we computed how often a participant inspected the example model (i.e., crossed the cursor from left

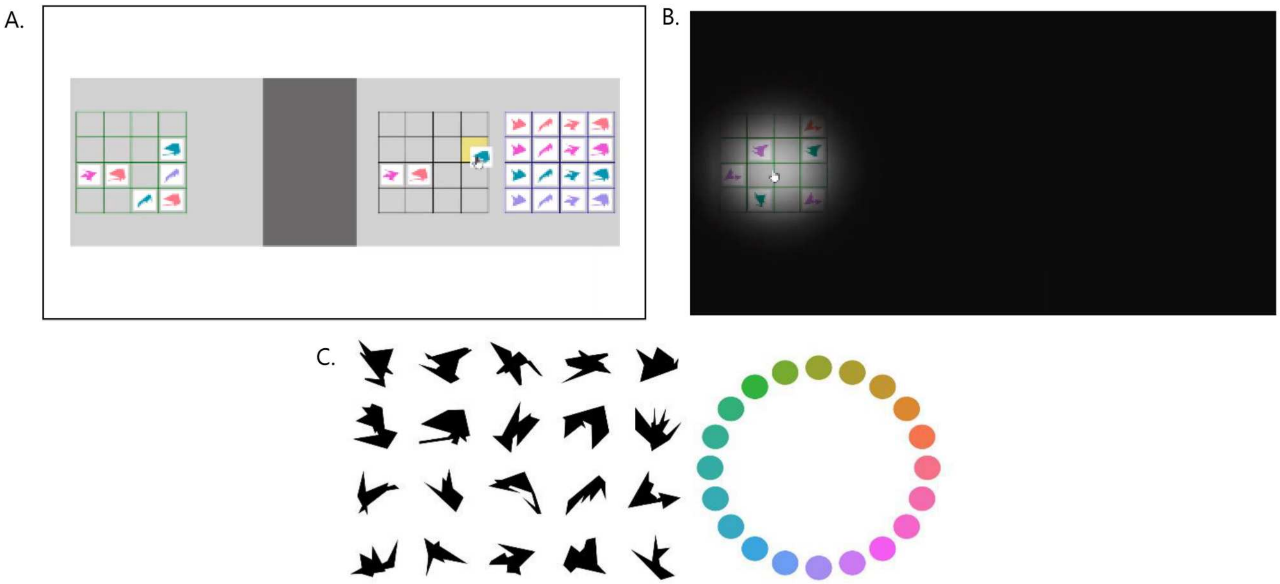

Figure 1. Experimental layout. (a) illustrates the experimental layout, consisting of the 4 × 4 example grid (model) shown on the left side of the screen, and the 4 × 4 grid (workspace) depicted on the inner right side and the resources grid (Resources) depicted on the outer right side. (b) Shows a still from an experimental trial, as seen by the participants. A cursor-directed opaque black overlay covered the display as depicted in Panel (a), allowing for full visibility of the Model at once. (c) The stimuli set (20 polygons x 20 colours; 400 unique stimuli (Sahakian et al., 2023)). Figures adapted with permission.

to right), averaged over trials, resulting in the mean *number of model inspections*. Second, we computed how long someone viewed (i.e., encoded) the example model per inspection. This was calculated by dividing the total duration of model inspections in seconds by the number of inspections per trial, and consecutively taking the median *dwell time at the model per inspection*. Conceptually, a sampling strategy translates to a high number of model inspections with shorter dwell times per inspection, whereas a memorization strategy translates to a low number of model inspections with longer dwell times per inspection. Therefore, participants were classified based upon the number of model inspections. These classification labels are solely introduced for descriptive purposes, and may serve as a rule of thumb to give some insight on the individual's general sampling tendency. Participants who made on average more than one inspection per correctly placed item were classified as low-loaders. For example, such participants may decide to first memorize the item, find the item in the resource grid, then reinspect the model to memorize the location, and only then place the item in the empty grid. Participants who correctly remembered one or more item(s) per inspection (i.e., made one or fewer than one inspection per correctly placed item) were classified as medium-loaders. Those who correctly remembered three or more items per inspection (i.e., made 0.33 inspections or fewer per correctly paced item) were classified as high-loaders (also see Analysis).

Participants were instructed to rebuild the example puzzle as quickly and accurately as possible. Although we thereby do not emphasize either outcome, we cannot guarantee that individuals have an equal attribution of importance for either speed or accuracy. This leaves room for individual differences in motivation to either perform the task without errors, or "quick but dirty". To control for potential individual differences in motivation and to deal with the presence of a speed-accuracy trade-off, performance was assessed by calculating a *linear integrated speed-accuracy score (LISAS;* Vandierendonck, n.d., 2017, 2021), depicted in Eq. (1).

$$LISAS = RTij + PEij\ x\frac{SRT}{SPE} \tag{1}$$

Here, RT_{ij} refers to the completion time on trial i divided by the number of correct placements on trial i for individual j. In the high-cost condition, the aperture delay was first subtracted from the completion time. The speed data was log transformed to account for skewness associated with time measures. PE_{ij} refers to the proportion error on trial i (1 minus the number of correct placements divided by the total attempts). We calculated the standard deviations S_{RT} and S_{PE} per individual over both conditions collapsed (Vandierendonck, n.d., 2017, 2021). A lower LISAS reflects better performance.

To investigate the degree to which each participant adapted their sampling strategy between conditions, we divided the number of inspections in the high-cost condition by those in the low-cost condition ("change factor number of model inspections"). The higher the change factor number of model inspections, the larger the adaptation from the low-cost condition to the high-cost condition. The same was done for the LISAS ("change factor LISAS"), to see how performance changed for the individual. A higher change factor LISAS reflected a larger drop in performance from the low-cost to the high-cost condition.

Pre-processing

The data were retrieved from https://osf.io/pkxdc (Sahakian et al., 2023). In the current study, we used the same filter and exclusion criteria as in the study of Sahakian et al. (2023); no additional participants were excluded from the analysis. The trials with probe questions (n = 1569, 59.2%) and without probe questions did not differ regarding the number of model inspections, dwell times per inspection, and the number of errors (Sahakian et al., 2023). Therefore, trials with and without probe questions were treated similarly. The main analyses were performed on all data (n = 4219 trials). The level of significance was set to an alpha of 0.05. To make sure that findings were not driven by outliers, we ran sensitivity analyses after removing those trials with scores ≥1.5 times the interquartile range apart from the group median for that specific outcome measure (i.e., number of model inspections or LISAS) in that specific condition (i.e., low-cost or high-cost) per order of condition (i.e., low-cost first, high-cost first). Information on outlier removal and sensitivity analyses can be found in the Supplementary Materials.

Analysis

Individual differences in strategy and effects of situational demands and prior experience

Statistical analyses were performed in R 4.1.2 (R Core Team, 2021).

First, we provided the number of model inspections for each condition (low-cost, high-cost) and both presentation orders (low-cost first, high-cost first), and we ran a non-parametric Kendall Rank correlation between the number of model inspections and dwell time at the model per inspection. We expected that a higher number of model inspections would relate to shorter dwell times per inspection, and vice versa. Fewer inspections with longer dwell times (i.e., encoding) would reflect a tendency towards memorization. Correlation coefficients were reported as tau (τ) and effect sizes as z.

To characterize the natural tendencies of individuals to either rely on external sampling or internal memorizing, we classified individuals as low-loaders, medium-loaders, or high-loaders based on how often they inspected the model to place one item correctly. We extracted the number of model inspections per trial (e.g., the number of times a participant moved their cursor to reveal the model) and divided this number by six (each trial had six items to copy). The choice of cut-offs for the different categories were partly data-driven (i.e., aiming for a substantial number of participants per category), partly theory-driven, and mostly based upon task constraints. Individuals who inspected the model more than once per correctly placed item were categorized as low-loaders. These participants were considered to employ the bare minimum working memory load in the copy task. To illustrate, participants may have memorized the polygon shape first, and may have memorized the location upon reinspection, thus memorizing the item in a feature-by-feature manner. Individuals who inspected the model once or less than once per correctly placed item, and thus memorized one bound item (shape plus location) or more items per inspection, were categorized as medium-loaders. Individuals who correctly placed three or more items per inspection were further classified as high-loaders. Our task did not have a high enough resolution to dissociate between people loading three, four or five items, as all of them would need an additional inspection for the remaining items, yielding two inspections for trial completion. Importantly, it has been claimed that people have an estimated working memory capacity of four items (Cowan, 2001), but for more complex shapes such as polygons, this maximum capacity seems to be decreased (Alvarez & Cavanagh, 2004; Luria et al., 2010; Luria & Vogel, 2011). We therefore cannot make conclusive statements but only speculate about whether or not individuals are fully loading their capacity. Importantly, these cut-offs are arbitrary, and were solely used to characterize individual strategies used in the current task.

First, we described the strategies used in the condition in which information was freely available, as this context resembles our daily life environment the most. We did this for the participants who started with this condition (i.e., low-cost condition first), as participants were not yet influenced by the situation where information availability was manipulated (high-cost condition). Next, we investigated whether in these participants, varying the situational demands led to an adaptation in sampling behaviour (e.g., relying more on the outside world or on memorizing). Therefore, we described whether participants used different strategies in this condition, to what extent they changed their strategy from the low-cost to the high-cost condition, and whether the previously established categories were still observable.

We also explored whether and how prior experience affected strategy choice. To this end, we explored data of the group that started with the high-cost condition to investigate whether introducing higher sampling costs at the outset of the task prompted participants to adopt a strategy biased towards heavier reliance on memory, and whether they stuck to this behaviour also when there was only a low cost to sampling. We assessed whether our previously introduced categories (low-loaders, medium-loaders and high-loaders) were still observable when the order of conditions was reversed.

In addition, we investigated the influence of the order of encountered situational demands *between* participants. To this end, all trials were fed to a linear mixed-effect model (LMM; Singmann & Kellen, 2019) by using the lmer function in R (lme4 package; Bates et al., 2014). The LMM is robust against deviations from normality of the outcome variables and takes individual differences within groups into account (Schielzeth et al., 2020). Factors included were order (low-cost first, high-cost first), condition (low-cost, high-cost), the interaction of order and condition, and random slope and intercept for individuals. We ran the model to predict the influence of these factors on the *number of model inspections*. The normality of the residuals was visually examined and confirmed. Effect sizes were reported as beta-coefficients with a 95% confidence interval. Post hoc pairwise comparisons were conducted using the Tukey–Kramer method using the contrast function within the emmeans package, which accounts for multiple comparisons and controls the family-wise error rate.

The effect of strategy on performance

The second aim was to study how a chosen sampling strategy influenced performance. We first investigated the group that started with the low-cost condition, as we were mostly interested in initial sampling preference in a situation where information was freely available. We checked whether there was a favourable strategy category in terms of performance (reported in the Supplementary Materials), and we used a non-parametric Kendall Rank correlation analysis to investigate whether and how initial sampling preference (number of model inspections in the low-cost condition) related to performance in the low-cost condition. Then, we evaluated whether individuals who showed larger changes in sampling frequency when moving to the high-cost condition (e.g., from low memory reliance in the low-cost condition to a higher memory load in the high-cost condition) showed more decline in performance compared to those who were more stable in strategy, or vice versa. We assessed the relation between the change factor number of model inspections and change factor LISAS with a Kendall Rank Correlation analysis. To more broadly assess the influence of order and condition on performance, we also fitted a linear mixed-effects model to LISAS.

Code and software

Experiment code, pre-processed data and raw data are publicly available and can be found at Open Science Framework: https://osf.io/pkxdc/ (accessed July 2023). Data and analyses for the current study are accessible via https://osf.io/cjgaf/.

Results

Conceptualization of memorization

We found that there was a moderate negative correlation between the number of model inspections and dwell time per inspection ($\tau = -.55$, $p < .001$, $z = -10.76$; Figure S1), indicating that fewer model inspections were related to longer inspection durations. With the current number of observations (176), for a one-tailed correlation test ($\alpha = .05$) with a power of 0.8, we should be able to reliably detect effects sizes of 0.18 (Faul et al., 2009).

Individual differences in strategy and the effects of situational demands and order

Within the group of participants that started with the low-cost condition (n = 43), 17 participants (39.5%) were classified as low-loaders as they inspected the model more than once to place one item correctly (Table 1; Figure 2(a), black dots). Twenty-four participants (55.8%) were categorized as medium-loaders, as they made one or fewer than one inspection to place one item correctly, implying that they relied relatively more on memorization. Importantly, only two participants (4.7%) placed three or more items correctly per inspection, thereby classifying as high-loaders. These results indicate that, generally, participants relied on the external world and used working memory capacity only to a limited extent. When moving to the high-cost condition, all participants (100%) changed their strategy in the sense that they made fewer inspections as compared to in the low-cost condition. Now, only one participant (2.3%) was classified as a low-loader, against 69.8% medium-loaders and 27.9% high-loaders (Table 1; Figure 2(a), red dots). This shows that when information was less readily available, participants tended to rely less on the external world and more on their

Table 1. Mean number of model inspections per order (low-cost first, high-cost first), per condition (low-cost, high-cost) and per strategy category (low-loader, medium-loader, high-loader).

		Order							
		Low-cost first (n = 43)				High-cost first (n = 45)			
Condition	Strategy category	n(%)	M	SD	Range	n(%)	M	SD	Range
Low-cost	Low-loader	17 (39.5%)	7.69	1.43	6.12–10.5	3 (6.7%)	7.26	0.17	7.21–7.46
	Medium-loader	24 (55.8%)	4.89	1.03	2.17–6	35 (77.8%)	4.02	1.12	2.08–6
	High-loader	2 (4.7%)	1.65	0.21	1.5–1.79	7 (15.5%)	1.60	0.38	1.04–2
High-cost	Low-loader	1 (2.3%)	7	–	–	0 (0%)	–	–	–
	Medium-loader	30 (69.8%)	2.86	0.53	2.17–4.58	33 (73.3%)	3.25	0.97	2.04–5.83
	High-loader	12 (27.9%)	1.39	0.38	1–2	12 (26.7%)	1.46	0.34	1.04–2

Notes: We provide the number of individuals (n) and percentages per category based upon the number of model inspections, and the mean (M), standard deviation (SD) and the range of the number of model inspections per trial. Note that the division in categories is arbitrary and is for descriptive purposes only.

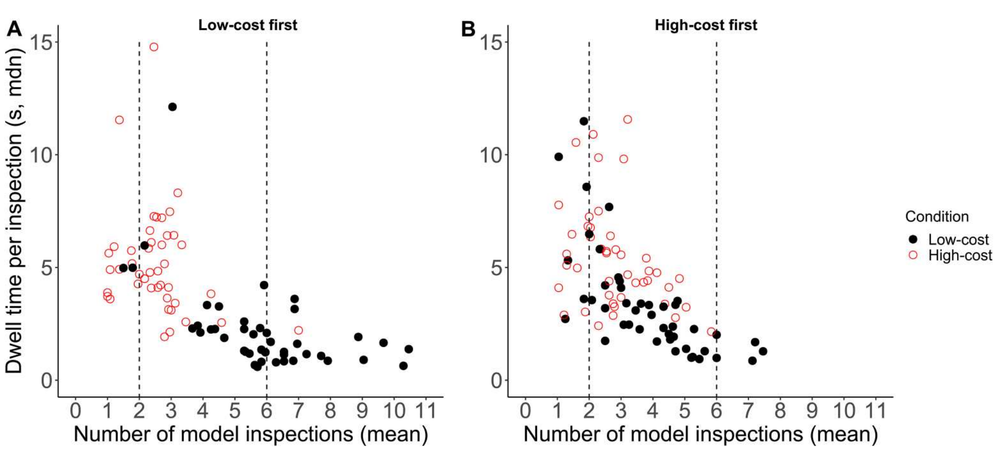

Figure 2. Copying behaviour, presented as median model dwell time per inspection per trial and the average number of inspections per trial. (a) Low-cost first. (b) High-cost first. Data points represent data of the individual in the low-cost condition (filled black dots) and high-cost condition (open red dots). The vertical dashed lines represent the cut-offs used to discriminate between low-loaders (dots to the right of the right dashed line), medium-loaders (dots on and to the left of the right dashed line), and high-loaders (dots on and to the left of the left dashed line).

internal memory capacity. However, the majority of people still did not employ a high-loading strategy.

Next, we investigated whether the presentation order of conditions affected the strategy that was used. In other words, did the context of the first condition prime the strategy that was used in the second condition? We here included the group of participants who started with the high-cost condition and completed the low-cost condition afterwards (n = 45). Interestingly, none (0%) of the 45 participants were categorized as low-loader in the high-cost condition, 73.3% was categorized as medium-loader and 26.7% as high-loader (Table 1; Figure 2(b), red dots). In the following low-cost condition, 3 participants (6.7%) shifted to a low-loading strategy (Table 1; Figure 2(b), black dots). A medium-loading strategy was used by 35 participants (77.8%), and 7 participants (15.5%) held on to the high-loading strategy. This means that the majority of participants was consistent in their strategy, even in a context in which information in the external world became freely available. Thus, strategy is not only dependent on the current context, but also seems influenced by prior experience.

To further investigate the effects of situational demands and the order they are encountered in on the number of model inspections, we fitted a linear mixed-effect model (LMM) to predict the number of model inspections with these factors, while controlling for individual differences (Figure 3). The LMM showed a significant effect of condition ($t = -14.22$, $p < .001$, beta = −3.31 [−3.76, − 2.85]), a significant effect of order ($t = -5.11$, $p < .001$, beta = −1.99 [−2.76, − 1.23]), and a significant interaction effect between condition and order ($t = 6.83$, $p < .001$, beta = 2.22 [1.58, 2.86]). Importantly, this interaction effect showed that prior experience (order) differentially affected sampling behaviour across conditions. Post-hoc tests revealed that the group starting with the low-cost condition made significantly more model inspections in the low-cost condition (M = 5.85, SD = 2.03) than the group that started with the high-cost condition ($M = 3.86$, $SD = 1.62$, $p < .001$). In the high-cost condition, the groups showed about the same number of model inspections (low-cost first: $M = 2.54$, $SD = 1.08$; high-cost first: M = 2.77, $SD = 1.1$, $p = .77$). Together, this implies that starting with the high-cost condition primed participants to keep using internal storage to a larger extent, also if the situation did not demand this per se.

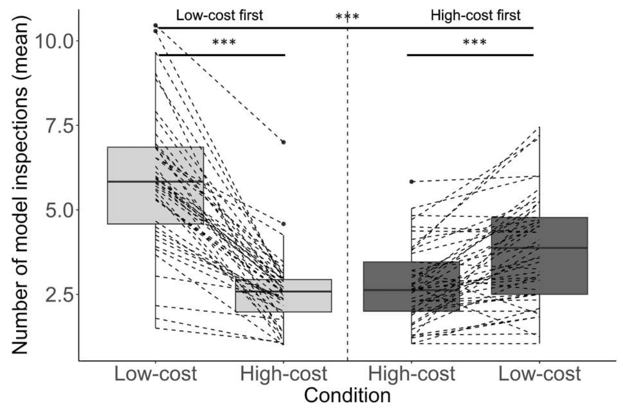

Figure 3. Data, presented as mean (+IQR) inspections per condition (low-cost, high-cost) for the different orders (low-cost first, high-cost first). There were more model inspections in the low-cost condition compared to the high-cost condition. Participants that started with the high-cost condition (dark grey) sampled less often than the group that started with the low-cost condition (light grey). The interaction-effect revealed that the low-cost condition differentially affected sampling behaviour across groups, with the high-cost first group making significantly fewer model inspections. *$p < .05$ **$p < .01$, ***$p < .001$.

The effect of strategy on performance

Next, we evaluated whether there was a relation between strategy and performance. For participants that started with the low-cost condition (n = 43), there were no differences in performance across low-loader, medium-loader, and high-loader categories in the low-cost condition (see Figure S2). When analysing continuous data from the same group, we found no significant correlation between the number of model inspections and LISAS in the low-cost condition ($\tau = .006$, $p = .96$; Figure 4(a)). So, when information was freely available, there was no favourable inspection frequency yielding better performance. However, when looking at the transition from the low-cost to the high-cost condition, some people made a greater behavioural adaptation compared to others (Figure 4(b), length of dashed lines). Therefore, we investigated whether the *magnitude of behavioural adaptation* (change factor number of model inspections) was related to performance within the individual (Figure 4(c)). For the low-cost first group, no significant correlation was found between the change in inspection frequency and the change in performance ($\tau = -.04$, $p = .69$), meaning that performing the task in a non-preferred approach (hence, larger adaptations; high change factor) when moving to the high-cost condition did

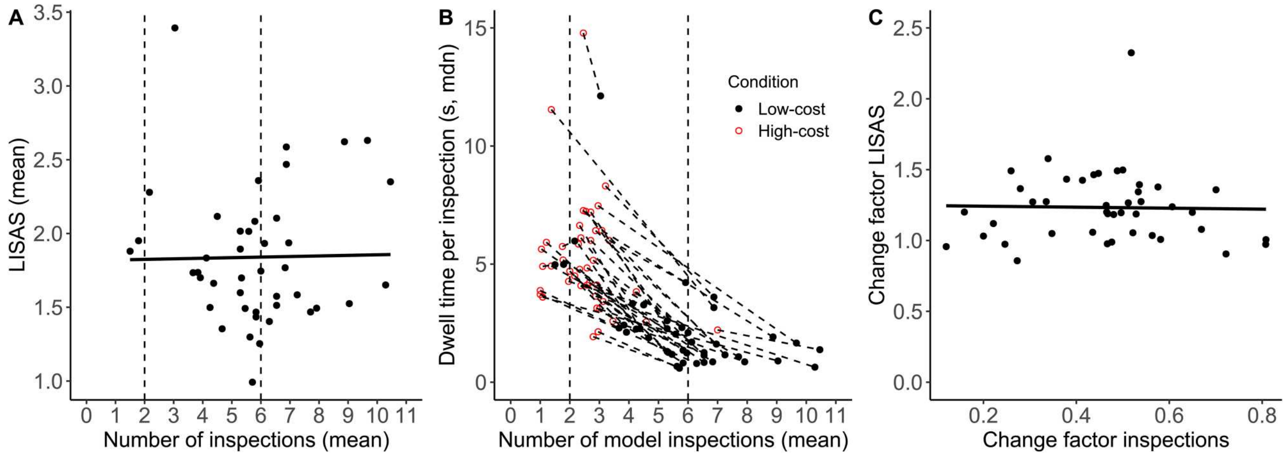

Figure 4. Effects of sampling behaviour on performance, behavioral strategy adaptations and effects of strategy adaptations on changes in performance for the low-cost first order group only. Each point reflects an individual participant (n = 43). (a) In the low-cost condition, the mean number of inspections is not correlated to performance (mean LISAS; higher scores reflecting worse performance). The vertical dashed lines represent the cut-offs used to discriminate between low-loaders (dots to the right of the right dashed line), medium-loaders (dots on and to the left of the right dashed line), and high-loaders (dots on and to the left of the left dashed line). (b) Representation of the behavioral shift based on mean number of model inspections per trial between the low-cost condition (closed black dots) and the high-cost condition (open red dots). The longer the line between two data points, the larger the behavioral adjustment for the individual. (c) No significant correlation was present for change factor number of model inspections and change factor LISAS. Scores towards zero on change factor number of model inspections reflect a larger adjustment from the low-cost condition to the high-cost condition. The higher the change factor LISAS, the more performance declined in the high-cost condition compared to the low-cost condition. Change factor LISAS <1 indicates performance improvement.

not affect performance more than for those who behaved more consistently across conditions.

We are aware that this analysis is underpowered (to be able to reliably detect a moderate effect size of 0.3 with a power of 0.8, we should have 64 observations instead of 43). However, due to the nature of our data we cannot treat all observations similarly. Collapsing the data across the low – and high-cost first group would bias results as their respective change factors indicate another transition (from low – to high-cost, or vice versa) and thus have different meaning. We have therefore chosen to do this analysis only for the low-cost first group as we assume this to be the most natural way of encountering information in everyday life.

However, we did want to assess the influence of the order of conditions on performance, and therefore investigated whether participants who started with the high-cost condition – and thus showed different sampling behaviour, as found in section 3.2 – performed differently than participants who started with the low-cost condition. Therefore, we decided to fit a linear mixed-effect model to LISAS to simultaneously analyse the influence of order and condition (Figure S3; Table S4). First, a significant main effect of condition ($t = 7.59$, $p < .001$, beta = 0.37 [0.27, 0.46]) was found, indicating overall worse performance in the high-cost condition compared to the low-cost (Figure S3). No significant main effect of order ($t = 0.4$, $p = .693$, beta = 0.03 [−0.13, 0.19]), nor an interaction effect between order and condition ($t = 0.2$, $p = .841$, beta = 0.014 [−0.12, 0.15]) was present.

To summarize the results on sampling behaviour and performance, both information availability (low-cost versus high-cost) and previous experience (low-cost first versus high-cost first) influenced sampling behaviour. There was no specific inspection frequency yielding better performance, and the magnitude of change in sampling behaviour did not affect performance within the individual. Note that including trial number in any of the reported LMM yielded qualitatively similar results, showing that when general practice effects were taken into account, the same conclusions could be drawn.

Sensitivity analyses

We detected and removed all outlier trials, and ran all analyses again. Detailed results can be found in the

Supplementary Materials (S5 – S11). Descriptively, some classifications changed (compare Table 1 and Table S6). Effectively, there were two additional participants classified as high-loader (1 in low-cost condition, low-cost first; 1 in low-cost condition, high-cost first). Three participants were initially classified as low-loader, but were now classified as medium-loader (all four in low-cost condition, high-cost first). Apart from these deviations in classification numbers, outlier removal did not yield different results when rerunning the analyses, and our interpretations again held when testing for general practice effects.

Discussion

Although estimations of maximum capacity have proven fruitful in visual working memory research (Luck & Vogel, 2013; Ma et al., 2014), this approach falls short in grasping the usage of working memory in everyday life. Not only do capacity tasks demand full use of memory capacity, which is often not required in everyday situations, they also fail to reveal individual differences in the use of strategies that might be employed to support memory functioning in different situations. Rather than considering working memory as a system that is defined by its maximum capacity, we here addressed visual working memory as a dynamic system that is used flexibly in response to changing situational demands, *and* is subject to individual differences in how one deals with these demands. Departing from traditional paradigms that force the use of full working memory capacity to successfully complete the task, we used pre-existing data from a paradigm (Sahakian et al., 2023) that allowed individuals to store information at the preferred load or otherwise to rely on (i.e., sample from) the outside world. Their results conceptually replicate previous group-level results in that people tend to make little use of visual working memory capacity when the situation allows for rapid sampling from the external world (Ballard et al., 1995; Böing et al., 2023, 2025; Draschkow et al., 2021; Fu & Gray, 2000; Somai et al., 2020). Reliance on sampling only decreased when information became less readily available (see Sahakian et al. [2023] for this group level analysis). While the tendency to decrease sampling in response to restricted information availability has been repeatedly observed, individual differences in (sampling) strategies have largely been ignored (Logie, 2023). In the current study, we partly look at aggregated strategy data, but place emphasis on individual differences.

In line with the sparse literature on the existence of individual differences in strategy use within *capacity* tasks (Morrison et al., 2016; Morrison & Richmond, 2020), we similarly identified individual differences in the extent to which people relied on external sampling across changing task demands. When initial information was continuously available – which often is the case in everyday tasks – more than a third of people sampled more than once to correctly place one item (i.e., *re*sampling). These *low-loaders* had a default setting to heavily rely on the outside world. Others used the external world to a lesser extent and loaded more information per iteration. Although individuals thus differed in the degree to which they relied on external sampling versus internal loading (low-loaders versus medium-loaders and high-loaders), very few participants showed infrequent sampling behaviour to a degree that they approached the limits of their working memory capacity (high-loading). Given that people *should* theoretically be able to load up three to four items in visual working memory (Cowan, 2016; Luck & Vogel, 2013), they *should* be able to place all six items with merely two inspections. Interestingly, when given the choice, people rarely do so: the majority of people inspects the model more often than strictly necessary, and few people load up three or more items at once.

When information availability was restricted, behaviour became more homogeneous. Individuals who initially *re*sampled reduced their sampling rate to such an extent that they would no longer be characterized as low-loaders. While *all* individuals reduced their sampling frequency (aligning with earlier group-level analyses), some individuals adapted more than others. Some of them even used only one or two inspections to complete a trial, meaning that they turned towards higher memory loads (three or more items). Still, even with changed task demands, the majority were reluctant to internalize information at high-load in working memory.

Interestingly, those who made greater adjustments in their sampling frequency from the low-cost to the high-cost condition did not perform worse than those who behaved more consistent across situations, suggesting that behavioural adaptations can be successfully incorporated by the individual. In other words, when information was freely available, there was no strategy that yielded better performance, and the magnitude of change in sampling frequency across conditions did not affect performance within the individual. Although we expected larger adaptations to negatively affect performance, the results imply that individuals can flexibly adjust their use of visual working memory to fit situational demands, even if this means that they approach the task in a way that does not match their initial tendency.

Notably, we found that the *order* in which one experienced changing situational demands differentially affected sampling behaviour: individuals who started with a task in which information availability was restricted, kept loading up more information during each inspection in a subsequent task in which information was freely available. This finding replicates a study of Patrick et al. (2015) that showed that starting with restricted access led to adhering to a memory-based strategy even when the restriction was lifted, whereas there was no adherence to a sampling strategy after the transition from non-restricted to restricted access. Given that some of the individuals in the high-cost first group would probably have shown low-loading behaviour if the order had been reversed (as we infer from the current study and Patrick et al. [2015]), participants in the high-cost first group may have learned that they could successfully rely on their working memory. This finding is reminiscent of previous work showing that people who had a successful memory usage experience were less inclined to externalize information as compared to people who had a less (or no) successful experience (Gray & Fu, 2004; Risko & Dunn, 2015). In these cases of changing task demands, the costs associated with strategy switching may be higher than the effort associated with (the continuation of) memorization (Gilbert, 2023; Kurzban et al., 2013; Xie & Zhang, 2023). Furthermore, for this high-cost first group, the *experienced* effort of using memory at higher load may have been lower than the *anticipated* effort that the other (low-cost first) group expected to encounter when using memory at higher load (Bambrah et al., 2019). Experience and expectation therefore seem to jointly explain why order influences the choice to either offload or not. Although these factors drove behaviour in the high-cost first group towards a larger tendency to memorize, only some individuals completed trials with only one or two inspections (thus loading three or more items). The majority of participants did not load working memory capacity to such a high load, and relied on the outside world more than necessary assuming a capacity of three to four items (Cowan, 2016; Luck & Vogel, 2013). Under this assumption, we conclude that deploying visual working memory at high-load more likely is a matter of willingness rather than ability. However, we acknowledge that more complex shapes such as polygons may give rise to a different (i.e., lower) maximum capacity value (Alvarez & Cavanagh, 2004; Luria et al., 2010; Luria & Vogel, 2011). On the other hand, this capacity may be increased again when the participant could draw from semantic representations in long-term memory to verbally label the polygons (Chung et al., 2024). For example, some participants may have recognized some of the polygons as a plane or a star, thereby using a verbal strategy to support visual working memory. We therefore cannot make conclusive statements but only speculate about the extent to which individuals were making use of the low-, medium-, or high-end spectrum of their maximum capacity for the used stimuli.

Related to the previous point, we emphasize that our category conceptualization is subject to debate. One should keep in mind that our definitions of low-, medium-, and high-loader may not cover the exact underlying mechanism. We interpret our results under the assumptions that low-loaders are individuals who, in general, heavily rely on (i.e., (re)sample from) the outside world and do not load much information into working memory. However, individuals may be classified as low-loader because they inspect external information relatively often, but that does not exclude the possibility that they *may* have loaded their memory fully, but just checked themselves multiple times before they acted upon that information. In a similar fashion, *within* trials, people can be classified differently at different points in time. Consider, for example, an action sequence in which a participant loads up three items on the first inspection, places all of them (high-loader behaviour), and then inspects and reinspects the model multiple times (low-loader behaviour) to the extent that this individual would be classified as a low-loader. The same may occur *between* trials. Further, it should be noted that some high-loaders exhibited only one inspection, implying they memorized all six items at once, while this single inspection was accompanied with a relatively short inspection time. It is certainly possible that these participants had very efficient memorization, but these data points could also reflect cheating behaviour (e.g., taking a picture of the model and using that picture to avoid inspecting the model in the online paradigm). In other instances, individuals dwelled relatively long at the model, which may reflect distractions or lapses during the task that we cannot rule out. We are aware of the ambiguities in our definition and acknowledge it's shortcomings, but are confident that our clustering approach still captures the individuals' *general* tendency. Similarly, although the relation between the number of inspections and dwell time may differ between participants, it does reflect the overall relationship.

We propose that the tendency to employ working memory at lower load is driven by effort expenditure. As information storage in visual working memory has been described as fragile (Cowan, 2001; Ricker & Cowan, 2010; Zhang & Luck, 2009) and effortful (Kardan et al., 2020; Xie & Zhang, 2023), people will likely try to circumvent the effort or uncertainty associated with maintaining multiple items. Indeed, it has been found that offloading behaviour – of which sampling can be seen as a subordinate – occurs as a result of subjective effort reduction (Risko & Dunn, 2015). Our findings are consistent with the literature describing the constant weighting of effort input and performance output at both the psychological and neurophysiological level (Kurzban et al., 2013). As there were no direct benefits (no specific rewards or importance) for the participant to fully tax memory, the decision to not memorize to high load seems a logical one. The brain is often described as a system that pursues optimal efficiency, and this leads us to speculate about optimality. Is it a matter of optimal performance or a question of optimal resource allocation? We dub a "battle of the trade-offs": in terms of resources, there is a continuous weighting to reach an optimal equilibrium between sampling or storing expenditure, but in terms of performance, there is also an optimal balance in the speed-accuracy trade-off. These two scales operate in synchrony and are constantly pushing and pulling their respective weights. Here, reluctance to commit to loading may be an important drive to opt for low-loading (effort trade-off), especially since there does not appear to be a performance benefit for either low-, medium – or high-loading (performance trade-off).

The elegance – but also the complexity – of our paradigm is exactly that one can always freely opt for whichever strategy is most appealing. This also means that we approach strategies from a volitional point of view, and do not explicitly instruct our participants to adopt either low-, medium – or high-loading. We intentionally did not do this, as this would align more with traditional capacity tasks where a particular strategy is enforced to successfully complete the task (i.e., adhering to the instructions). While this volitional approach shows natural tendencies in engaging visual working memory and shows not to affect performance, it is possible that explicitly *instructing* a strategy might yield different performance outcomes, and that there might actually be performance benefits (or disadvantages) if we would instruct people to make more extreme changes or show a more consistent strategy.

Here, the only instruction was to complete the task as quickly and accurately as possible. Performance was therefore calculated as a linear integrated speed-accuracy trade-off score (LISAS; Vandierendonck, 2017). This measure corrects reaction time for the number of errors that were made, and therefore considers both speed and accuracy as equally favourable outcomes. Consequently, effects on either speed or accuracy of any of the strategies (one being faster but the other being more accurate) are balanced out with this measure, and do not yield a "better" strategy. However, in everyday settings – or due to participants' convictions about what constitutes good performance – accuracy may be favoured over speed. In such cases, it may be beneficial to *re*sample to safeguard an activated representation of the information to-be-used, even if this takes up additional time. Ensuring that the representation of memory content matches the target item by resampling then strengthens the decision to act upon the internal representation and place the item (e.g., reaching an action threshold; Sahakian et al., 2023). Potentially, low-loaders simply need more certainty to reach their action threshold as compared to medium – or high-loaders. A task-specific factor that may have influenced the perceived unequal importance of speed or accuracy, is that the paradigm incorporated 2-AFC probes to assess participants on any residual memory traces that were or were not put to use (Sahakian et al., 2023). Being aware of the possibility to be examined, participants may have placed more weight on accuracy than on speed. Unfortunately, we have no insight into participants' motivation to adhere to the task instructions, as the task was carried out online. If accuracy was deemed more important than speed, we hypothesize that frequent (re)inspecting would occur in order to check oneself, even when sampling costs were imposed. Future research should point out how changing the speed-accuracy scale (e.g., by placing more weight on either of the two outcomes) affects loading behaviour. Instructions and personal motivation should be separated in order to gain a better understanding of the variables involved.

Our study provides insight into how individuals allocate visual working memory resources to different situational demands when given the freedom of choice. Future research could investigate what exact factors underly an individual's allocation of external versus internal resources in visual working memory. Furthermore, it would be valuable to explore what additional factors might help understand this allocation question, such as stimulus familiarity and practice effects. Familiarity with specific objects could reduce effort and facilitate memorization (Blalock, 2015; Poppenk et al., 2010; Xie & Zhang, 2018), and thus could lead to reduced sampling frequency and/or improved performance *over the course of the task* (even so within blocks). Intriguingly, effects of training and experience on strategy continuation and adjustments seem to be time and frequency dependent. For instance, Patrick et al. (2015) observed that one "reversal" trial provoking memorization was already sufficient for people to opt for and continue with a memorization strategy for at least ten consecutive trials, even though the individuals were initially trained to use an external sampling strategy. Extending this research in an intention offloading task, Scarampi and Gilbert (2020) assessed the time-course of choice or a certain strategy after primed strategies and reversal trials. They found that a reversal trial provoked an immediate adaptation to that strategy, but that this effect wore off after a number of trials, after which people turned to the primed strategy again. To get a broad idea of behaviour and performance over the course of the current copying task, we have visualized the number of model inspections and performance as a function of trial number in Supplementary Figures S12 and S13. Similar to adaptation effects, more general task practice effects could have led to altered behaviour towards the end of the block or task. When checking for this, we indeed confirmed that practice effects influenced sampling frequency and performance, but not to the extent that the other factors were not meaningful anymore. Yet, potential detrimental effects of strategy adaptation on performance (as the individual moves from one condition to another) are not as clear-cut across conditions, and could be overshadowed by improved performance due to these learning and habituation effects.

Finally, it would be interesting to assess under what circumstances someone would switch from being a medium-loader to a high-loader, or even a full-loader, by incorporating a continuum of changing task demands rather than only two discrete conditions. Such a study could reveal a "tipping point" for strategy adaptation. Findings regarding the linear relationship between (working) memory capacity scores and offloading behaviour reveal multifaceted contributions of memory subsystems: some studies find an effect of capacity while others do not, or only partly (in healthy subjects: Meyerhoff et al., 2021; Morrison & Richmond, 2020; Risko & Dunn, 2015; in patients: Böing et al., 2023, 2025). Differences between studies can partly be attributed to different capacity tasks used to estimate capacity. It would be interesting to see which of these findings could be replicated when the capacity scores are not derived from traditional neuropsychological tasks, but directly reflect the capacity for the stimuli used in the paradigm at hand (in the current study, those of Arnoult, 1956). Other factors, such as personality traits or neuropsychiatric tendencies (e.g., compulsive checking in obsessive-compulsive disorder) could also be of influence.

The current research only covers a small part of the strategy palette: strategically using the outside world to aid visual working memory task performance is one option, but there are many other strategies (see Gonthier (2021) for a review) that may in itself be strategically employed during our Copy Task. To illustrate, it is possible that individuals do not increase the mere *visual* representation load when the sampling cost is elevated – as implied in this study – but that the individual more actively engages in verbal recoding, thereby expanding the capacity that could be loaded (Chung et al., 2024). Although such strategy usage may have occurred within our participants, individuals still had the option to not use any of those and instead rely on the external world to circumvent (the effort associated with) internal strategies and high-loading. Thus, irrespective of whether or not participants used any of the internal strategies, we find that people tend to heavily rely on the external world.

In summary, while many factors complicate the attribution of *why* a sampling strategy occurs, having an eye for strategy use can still be insightful in seeing *how* one uses working memory. Recognizing that each individual has their own preferred approach may help future researchers and clinicians to understand the complex dynamic nature of visual working memory use outside of the lab or clinic.

Conclusion

Visual working memory use is clearly not solely determined by an individual's visual working memory capacity. Individuals tend to rely on the outside world more than strictly necessary given capacity, and they flexibly adapt this degree of reliance to changing situational demands. We identified low-loaders, medium-loaders, and high-loaders, and although we could distinguish these individual differences in reliance on the outside world, the majority of people is – and remains – reluctant to approach the higher end of memory capacity use. We suggest that this individual variation is the result of an ongoing weighting of resource allocation (the effort of sampling vs. storing) relative to optimizing performance (speed vs. accuracy). Prior experience, underlying personal characteristics (e.g., motivation or confidence) and the recruitment of other strategies can in turn influence these trade-offs. We conclude that visual working memory is an adaptive system that is employed based on situational demands, effort and performance expenditure, and underlying individual tendencies.

Acknowledgements

We thank Andre Sahakian and colleagues for collecting and providing their dataset on: https://osf.io/pkxdc/ (Sahakian et al., 2023).

Author contributions

Sanne Böing: Conceptualization, Formal analysis, Methodology, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing. **Beleke de Zwart:** Conceptualization, Data Curation, Formal analysis, Methodology, Visualization, Writing – Original Draft,. **Antonia F. Ten Brink:** Conceptualization, Supervision, Writing – Review & Editing. **Tanja C.W. Nijboer:** Conceptualization, Supervision. **Stefan Van der Stigchel:** Conceptualization, Funding acquisition, Project Administration, Supervision, Writing – Review & Editing

Disclosure statement

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

Funding

This project was supported by an ERC Consolidator Grant [grant number ERC-CoG-863732] awarded to Stefan Van der Stigchel.

Institutional review board statement

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the Faculty Ethics Review Board of the Faculty of Social and Behavioural Sciences at Utrecht University. The approval is filed under number 21-0297.

Informed consent statement

Informed consent was obtained from all subjects involved in the study. Written informed consent was obtained from the patient(s) to publish this paper.

Open practices statement

The data presented in this study are now available in Open Science Framework at https://osf.io/cjgaf/. The experiment was not pre-registered.

ORCID

S. Böing http://orcid.org/0000-0002-2410-8708

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Anwyl-Irvine, A. L., Armstrong, T., & Dalmaijer, E. S. (2022). MouseView.js: Reliable and valid attention tracking in web-based experiments using a cursor-directed aperture. *Behavior Research Methods*, *54*(4), 1663–1687. https://doi.org/10.3758/s13428-021-01703-5

Arnoult, M. (1956). Familiarity and recognition of nonsense shapes. *Journal of Experimental Psychology*, *51*, 269–276. https://doi.org/10.1037/h0047772

Ballard, D., Hayhoe, M., & Pelz, J. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*, 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Bambrah, V., Hsu, C.-F., Toplak, M. E., & Eastwood, J. D. (2019). Anticipated, experienced, and remembered subjective effort and discomfort on sustained attention versus working memory tasks. *Consciousness and Cognition*, *75*, 102812. https://doi.org/10.1016/j.concog.2019.102812

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4 (arXiv:1406.5823). arXiv. http://arxiv.org/abs/1406.5823.

Blalock, L. D. (2015). Stimulus familiarity improves consolidation of visual working memory representations. *Attention, Perception, & Psychophysics*, *77*(4), 1143–1158. https://doi.org/10.3758/s13414-014-0823-z

Böing, S., Brink, A. F. T., Ruis, C., Schielen, Z. A., Van Den Berg, E., Biesbroek, J. M., Nijboer, T. C. W., & Van Der Stigchel, S. (2025). Inspecting the external world: Memory capacity, but not memory self-efficacy, predicts offloading in working memory. *Journal of Clinical and Experimental Neuropsychology*. https://doi.org/10.1080/13803395.2024.2447263

Böing, S., Ten Brink, A. F., Hoogerbrugge, A. J., Oudman, E., Postma, A., Nijboer, T. C. W., & Van der Stigchel, S. (2023). Eye movements as proxy for visual working memory usage: Increased reliance on the external world in Korsakoff Syndrome. *Journal of Clinical Medicine*, *12*(11), Article 11. https://doi.org/10.3390/jcm12113630

Chung, Y. H., Brady, T. F., & Störmer, V. S. (2024). Meaningfulness and familiarity expand visual working memory capacity. *Current Directions in Psychological Science*, *33*(5), 275–282. https://doi.org/10.1177/09637214241262334

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain* (p. 85) [PhD thesis]. Montreal: McGill University, Montreal.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2016). *Working memory capacity: Classic edition*. Psychology Press.

Draschkow, D., Kallmayer, M., & Nobre, A. C. (2021). *When natural behavior engages working memory*. https://doi.org/10.1016/j.cub.2020.11.013

Droll, J., & Hayhoe, M. (2008). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology. Human Perception and Performance*, *33*(6), 1352–1365. https://doi.org/10.1037/0096-1523.33.6.1352

Faul, F., Erdfelder, E., Lang, L.-G., & Buchner, A. (2009). *G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences*. https://link.springer.com/article/10.3758BF03193146.

Fu, W.-T., & Gray, W. D. (2000). Memory versus perceptual-motor tradeoffs in a blocks world task. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (p. 22). https://escholarship.org/uc/item/5mn4s2zt.

Gilbert, S. (2015). Strategic use of reminders: Influence of both domain-general and task-specific metacognitive confidence, independent of objective memory ability. *Consciousness and Cognition*, *33*, 245–260. https://doi.org/10.1016/j.concog.2015.01.006

Gilbert, S. J. (2023). *Cognitive offloading is value-based decision making: Modelling cognitive effort and the expected value of memory.*

Gonthier, C. (2021). Charting the diversity of strategic processes in visuospatial short-term memory. *Perspectives on Psychological Science*, *16*(2), 294–318. https://doi.org/10.1177/1745691620950697

Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science*, *28*(3), 359–382. https://doi.org/10.1207/s15516709cog2803_3

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review*, *113*(3), 461–482. https://doi.org/10.1037/0033-295X.113.3.461

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, *74*(9), 1477–1496. https://doi.org/10.1177/17470218211008060

Hoogerbrugge, A. J., Strauch, C., Böing, S., Nijboer, T. C. W., & Van der Stigchel, S. (2024). Just-in-time encoding into visual working memory is contingent upon constant availability of external information. *Journal of Cognition*, *7*(1), Article 1. https://doi.org/10.5334/joc.364

Hoogerbrugge, A. J., Strauch, C., Nijboer, T. C. W., & Van der Stigchel, S. (2023). Don't hide the instruction manual: A dynamic trade-off between using internal and external templates during visual search. *Journal of Vision*, *23*(7), 14. https://doi.org/10.1167/jov.23.7.14

Janssen, C. P., & Gray, W. D. (2012). When, what, and how much to reward in reinforcement learning-based models of cognition. *Cognitive Science*, *36*(2), 333–358. https://doi.org/10.1111/j.1551-6709.2011.01222.x

Kardan, O., Adam, K. C. S., Mance, I., Churchill, N. W., Vogel, E. K., & Berman, M. G. (2020). Distinguishing cognitive effort and working memory load using scale-invariance and alpha suppression in EEG. *NeuroImage*, *211*, 116622. https://doi.org/10.1016/j.neuroimage.2020.116622

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*(6), 661–679. https://doi.org/10.1017/S0140525X12003196

Kvitelashvili, S., & Kessler, Y. (2024). *Memory at will: Investigating voluntary utilization of visual working memory capacity*. https://doi.org/10.21203/rs.3.rs-3834000/v1

Logie, R. H. (2023). Strategies, debates, and adversarial collaboration in working memory: The 51st Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, *76*(11), 2431–2460. https://doi.org/10.1177/17470218231194037

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Luria, R., Sessa, P., Gotler, A., Jolicœur, P., & Dell'Acqua, R. (2010). Visual short-term memory capacity for simple and complex objects. *Journal of Cognitive Neuroscience*, *22*(3), 496–512. https://doi.org/10.1162/jocn.2009.21214

Luria, R., & Vogel, E. K. (2011). Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia*, *49*(6), 1632–1639. https://doi.org/10.1016/j.neuropsychologia.2010.11.031

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Melnik, A., Schüler, F., Rothkopf, C. A., & König, P. (2018). The world as an external memory: The price of saccades in a sensorimotor task. *Frontiers in Behavioral Neuroscience*, *12*, 253. https://doi.org/10.3389/fnbeh.2018.00253

Meyerhoff, H. S., Grinschgl, S., Papenmeier, F., & Gilbert, S. J. (2021). Individual differences in cognitive offloading: A comparison of intention offloading, pattern copy, and short-term memory capacity. *Cognitive Research: Principles and Implications*, *6*(1), 34. https://doi.org/10.1186/s41235-021-00298-x

Morgan, P., Patrick, J., Waldron, S., King, S., & Patrick, T. (2009). Improving memory after interruption: Exploiting soft constraints and manipulating information access cost. *Journal of Experimental Psychology. Applied*, *15*(4), 291–306. https://doi.org/10.1037/a0018008

Morrison, A. B., & Richmond, L. L. (2020). Offloading items from memory: Individual differences in cognitive offloading in a short-term memory task. *Cognitive Research: Principles and Implications*, *5*(1), 1. https://doi.org/10.1186/s41235-019-0201-4

Morrison, A. B., Rosenbaum, G. M., Fair, D., & Chein, J. M. (2016). Variation in strategy use across measures of verbal working memory. *Memory & Cognition*, *44*(6), 922–936. https://doi.org/10.3758/s13421-016-0608-9

Patrick, J., Morgan, P. L., Smy, V., Tiley, L., Seeby, H., Patrick, T., & Evans, J. (2015). The influence of training and experience on memory strategy. *Memory & Cognition*, *43*(5), 775–787. https://doi.org/10.3758/s13421-014-0501-3

Poppenk, J., Köhler, S., & Moscovitch, M. (2010). Revisiting the novelty effect: When familiarity, not novelty, enhances memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1321–1330. https://doi.org/10.1037/a0019900

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Ricker, T. J., & Cowan, N. (2010). Loss of visual working memory within seconds: The combined use of refreshable and non-refreshable features. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(6), 1355–1368. https://doi.org/10.1037/a0020356

Risko, E. F., & Dunn, T. L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. *Consciousness and Cognition*, *36*, 61–74. https://doi.org/10.1016/j.concog.2015.05.014

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Sahakian, A., Gayet, S., Paffen, C. L. E., & Van der Stigchel, S. (2023). Mountains of memory in a sea of uncertainty: Sampling the external world despite useful information in visual working memory. *Cognition*, *234*, 105381. https://doi.org/10.1016/j.cognition.2023.105381

Scarampi, C., & Gilbert, S. J. (2020). The effect of recent reminder setting on subsequent strategy and performance in a prospective memory task. *Memory (Hove, England)*, *28*(5), 677–691. https://doi.org/10.1080/09658211.2020.1764974

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434

Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In D.H. Spieler & E. Schumacher (Eds.), *New methods in cognitive psychology* (pp. 4–31). Psychology Press.

Somai, R. S., Schut, M. J., & Van der Stigchel, S. (2020). Evidence for the world as an external memory: A trade-off between internal and external visual memory storage. *Cortex*, *122*, 108–114. https://doi.org/10.1016/j.cortex.2018.12.017

Storm, B. C., Stone, S. M., & Benjamin, A. S. (2017). Using the Internet to access information inflates future use of the Internet to access other information. *Memory (Hove, England)*, *25*(6), 717–723. https://doi.org/10.1080/09658211.2016.1210171

Van der Stigchel, S. (2020). An embodied account of visual working memory. *Visual Cognition*, *28*(5–8), 414–419. https://doi.org/10.1080/13506285.2020.1742827

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research*

Methods, *49*(2), 653–673. https://doi.org/10.3758/s13428-016-0721-5

Vandierendonck, A. (2021). On the utility of integrated speed-accuracy measures when speed-accuracy trade-off is present. *Journal of Cognition*, *4*(1), 1–26. https://doi.org/10.5334/joc.154

Vandierendonck, A. (n.d.). *Note on the usage of the linear integrated speed-accuracy score (LISAS)*. https://www.researchgate.net/profile/Andre-Vandierendonck/publication/330483132_lisas-usage-notepdf/data/5c41ffba458515a4c72f8a5a/lisas-usage-note.pdf?origin=publication_list.

Waldron, S. M., Patrick, J., Morgan, P. L., & King, S. (2007). Influencing cognitive strategy by manipulating information access. *The Computer Journal*, *50*(6), 694–702. https://doi.org/10.1093/comjnl/bxm064

Xie, W., & Zhang, W. (2018). Familiarity speeds up visual short-term memory consolidation: electrophysiological evidence from contralateral delay activities. *Journal of Cognitive Neuroscience*, *30*(1), 1–13. https://doi.org/10.1162/jocn_a_01188

Xie, W., & Zhang, W. (2023). *Effortfulness of visual working memory: gauged by physical exertion*. https://oce-ovid-com.proxy.library.uu.nl/article/00004785-202307000-00015/HTML.

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*(4), 423–428. https://doi.org/10.1111/j.1467-9280.2009.02322.x