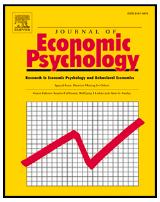
Journal of Economic Psychology 96 (2023) 102622

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep



Predicting serial position effects and judgment errors in retrospective evaluations from memory recall[☆]

Janina A. Hoffmann [a,b,*], Ann-Katrin Hosch [c,b]

[a] *University of Bath, United Kingdom*
[b] *University of Konstanz, Germany*
[c] *University of Bremen, Germany*

ARTICLE INFO

Dataset link: https://osf.io/baqv7/

Keywords:
Retrospective evaluation
Judgment
Episodic memory
Free recall

ABSTRACT

When forming global impressions in retrospect, the first, the last, and the most outstanding experience often have a lasting impact on the final evaluation of an event, as most prominently captured in the peak-end rule. Such serial position effects in impression formation provide indirect evidence that individuals reconstruct their evaluations by retrieving previous experiences from memory, instead of updating their impression online. Yet, latest work sheds doubt on the ability to predict global evaluations from memory retrieval on the individual level. In three experiments, we aim to quantify how much variability in retrospective evaluations can be attributed to memory retrieval by relating serial position effects in retrospective averaging judgments to serial recall curves from memory. The experiments revealed serial position effects in memory recall and corresponding, but less consistent effects in averaging judgments, demonstrating that individuals better recalled and more heavily weighted the first and last item. For long sequences, memory recall permitted to predict individuals' averaging error to a moderate to strong degree, even if individuals were unaware of number recall as a potential averaging strategy (Experiment 2). Yet, shorter sequences fail to evidence the same relationship, possibly because individuals attempt to apply more optimal averaging strategies (Experiment 3). We discuss retrieval patterns as markers for distinct evaluation strategies.

Making accurate decisions in daily life often necessitates recalling previous experiences from memory. When deciding which new employee to hire, for instance, personnel managers may recall at the time of the decision how well several job candidates performed on a standardized interview and form a global impression of each candidate. Indeed, evidence converges that retrieval from memory plays a key role across a broad range of decision tasks from performance evaluation to preferential choice to economic games (Bornstein & Norman, 2017; Duncan & Shohamy, 2016; Hoffmann, von Helversen, & Rieskamp, 2014; Hoffmann, von Helversen, Weilbächer, & Rieskamp, 2018; Murty, FeldmanHall, Hunter, Phelps, & Davachi, 2016; Volstorf, Rieskamp, & Stevens, 2011). The serial position effect in evaluation vividly demonstrates this interplay between memory recall and decision making: Candidates on TV song contests are evaluated more favorably if they performed in the first or the last position — those positions for which performance is remembered best (Page & Page, 2010). Judgment and decision research has documented such serial position effects in a variety of evaluation contexts from music and sports competitions (Antipov & Pokryshevskaya, 2017; Bruine de Bruin, 2005; Bruine de Bruin & Bruin, 2006) to political elections (Blom-Hansen, Elklit, Serritzlew, & Riis Villadsen, 2016; Chen, Simonovits, Krosnick, & Pasek, 2014) to commercials (Li, 2010) to the evaluation of word lists (Aldrovandi, Poirier, Kusev, & Ayton, 2015).

[☆] This research was supported by German Research Foundation Grant HO 5715/1-1. Data, code and materials are available for download at the Open Science Framework, https://osf.io/baqv7/.

[*] Correspondence to: Department of Psychology, University of Bath, 10W Claverton Campus, Bath, BA27AY, United Kingdom.
E-mail address: j.a.hoffmann@bath.ac.uk (J.A. Hoffmann).

https://doi.org/10.1016/j.joep.2023.102622
Received 29 July 2022; Received in revised form 17 February 2023; Accepted 8 March 2023

Position effects in evaluation and choice may arise because people evaluate the options retrospectively after having experienced all of them, thereby reconstructing their evaluations from memory (Payne, Bettman, & Johnson, 1992; Slovic, 1995). Retrieval processes may underpin how individuals summarize single experiences to a global impression, too. The peak-end rule posits that individuals form an overall impression of an event, such as the painfulness of a medical treatment, depending upon selected memorable episodes, such as the most painful experience during the treatment and the final moments of the treatment (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993). Similarly, the climate crisis is evaluated as more pressing, the more recently and the more often individuals experienced extreme weather events, such as floods or heat waves (Sambrook, Konstantinidis, Russell, & Okan, 2021). Although observational and experimental evidence converges towards the view that memory retrieval plays a key role in global evaluations (Bornstein & Norman, 2017; Duncan & Shohamy, 2016; Murty et al., 2016), previous work has rarely quantified how much variability in global evaluations can be attributed to successful or failed retrieval and to distinct retrieval processes. In this article, we first determine how much of the variability in global evaluations can be explained at maximum by retrieval processes from memory. Next, we assess which retrieval component, serial position, distinctiveness, or retrieval order, affects evaluation biases the most. Finally, we investigate to what degree those findings translate to more applied evaluation tasks, here price evaluations of shopping baskets.

1. Phenomena in episodic memory and their effects on global evaluations

When individuals are asked to recall events, facts, or items from memory, not all these episodes are equally well remembered. Individuals remember episodes better, the more frequently and recently they have encountered these episodes and the more distinctive the episodes are compared to surrounding events (Anderson & Schooler, 1991; Brown, Neath, & Chater, 2007). Individual recall patterns are typically marked by so-called serial position effects: Items presented first or last in a sequence have a recall advantage over items in mid-position, coined a primacy or recency effect, respectively (Glanzer & Cunitz, 1966; Murdock, 1962). Better recall for mid-position items is rather the exception than the rule (Jones & Oberauer, 2013). Primacy effects are explained by enhanced storage of items presented first in long-term memory (Roediger & Crowder, 1976), whereas recency effects result from reading out the last items from a short-term working memory store (Glanzer, 1972).

When forming global average impressions of sequentially presented episodes, such as averaging a series of numbers, converging evidence suggests that people are impressively accurate (Brezis, Bronfman, & Usher, 2015; Pollard, 1984), but are susceptible to consistent biases. Recent research suggests that individuals tend to underestimate sequence averages, if the variance is high or distributions are left-skewed, potentially indicating that single numbers are compressed on a mental number line (Olschewski, Newell, Oberholzer, & Scheibehenne, 2021). Memory phenomena, akin to the ones observed in free recall, have been also noted in sequential averaging. First, if numbers are presented sequentially, the last numbers in the sequence strongly influence participants' estimated average suggesting recency effects (Weiss & Anderson, 1969). Second, delaying the judgment diminishes the recency effect (Aldrovandi et al., 2015; Montgomery & Unnava, 2009). Third, numbers presented first can influence the estimated average as well (Aldrovandi et al., 2015; Hendrick & Costantini, 1970), but these effects are typically less pronounced. Yet, it still remains an open question if serial position effects in averaging reflect the same phenomena as serial recall from memory (Mason, Brown, Ward, & Farrell, 2021).

When forming summative impressions of sequentially presented episodes, such as summing up a series of numbers, individuals are far less accurate than in averaging and typically underestimate the sum (Goswami, Greenberg, & Schley, 2021; McGowan, Denny, & Lunn, 2022; Van Ittersum, Pennings, & Wansink, 2010). Longer sequences are more strongly underestimated than shorter ones (McGowan et al., 2022), but also presentation order influences the degree of underestimation. Sequences ending with high numbers are typically underestimated more than sequences ending with low numbers, indicating a reversed serial position effect (Olschewski et al., 2021; Scheibehenne, 2019). Yet, while underestimation of sums has been traced back to perceptual and working memory processes (Dougherty & Hunter, 2003; Olschewski et al., 2021; Scheibehenne, 2019; Sprenger & Dougherty, 2006), reversed serial position effects lack a profound cognitive basis, making it worthwhile to explore their connection to recall from memory.

2. Mechanisms underpinning numerical summation and averaging

Limited recall from memory has been offered as one explanation for serial position effects observed in global evaluations, most prominently encapsulated in the peak-end rule (Kahneman et al., 1993). This heuristic states that individuals overweigh recent and intense experiences in their global evaluations. Consequently, the peak-end rule explains recency and distributional effects, but cannot account for primacy effects.

Recently, researchers have argued that perceptional processes supplement retrospective retrieval and working memory updating (Brezis et al., 2015; Olschewski et al., 2021; Scheibehenne, 2019). To identify the independent contribution of retrieval and perceptual processes, previous work has formulated distinct mathematical models to infer these weighing processes (Mason et al., 2021; Scheibehenne, 2019). Optimally, individuals should weigh all numbers in a sequence equally to determine the sequence average (the optimal or expected-value model). Thus, participants' predicted judgment $\hat{y}_p$ for each sequence p is the average of all presented numbers in this sequence $x_{p,s}$, $\hat{y}_p = \frac{\sum^{n_p} x_{p,s}}{n_p}$, with n_p representing the amount of presented numbers and $x_{p,s}$ the magnitude of the number at each serial position s.

If primacy and recency effects modulate how much numbers presented at first and last position influence averaging, participants should weight each presented number with a positional weight w_s to form the average.

$$\hat{y}_p = \frac{\sum^{n_p} w_s * x_{p,s}}{n_p} \tag{1}$$

In quadratic weighting models, the positional weights w_s are a quadratic function of this serial position s to allow numbers presented first and last to influence individuals' averages more.

$$w = \beta_1 s^2 + \beta_2 s + \beta_0 \tag{2}$$

Quadratic scaling models account additionally for a biased perceptual representation of the mental number line by compressing the magnitude of single numbers with a scaling parameter α.

$$\hat{y}_p = \frac{\sum^{n_p} w_s * x_{p,s}^{\alpha}}{n_p} \tag{3}$$

Matching the idea that symbolic numbers are compressed on the mental number line, Scheibehenne (2019) found that underestimation of summation judgments was best captured by such a quadratic scaling model, comprising both a power law to compress single numbers and a quadratic function to account for position effects.

Yet, those modeling attempts have rarely measured both recall performance and estimation behavior on the individual level and were thus unable to jointly model recall and estimation strategies (but see Mason et al., 2021). Initial attempts to predict averaging behavior from serial recall probabilities have found thus far limited support for a memory-based recall model compared to an expected value averaging model (Mason et al., 2021). In three experiments, Mason et al. presented participants with seven numbers drawn form a uniform distribution and asked them to estimate their average and to recall them. A recall-based model fared only better at predicting estimations than an expected value model, if participants first had to recall the numbers from memory, but not if participants first predicted sequence averages. Possibly, expected-value models described averaging more parsimoniously than recall-based models because recency effects were absent in serial recall whenever averaging preceded the recall task. Further, the short sequence length may have allowed some individuals to rehearse all numbers in short-term memory, leading to a mixture of recall strategies in evaluation and serial recall. Finally, the uniform distribution made it harder to diagnose primacy and recency effects in averaging, clear markers of memory effects in evaluations. We address these shortcomings in our experiments by (a) investigating longer sequences so that we increase the need to recall information from episodic memory and reduce mixed recall strategies and by (b) drawing the numbers systematically from different distributions with a sufficient number of trials (Scheibehenne, 2019) so that we can estimate serial position curves on the individual level for number recall and averaging. This alterations allow us to precisely estimate if and how strongly individual recall strategies shape averaging and summation judgments.

2.1. Aims and analysis rationale of the current study

The current study aims to determine to what degree recall processes from memory can predict retrospective global evaluations in the form of averaging (Experiment 1 to 3) and summation judgments (Experiment 3). Experiment 1 sought to determine how much variability in judgment biases can maximally be explained by recall from memory. Experiment 2 assessed to what degree recall processes still predict averaging if individuals are unaware of number recall as a potential averaging strategy. Experiment 3 then pursued to generalize these findings to summation judgments and the shorter sequences typically used in previous studies (Mason et al., 2021). In doing so, we evaluate for each experiment four aspects that allow us to infer how much recall from memory contributes to retrospective global evaluations and which recall components can account for the observed relationship.

First, we expect to replicate classical serial position effects when individuals recall numbers from memory (Glanzer & Cunitz, 1966; Murdock, 1962). For this purpose, we construct for each experiment a serial position curve, displaying how likely participants recall a number that had been presented at a certain position (Fig. 2). We also report which serial position participants most often retrieved in their first successful retrieval attempt.

Second, if memory recall underpins retrospective evaluations, we expect to replicate the finding that judgment accuracy depends on the serial position at which high or low numbers are displayed (distributional effects). Specifically, higher numbers presented at the beginning and at the end of the sequence should lead to a stronger overestimation of the sequence average than presenting high numbers in the middle of the sequence. In line with Scheibehenne (2019), we calculated the amount of over- and underestimation as the difference between each participant's estimate $\hat{y}$ and the correct sequence mean y proportional to the sequence mean bias $= \frac{(\hat{y}-y)}{y}$. We test for distributional effects on over- and underestimations with a one-way ANOVA with the independent factor distributional shape of the number sequence. A planned contrast compares if individuals overestimate sequences with higher numbers in the end (linearly increasing and U-shaped sequences) more than sequences with higher numbers in the beginning or middle (linearly decreasing and inversely U-shaped sequences).

Third, individual recall patterns should correlate with the positional weight persons assign to each number in averaging (or summation) judgments. To identify these positional weights, we first estimated the parameter values for the optimal model, the quadratic weighting model, the quadratic scaling model, and a baseline model by minimizing the deviance, $D = -2 * LL$. The baseline model accounted for guessing behavior by proposing that a single value best represented participants' judgments. All models assumed a normally distributed error around the mean and an estimated standard deviation σ. We next compared all models using the Bayesian Information Criterion (BIC, Schwarz, 1978) that penalizes the models for the number of free parameters. This model

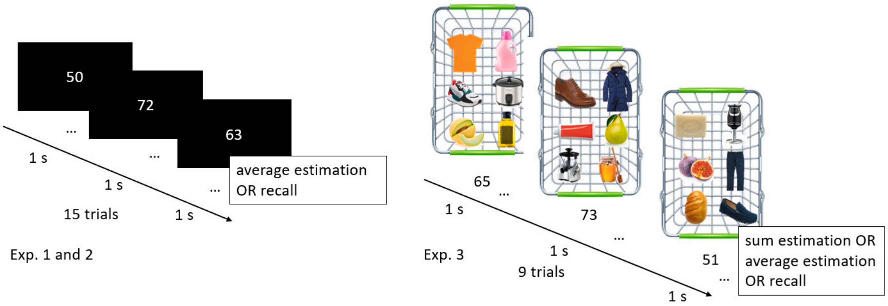


Fig. 1. Example of stimuli and flow in each single trial in experiment 1 and 2 (left plot) and Experiment 3 (right plot).

comparison allows to test whether it is necessary to impose positional weights or a compression of the mental number line. We inferred the positional weights from the quadratic weighting model that described the majority of participants best in Experiment 1. To quantify the impact of recall, we finally calculated for each participant the correlation between the positional weights and their corresponding recall percentages for serial recall and first recall (see Fig. 3).

Fourth, individual recall patterns should ultimately allow to predict which sequences individuals over- or underestimate beyond the observed distributional effects. For this purpose, we constructed expected over- and underestimations for three memory indicators: For serial position effects, we weighted each sequence number, standardized to the correct average, with their respective serial recall percentage. For first recall effects, we similarly weighted each standardized sequence number with their corresponding percentage of being recalled first. For extremity effects, we determined how much each number differed from their neighbors when sorted by magnitude. The relative distance was then used as a weight to predict over- and underestimations. The three predictors were entered into a linear mixed model, controlling for random effects of participants and distribution, to infer which memory phenomena explain accuracy in averaging judgments.

3. Experiment 1

Experiment 1 maximized the overlap between recall strategies and averaging strategies by randomly asking participants after seeing a number sequence to estimate its average or to recall the numbers. To directly observe serial position effects in averaging, we manipulated the distributional shape of the number sequence and presented random, linearly increasing, linearly decreasing, U-Shaped, and inversely U-Shaped sequences.

3.1. Method

3.1.1. Participants

50 students from the University of Konstanz (35 females, $M_{\text{Age}} = 23.0$, $SD_{\text{Age}} = 4.0$) participated in the experiment. We decided a priori to collect data from 50 participants based upon an expected correlation between number recall and judgment accuracy, r = .6, one-sided CI 95% = [.42; 1]. Participants received an hourly wage of 8 € for their participation plus a performance-based bonus (M = 1.16 €, SD = 0.36 €). We excluded 4 participants because they estimated implausibly low (below 10) or high values (above 100) in the judgment task or had several missings.

3.1.2. Design

Each participant estimated the average of 40 number sequences and recalled the numbers for another 40 number sequences. Each number sequence consisted of 15 numbers (see Fig. 1 for a display of the stimuli and experimental flow). Out of the 40 sequences, 8 sequences were generated for each distribution with three restrictions: all numbers had to range from 10 to 99, none of the numbers was repeated in the sequence, and the standard deviation of the sequences was comparable across distributions. In random sequences, all numbers within a certain range appeared at all positions within the sequence with the same probability. In linearly increasing sequences, lower numbers more likely appeared in the beginning of the sequence than at the end, whereas in linearly decreasing sequences they appeared more likely in the end. In U-shaped sequences, lower numbers more likely appeared in the middle, whereas in inverse U-shaped sequences they were more likely to appear in the beginning and the end of the sequence. A number could appear in several sequences, but was not repeated twice within the same sequence.

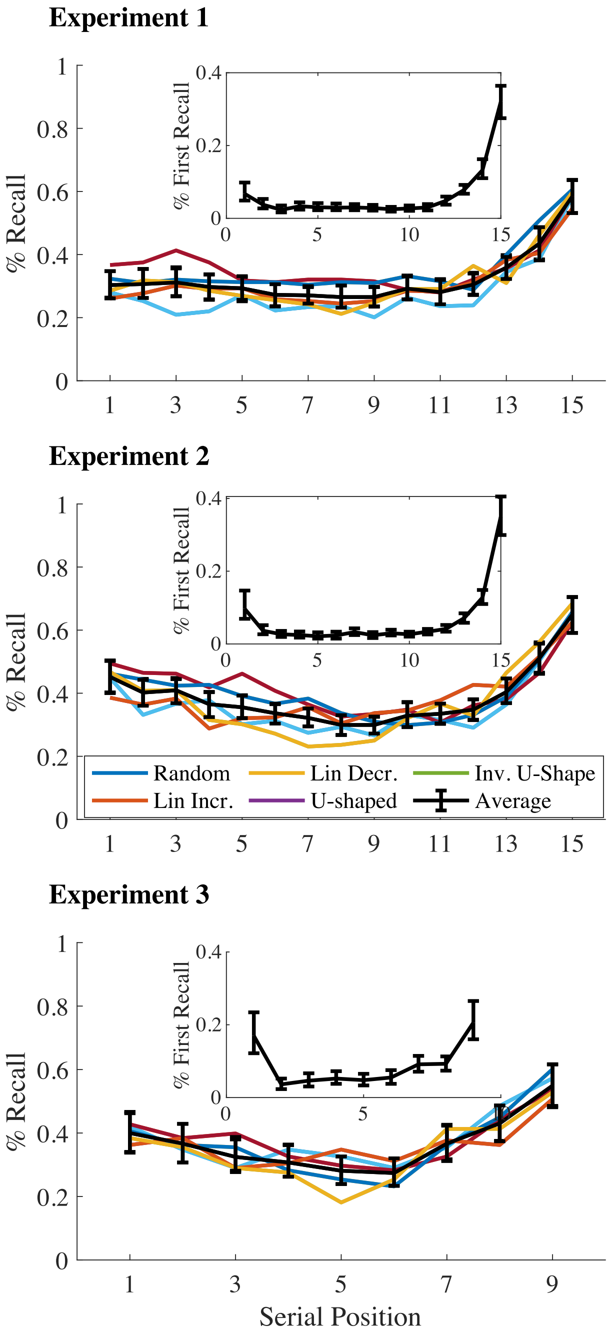


Fig. 2. Recall percentages for each serial position in the number sequence in Experiment 1 (upper plot), Experiment 2 (middle plot), and Experiment 3 (lower plot). The black line shows average serial recall percentages, colored lines display serial recall percentages separately for each distribution. Smaller plots display the percentages that a number is recalled first, depending upon their serial position. Error bars plot bootstrapped confidence intervals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.1.3. Procedure

Participants were presented with 80 sequences consisting of 15 numbers each. The numbers were sequentially displayed on a computer screen for 1 s each and a pause of 0.2 s between them. After all numbers had been displayed, participants were instructed to either estimate the average of the sequence or to recall as many numbers as they could. Averaging and recall trials were randomly interspersed so that participants could not predict which task they will have to perform.

Table 1
Descriptive statistics for recall and judgment bias in experiment 1, 2, and 3 (averaging and summation). SD in parentheses; 95% confidence interval in brackets.

	Exp. 1	Exp. 2	Exp. 3 (Averaging)	Exp. 3 (Summation)
N participants	46	46	46	(see Averaging)
Recall % correct	32.3 (9.2)	38.8 (7.6)	36.7 (12.7)	(see Averaging)
Judgment bias				
Overall	−0.02 (0.02)	−0.01 (0.02)	−0.00 (0.02)	−0.02 (0.03)
Random	−0.03 (0.03)	−0.02 (0.03)	−0.00 (0.04)	−0.01 (0.06)
Linear inc.	0.01 (0.04)	0.01 (0.03)	0.01 (0.04)	−0.01 (0.06)
Linear dec.	−0.05 (0.04)	−0.03 (0.04)	−0.01 (0.05)	−0.02 (0.04)
U-shaped	0.01 (0.04)	0.00 (0.03)	0.00 (0.05)	−0.02 (0.05)
Inverse U	−0.04 (0.04)	−0.02 (0.03)	−0.02 (0.04)	−0.02 (0.10)
Corr. Pos. weights				
Numbers recalled	0.60 [0.46; 0.70]	0.38 [0.19; 0.55]	0.13 [−0.09; 0.35]	0.01 [−0.19; 0.21]
First recall	0.61 [0.52; 0.69]	0.30 [0.13; 0.45]	0.11 [−0.08; 0.30]	0.17 [−0.03; 0.35]

Note. Linear Inc. = Linear Increasing; Linear Dec. = Linear Decreasing; Inverse U = Inversely U-Shaped; Corr. Pos. Weights = Correlation with position weights.

Table 2
Model fits and classifications to each model in experiment 1, 2, and 3 (averaging and summation). SD in parentheses.

	Exp. 1	Exp. 2	Exp. 3 (Averaging)	Exp. 3 (Summation)
Model	Model fit (BIC)			
Baseline	356 (6)	356 (7)	131 (5)	197 (6)
Optimal	227 (28)	196 (31)	76 (14)	162 (20)
Quadratic weighting	224 (27)	196 (31)	77 (15)	163 (21)
Quadratic scaling	226 (27)	198 (31)	79 (16)	164 (20)
	Model classification (N)			
Baseline	0	0	0	3
Optimal	24	24	29	30
Quadratic weighting	19	21	11	11
Quadratic scaling	3	1	6	2

Note. BIC = Bayesian Information Criterion.

To incentivize accuracy, participants could earn and lose points. For averaging trials, the number of points depended upon the deviance between participant's estimates $\hat{y}$ and the true sequence mean y, standardized by the standard deviation of the sequence (SD):

$$\text{Points} = 6 - \frac{(\hat{y} - y)^2}{\text{SD}} \tag{4}$$

In number recall, participants earned one point for each correctly recalled number, but lost one point for each incorrect recall attempt. The total amount of points was divided by 200 and paid out in €. At the end of every 20th trial, participants received feedback about how many points they earned in total, how many of the numbers they had recalled correctly and incorrectly, and how accurate their estimates were in the preceding trials, as measured by the mean absolute difference of their estimate from the true sequence mean.

3.2. *Results*

3.2.1. *Recall performance*

On average, participants recalled 32.3% of the numbers correctly (SD = 9.2). They rarely recalled a number not presented to them ($M = 0.7$, $SD = 0.7$) or repeatedly retrieved the same number within the sequence ($M = 0.1$, $SD = 0.1$). Fig. 2 displays the serial position curve averaged across participants and as a function of the distribution. Overall, participants are more likely to retrieve the numbers presented last correctly, suggesting a strong recency effect. However, participants only remembered numbers presented in the first position better for linearly increasing sequences, suggesting that the primacy effect was largely absent in serial recall. Measuring which serial position participants most often retrieved in their first successful retrieval attempt (% First Recall in the smaller upper graph in Fig. 2) revealed that participants often report the last numbers first, but also more likely report the first number in a sequence first.

3.2.2. *Judgment bias and serial position curves in estimation*

On average, participants slightly underestimated the sequence mean (M = −0.02, SD = 0.02) with 39 out of 46 participants underestimating the sequence mean.

Descriptively, participants underestimated linearly decreasing and inversely U-shaped distributions more than linearly increasing or U-shaped distributions. Table 1 summarizes descriptive statistics for recall, over- and underestimations and correlations for all

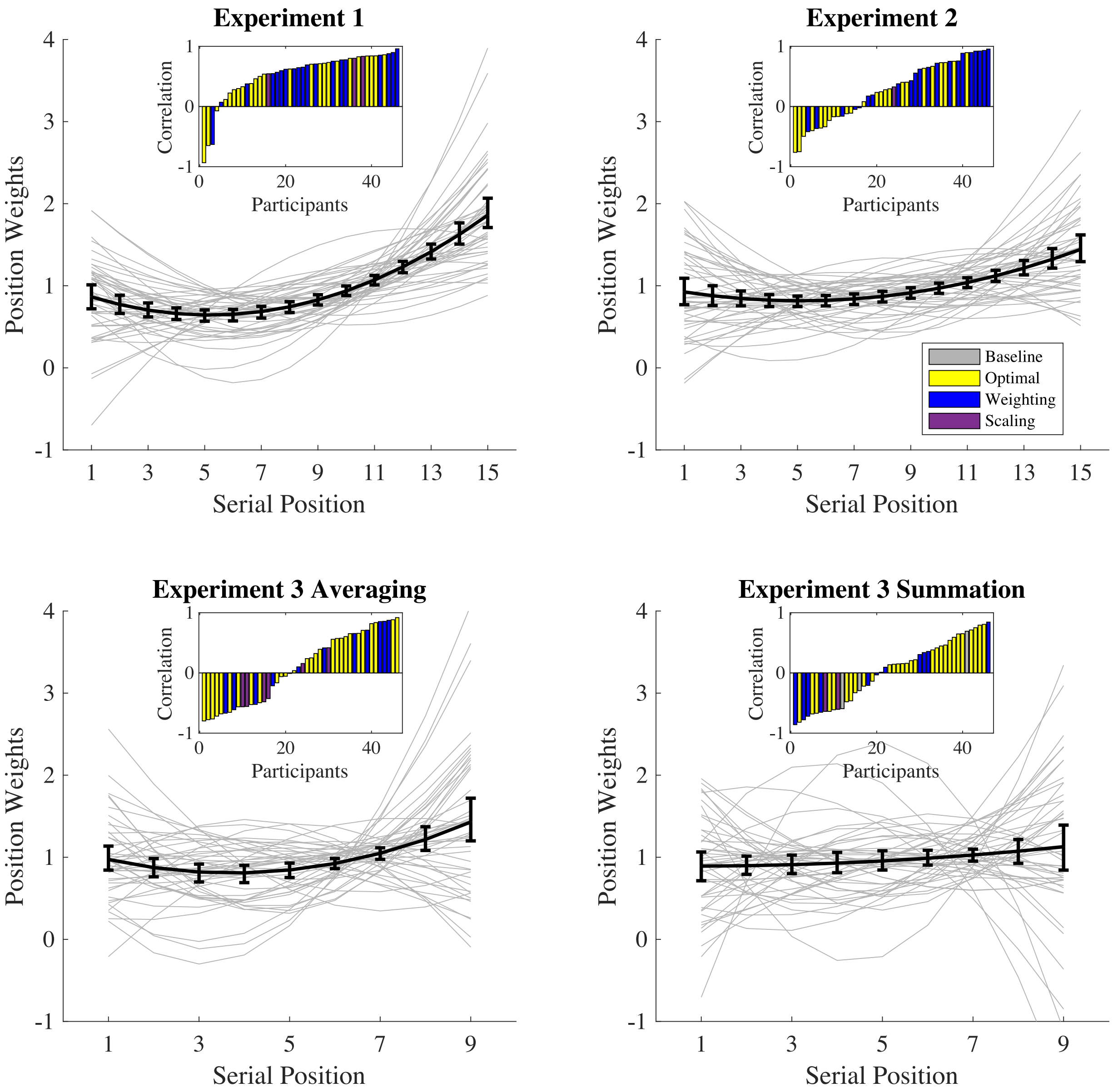


Fig. 3. Relative weight of each presented number in the estimation task depending on its serial position in Experiment 1 (upper left plot), Experiment 2 (upper right plot), and Experiment 3 (Averaging in lower left plot and Summation in lower right plot). The black line shows average weights, gray lines show weights for each individual participant. Error bars plot bootstrapped confidence intervals. The smaller bar plot depicts the correlation between recall percentages and the estimated position weight for each participant, sorted by the magnitude of the correlation.

experiments. A Greenhouse–Geisser corrected repeated-measures ANOVA on judgment bias with distributional shape as a within-factor suggested that the magnitude of judgment bias varied depending on distributional shape, $F(3.2,\ 143.4) = 32.4$, $p = 0.001$, $\eta^2 = 0.32$. A planned contrast indicated that participants overestimated sequences with higher numbers in end more than sequences with higher numbers in the middle or only the beginning, $\Delta M = 0.03$, $SE = 0.00$, $t(45) = 8.46$, $p = 0.001$.

On average, the quadratic weighting and the quadratic scaling model better captured participants' judgments than the baseline model, but fared only slightly better than the optimal model (see Table 2 for model fits and number of participants classified to each model). The quadratic scaling model did not provide a better fit to participants' judgments than the quadratic weighting model. Classifying participants to the best-fitting model, that is the model with the lowest BIC, provided a similar picture: A similar number of participants was best described by the quadratic weighting and the optimal model, whereas the quadratic scaling model only described judgments of a few participants best. The baseline model did not describe a single participant best.

Fig. 3 depicts the position weights from the quadratic weighting model for each participant (gray lines) and the average position weights across all participants (black line). Reflecting the serial position curve from recall, the positional weighing curves suggest

Table 3
Linear mixed model summary comparing the model with memory effects as predictors against a model with only random intercept. SE in parentheses.

Model	Effect	Exp. 1	Exp. 2	Exp. 3 (Averaging)	Exp. 4 (Summation)
Memory	*D*	−3610	−5078	−1637	−468
	Serial position	0.11 (0.11)	0.32 (0.08)	0.10 (0.11)	0.32 (0.26)
	First recall	0.18 (0.04)	0.00 (0.03)	0.09 (0.05)	0.09 (0.11)
	Extremity	−0.01 (0.05)	−0.06 (0.03)	0.03 (0.07)	0.01 (0.16)
Intercept-only	*D*	−3565	−5047	−1625	−462

Note. *D* = Deviance.

a strong recency effect with numbers presented last influencing the mean estimates more than numbers presented in the middle of the sequence. In addition, the average curves suggest a slight primacy effect.

3.2.3. Predicting judgments from memory recall

To what degree do the positional weighing curves from estimation reflect recall of numbers from memory? On average, the positional weights and recall percentages were strongly positively correlated, $r = 0.60$, 95% CI $= [0.46; 0.70]$, $t(45) = 7.5$, $p < 0.001$. Correlations ranged from −0.93 to 0.96 with 4 out of 46 participants showing negative correlations between the positional weights and recall percentages. Similarly, how likely participants were to report a number first correlated strongly with the positional weights, $r = 0.61$, 95% CI $= [0.52; 0.69]$.

In a last step, we investigated if serial position, first recall, or extremity effects best predict why individuals overestimate some sequences, but underestimate others. Table 3 summarizes model deviances in effects of each predictor in all experiments. The linear mixed model involving participant-level predictors for serial position, first recall, and extremity effects outperformed an intercept-only model, $\chi^2(3) = 45.3$, 0.001. The magnitude of the number recalled first influenced the degree of over- and underestimations the most, but over- and underestimations were not predicted by serial position effects or extremity effect. Taken together, if individuals recalled higher numbers first in free recall, they were more likely to overestimate the sequence average in the averaging task.

3.3. Discussion

Experiment 1 provides strong evidence that the order in which individuals recall information from memory can systematically affect their global evaluation of these number sequences. We replicated previously found distributional effects on averaging showing that if higher numbers appear last, sequence estimates are overestimated (Aldrovandi et al., 2015). Position weights indicated that these over- and underestimations occurred, because numbers presented first or last were overweighted compared to numbers presented in the middle and the amount of overweighting correlated strongly with how likely participants were to recall the number displayed at a position from memory. Finally, which number individuals recalled first predicted directly how strongly they over- or underestimated sequence averages. In sum, if individuals expect to perform a global evaluation task interchangeably with recall from memory, their global evaluations can be well predicted from the information they recall from memory, in particular the number they recalled first.

4. Experiment 2

Experiment 1 sought to quantify how much variability in global evaluations can maximally be explained by recall from memory on the individual level because recent work has questioned a consistent relationship between recall and averaging judgments. For this purpose, we intentionally increased the overlap between the global evaluation and the recall task. Retrospective evaluations and recall may have displayed a strong relationship in Experiment 1 because individuals relied upon memory recall as a strategy for forming global evaluations. Yet, it is likely that in typical evaluation tasks, individuals rarely recall numbers strategically from memory, but rather average the numbers intuitively or update sequence averages on the fly (Brezis et al., 2015; Brusovansky, Vanunu, & Usher, 2019). Meta-cognitive research similarly points towards the possibility that individuals adjust their study time and learning strategy when they are aware of recall phenomena, for instance, after experiencing prior training (Castel, 2008; Murphy, Friedman, & Castel, 2022). Thus, individuals may less often recall numbers strategically from memory in evaluation tasks when recall is not highlighted as a feasible averaging strategy. Experiment 2 intended to determine the degree to which memory recall still explains variability in retrospective evaluations if individuals resort to their default strategy to form evaluations instead.

4.1. Method

4.1.1. Participants

50 students from the University of Konstanz (39 females, $M_{\text{Age}} = 24.0$, $SD_{\text{Age}} = 5.0$) were recruited to participate in the experiment. We decided in advance to collect data from 50 participants based upon an expected correlation r = .35 and the lower boundary confidence interval, one-sided CI-95% = [.12; 1]. Participant received a payment of 8 € per hour and a performance-dependent bonus (M = 1.25 €, SD = 0.23 €). We excluded 4 participants because they estimated implausibly low or high values in the judgment task or had several missings.

4.1.2. Materials, design and procedure

In Experiment 2, all participants first completed the 40 estimation trials before they moved on to the 40 recall trials. Participants were informed in the beginning of each block which task they should perform. Hence, the design reduced the need to keep the numbers in memory in estimation tasks. All other experimental details were left unchanged.

4.2. Results

4.2.1. Recall performance

Compared to Experiment 1, participants recalled a slightly higher percentage of the numbers correctly (38.8%, $SD = 7.6$). They rarely recalled a number not presented to them ($M = 0.6$, $SD = 0.5$) or repeatedly retrieved the same number within the sequence ($M = 0.1$, $SD = 0.1$). Fig. 2 displays the average serial position curve as well as recall probabilities for each individual participant. Participants recalled more likely numbers presented first and last, suggesting a more pronounced primacy effect compared to Experiment 2.

4.2.2. Judgment bias and serial position curves in estimation

Like in Experiment 1, participants on average underestimated the sequence mean slightly ($M = -0.01$, $SD = 0.02$) with 34 out of 46 participants underestimating the sequence mean. Participants underestimated linearly decreasing and inversely U-shaped distributions more than linearly increasing or U-shaped distributions (Table 1). A Greenhouse–Geisser corrected repeated-measures ANOVA indicated that distributional shape influenced the magnitude of judgment bias, $F(3.3,148.5) = 12.8$, $p = 0.001$, $\eta^2 = 0.17$. A planned contrast comparing linearly decreasing and inverse U-shaped distributions to linearly increasing and U-shaped distributions similarly suggested that participants more likely underestimated sequences presenting smaller numbers first and last, $\Delta M = 0.01$, $SE = 0.002$, $t(45) = 5.78$, $p = 0.001$.

As in Experiment 1, the quadratic weighting model captured participants' judgments better than the quadratic scaling model or the baseline model (see Table 2 for model fits and number of participants classified to each model). The quadratic weighting model fared as well as the optimal model. Classifying participants to the best-fitting model according to the BIC suggested that only a few participants were better described by the quadratic scaling modeling and most participants were either best described by quadratic weighting or the optimal model. As in Experiment 1, the positional weighing curves derived from this quadratic weighting model demonstrate a strong recency effect, whereas the primacy effect is less pronounced.

4.2.3. Predicting judgments from memory recall

We quantified the relationship between recall processes and averaging judgments by correlating the positional weights with the recall percentages. On average, the positional weights and recall percentages were positively correlated, but —as expected —the correlation declined somewhat compared to Experiment 1, $r = 0.38$, 95% CI $=\left[0.19;\ 0.55\right]$, $t(45) = 3.8$, $p < 0.001$. Correlations ranged from −0.76 to 0.95 with 16 out of 46 participants showing negative correlations between the positional weights and recall percentages. Finally, the likelihood of recalling a number first also correlated moderately with positional weights, $r = 0.30$, 95% CI $=\left[0.13;\ 0.45\right]$.

Finally, we tested which memory components explain over- and underestimations of the sequence average (see Table 3). A linear mixed model including serial position, first recall, and extremity effects fared better than a linear mixed model that controlled for random effects of participants and distribution, $\chi^2(3) = 31.1$, 0.001. Which serial position participants remembered best predicted over- and underestimations, but not which number they recalled first on average or which number was the most extreme one. In sum, serial position effects predicted over- and underestimations better than first recall or extreme numbers.

4.3. Discussion

Overall, Experiment 2 replicated the finding in Experiment 1 that recall can explain a decent amount in variability in evaluations, although effect sizes were smaller. We found that which distribution individuals had to judge influence the amount of over- and underestimations systematically. On the individual level, which number the person was more likely to recall correlated moderately with how strongly this number influenced global evaluations. Finally, the probability to recall a number at this serial position predicted why individuals overestimated linearly increasing and U-shaped functions compared to linearly decreasing and inverse U-shaped functions, whereas which number individuals recalled first was not associated with averaging biases. One reason is possibly that participants developed independent strategies for memory retrieval and averaging. In memory recall, this leads to a better encoding and recall of items presented at the first position. Possibly, this more pronounced primacy effect explains why the serial position effect better accounts for evaluation biases than first recall in Experiment 2. Taken together, these results suggest that recall from memory is at least partly a viable strategy that individuals engage in when solving retrospective evaluation tasks.

5. Experiment 3

So far, our experiments demonstrated that individual recall patterns can successfully predict retrospective evaluations when individuals have to average abstract number sequences. Experiment 3 intended to generalize these findings in three ways. First, on many occasions individuals have to form global impressions by summing up the presented evidence, for instance, in valuation. Individuals typically underestimate sums more than averages (Olschewski et al., 2021), but also distributional effects may differ between summation and averaging judgments. Inferred positional weights from summation judgments suggest that individuals underestimate, instead of overestimate, the influence of numbers presented first or last (Scheibehenne, 2019), potentially indicating that individual recall from memory plays a different role in summation than in averaging. Second, we contextualized the tasks by transferring them into a shopping context in which individuals had to recall the price of a shopping basket, estimate the average price of a shopping basket or the summed price of all baskets. Third, previous work has typically studied recall and retrospective evaluations in shorter sequences, but failed to find a consistent relationship (Mason et al., 2021). For this reason, we shortened the presented sequences to 9 sequentially presented baskets with their corresponding prices.

5.1. Method

5.1.1. Participants

50 students of the University of Bremen (39 females, $M_{\text{Age}} = 25.6$, $SD_{\text{Age}} = 6.4$) participated in the experiment. Participant were reimbursed with course credit and had the chance to participate in a lottery for a 10 € shopping voucher. We excluded 3 participants because they provided implausibly low or high estimates for sequence averages on more than five trials (as defined in a preregistration as 1 SD below the smallest number in the sequence or 1 SD above the highest number in the sequence). We excluded 1 participant because he was assigned to a previous version of the experiment.

5.1.2. Materials, design and procedure

The experiment was preregistered on the Open Science Framework (https://osf.io/f9egs). Stimuli in this task were shopping baskets containing six products (see Fig. 1 for a display of the stimuli and experimental flow). Each of the items belonged to a different product category (clothing, shoes, tools, kitchen devices and food). All categories contained six different products out of which one was randomly displayed in the shopping basket. In each sequence, nine shopping baskets were shown one after another and each basket was displayed with a price for the full basket underneath. Single products were not associated with a price tag. Each basket was presented for 1000 ms and followed by the next basket, until a sequence of nine baskets (and their respective prices) had been presented. Prices were selected in the same fashion as the numbers in previous experiments and randomly allocated to the shopping baskets.

Participants completed three tasks in randomized task order with each task containing 15 number sequences: An averaging task, a summation task, and a cued recall task. In the averaging task, participants estimated the average price of the 9 presented baskets. In the summation task, participants estimated the summed price of the 9 presented baskets. In the cued recall task, all 9 shopping baskets from the current sequence were simultaneously displayed on screen and participants could type in the recalled price underneath each shopping basket. Participants were instructed which task they had to complete for the next 15 sequences. As in the previous experiments, the 15 number sequences were generated from five different distributions (decreasing, increasing, U-shaped, inversely U-shaped and uniform distribution) and 3 sequences were generated for each distribution.

In order to incentivize accuracy, participants earned points in every trial. The closer participants' estimates $\hat{y}$ were to the true average (or true sum) y, the more points they earned.

$$\text{Points} = 4 - \frac{|\hat{y} - y|}{y} \tag{5}$$

Participants also earned 1 point for each correctly recalled number. The total amount of points won determined participants' chances to win the lottery.

5.2. Results

5.2.1. Recall performance

Despite the shorter sequence of numbers to remember, participants recalled on average a similar percentage of numbers correctly as in Experiment 2 (36.7%, $SD = 12.7$). They seldom recalled a number not presented to them ($M = 0.8$, $SD = 0.6$) or committed repetition errors ($M = 0.5$, $SD = 0.4$). Fig. 2 displays the average serial position curve as well as recall percentages for each individual participant. As in previous experiments, participants recalled more likely numbers presented first and last, but the serial position effect was less pronounced.

5.2.2. Judgment bias and serial position curves in averaging and summation

Participants on average underestimated the sequence sum ($M = -0.02$, $SD = 0.03$), but did not underestimate the sequence mean ($M = -0.00$, $SD = 0.02$). For sequence averages, a planned contrast comparing linearly decreasing and inverse U-shaped distributions to linearly increasing and U-shaped distributions suggested that participants more likely underestimated sequences presenting smaller numbers first and last, $\Delta M = 0.01$, $SE = 0.004$, $t(45) = 2.35$, $p = 0.02$, although the main effect of distribution was not significant and smaller than in the previous experiments, $F(2.9, 128.8) = 2.5$, $p = 0.067$, $\eta^2 = 0.04$ (Table 1 for descriptive statistics). Whether smaller numbers were presented first or last did not influence how individuals judged sums, $F(2.5, 112.4) = 0.3$, $p = 0.821$, $\eta^2 = 0.00$.

In both tasks, averaging and summation, more participants were classified to the optimal model than to the quadratic weighting or quadratic scaling model (see Table 2 for model fits and number of participants classified to each model). The optimal model also possessed a slightly lower BIC than the quadratic weighting or the quadratic scaling model.

5.2.3. Predicting judgments from memory recall

Fig. 3 displays the relationship between positional weights in averaging and estimation with recall probabilities assessed in memory recall. Although positional weights from averaging and serial position weights were still on average positively correlated, this average correlation was considerably lower than in the previous experiments, $r = 0.13$, 95% CI $= [-0.09; 0.35]$, and not distinguishable from a correlation of 0, $t(45) = 1.2$, $p = 0.244$. 21 out of 46 participants showed negative correlations between the positional weights and recall probabilities with individual correlations ranging from −0.80 to 0.92. Which number participants recalled first, likewise correlated only lightly with the position weights, $r = 0.11$, 95% CI $= [-0.08; 0.30]$.

Summations were on average uncorrelated with serial position weights, $r = 0.01$, 95% CI $= [-0.19; 0.21]$, $t(45) = 0.1$, $p = 0.920$. As for averaging judgments, 20 out of 46 participants displayed negative correlations with individual correlations ranging from −0.86 to 0.84. Positional weights were only mildly correlated with first recall, $r = 0.17$, 95% CI $= [-0.03; 0.35]$. Interestingly, positional weights for averaging correlated on average negatively with positional weights for summation, $r = -0.41$, but confidence intervals were very large and covered 0, 95% CI $= [-0.71; 0.03]$. Only 15 out of 46 participants who put a high weight on a certain position in averaging also weighted the same position higher in summation ($r > .30$); instead, 25 out of 46 participants who put a high weight on a certain position in averaging underweighted the same position in summation ($r < -.30$).

At last, we attempted to predict over- and underestimations observed in averaging and summation judgments from the observed memory phenomena (see Table 3). Over- and underestimations in summation judgments could not be predicted from the serial position effect, first recall, or the extremity effect, as indicated by the finding that a linear mixed model including these phenomena did not fare better than a linear mixed model that controlled for random effects of participants, $\chi^2(3) = 6.1$, $p = 0.106$. In contrast, over- and underestimations in averaging were better explained by a linear mixed model comprising these memory phenomena compared to a linear mixed model containing only random effects for participant and distribution, $\chi^2(3) = 11.9$, $p = 0.008$. If participants recalled a higher number first, they more likely overestimated the sequence mean. Overestimations did, however, not increase with serial position, nor with extreme numbers.

5.3. Discussion

When generalizing the numerical averaging results to a consumer choice context, we only partially replicated the key findings from previous studies. A planned linear contrast indicated that participants underestimated linearly decreasing and inverted U-Shaped price distributions and overestimated linearly increasing and U-Shaped price distributions, however, we did not replicate the main effect of distribution on judgment error. Further, the correlations between first (and serial) recall and positional weights could not be distinguished from 0, nor was averaging of shopping basket prices better described by a quadratic model with positional weights, although individual recall strategies somewhat predicted over- and underestimations. When individuals had to sum those prices, over- and underestimations were less systematic and neither depended on the distribution, nor were correlated with individual recall curves, potentially indicating that memory phenomena contribute less to summing up evidence than to forming an average opinion.

6. General discussion

Forming retrospectively an aggregated impression of an event, person, or group from single observations is a common daily task, ranging from evaluating the pleasure of a music performance to the overall stress experienced during the day. These retrospective impressions often provide the evaluation basis for subsequent economic behavior, such as the willingness-to-pay for future concerts or the choice of a recreational activity after work. Influential theories in impression formation, such as the peak-end-rule, identified selective retrieval of previously experienced episodes as the key determinant of global evaluations, but so far, little research has directly traced how retrieval processes contribute to global evaluations. Our study fills this gap by simultaneously assessing individual recall patterns and averaging behavior for number sequences and prices of shopping baskets.

6.1. Serial position effects in retrospective evaluations

In which sequential order high and low numbers were presented influenced how strongly individuals over- and underestimated the sequence average across Experiments 1 and 2. Linearly increasing and U-shaped sequences were more likely overestimated than linearly decreasing or inversely U-shaped sequences. Further, quadratic weighting models captured participants' averaging judgments well in Experiments 1 and 2, indicating that numbers presented first and last were more heavily weighted in participants' judgments. In contrast, a reverse pattern with an underestimation of increasing and U-shaped sequences is displayed in summation judgments (Scheibehenne, 2019). In Experiment 3, averaging and summation judgments were less systematically over- or underestimated and participants' judgments were best described by an optimal weighting model that assigned all prices equal importance, matching averaging behavior in short sequences (Mason et al., 2021).

When predicting over- and underestimations, we found that which number individuals recalled first predicted over-and underestimations in averaging best for Experiment 1 and 3, but serial recall percentages better predicted over-and underestimations in Experiment 2. This discrepancy can possibly be explained by a stronger primacy effect in Experiment 2. A primacy effect is consistently displayed in first recall in all three experiments, but absent or less pronounced in serial recall in Experiment 1 and 3. Taken together, these results imply that not only last, but also first impressions are essential for predicting average global evaluations from memory.

6.2. Sequence length as mediator

The diverging results of Experiments 1 and 2 versus Experiment 3 hint towards the possibility that sequence length serves as mediator for optimal averaging (Brezis et al., 2015). While individuals may weigh all observations equally in short sequences (Experiment 3, Mason et al., 2021), whenever the sequence length surpasses the individual's digit span, as for longer sequences of 15–20 numbers, individuals may begin to overweight first and last numbers more heavily, eliciting serial position effects in evaluations. In line with this idea, we found in Experiment 1 and 2 that positional weights correlated positively with a medium to strong effect size with the likelihood to recall a number from this position correctly or at the first retrieval attempt. This finding dovetails previous insights from dual-process modeling proposing that individuals shift to intuitive processing for long number sequences (Brezis et al., 2015).

Still, Experiment 3 introduced more variations than sequence length which may have contributed to the observed lower correlation between evaluation and recall in Experiment 3. First, the prices were displayed underneath the picture of a shopping basket and these shopping baskets might have drawn attention to them. Attending to the baskets is, however, unnecessary to form an accurate average impression in the averaging task and a heightened amount of concentration would further not be needed if people form these averages intuitively (Brezis et al., 2015). Second, individuals performed a cued recall task in Experiment 3. They saw all shopping baskets on the screen in random order and had to recall their price. This change in recall task may have provoked a shift during the learning process and participants potentially focused more on associating each basket with its price than the individual prices. Yet, serial position curves do not display a systematic shift in recall patterns. Third, we varied the order of the recall, summation and averaging task between participants instead of randomly interspersing the trials (Experiment 1) or keeping it constant (Experiment 2). However, we could not detect a systematic effect of order. Taken together, we consider sequence length as the most plausible reason for the weakened relationship between recall and evaluation.

6.3. Limitations and future directions

When individuals had to recall and evaluate shorter sequences in Experiment 3, individual recall patterns were only positively associated with position weights in averaging and summation for half of the participants. Indeed, some participants displayed strong negative correlations between recall patterns and averaging or summation behavior, likely emerging from reversed serial position effects in retrospective evaluations (Scheibehenne, 2019). Reversed serial position effects are rarely reported in recall from long-term memory, but simulations with short-term memory models indicate that items in the middle may gain a retrieval advantage if their positional markers overlap strongly (Jones & Oberauer, 2013). In summation and averaging, this advantage may arise from presenting highly similar numbers in the middle that are less distinguishable from one another, leading to the hypothesis that positional weighting schemes in summation and averaging might vary with distributional shape. Alternatively, positional weighting curves might serve as an easily accessible marker for distinct evaluation strategies in averaging and summation. While the classical serial position effect may characterize intuitive averaging strategies that hinge upon retrospective recall from memory, reversed serial position effects may characterize attempts to serially update the presented numbers (Brezis et al., 2015). Future research may seek to elicit both types of retrieval patterns within the same person to systematically study strategy switches as a function of evaluation task and complexity.

So far, our research has emphasized the question to what degree recalling numbers within the same sequence underpins the ability to form a global impression of the sequence. It is possible, however, that individuals may also falsely recall a number from previous sequences and these intrusions may hamper global evaluations (Carretti, Cornoldi, & Pelegrina, 2007). To explore the impact of intrusions on global evaluations, we determined if participants were more prone to overestimate the sequence average when the preceding sequence possessed a higher average (compared to a lower average) than the current sequence. Overall, participants did not overestimate the sequence average more when preceding sequences possessed a higher average ($M = -0.01$, $SD = 0.02$) compared to a lower average, $M = -0.01$, $SD = 0.02$, $t\ (45) = 0.3$, $p = 0.752$. Further, the number of intrusions in number recall did not correlate with the difference in judgment biases between preceding sequences with high and low means, $r = -0.09$. These results indicate that intrusions may play a smaller role in retrospective evaluative judgments than in number recall. It might be interesting to explore the role of intrusions in retrospective evaluations more carefully in future work and explore if individuals suppress intrusions more successfully in global evaluations than in memory recall or updating (Carretti et al., 2007).

Serial position effects in impression formation have been oftentimes conceived as an illustrative example of fundamental judgment fallacies. Trusting one's memory can, however, also serve as an adaptive evaluation strategy when coping with naturalistic environments in which high gains are rare, but small losses appear frequently (Anderson & Schooler, 1991; Stewart, Chater, & Brown, 2006). The shopping task highlights that understanding these strategies is key to better explain fundamental financial decisions, such as budgeting or household decision making (Hilbert, Noordewier, & van Dijk, 2022; Thaler, 1999). Memory contributions to mental accounting strategies might prove a promising avenue for future research, potentially shifting the focus from summative to subtractive retrospective evaluations.

Data availability

Data, code and materials are available for download at the Open Science Framework, https://osf.io/baqv7/.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.joep.2023.102622.

References

Aldrovandi, S., Poirier, M., Kusev, P., & Ayton, P. (2015). Retrospective evaluations of sequences. *Experimental Psychology*, *62*(5), 320–334. http://dx.doi.org/10.1027/1618-3169/a000301.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396–408. http://dx.doi.org/10.1111/j.1467-9280.1991.tb00174.x.

Antipov, E. A., & Pokryshevskaya, E. B. (2017). Order effects in the results of song contests: Evidence from the eurovision and the new wave. *Judgment and Decision Making*, *12*(4), 415–419.

Blom-Hansen, J., Elklit, J., Serritzlew, S., & Riis Villadsen, L. (2016). Ballot position and election results: Evidence from a natural experiment. *Electoral Studies*, *44*, 172–183. http://dx.doi.org/10.1016/j.electstud.2016.06.019.

Bornstein, A. M., & Norman, K. A. (2017). Reinstated episodic context guides sampling-based decisions for reward. *Nature Neuroscience*, *20*(7), 997–1003. http://dx.doi.org/10.1038/nn.4573.

Brezis, N., Bronfman, Z. Z., & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Scientific Reports*, *5*, 1–11. http://dx.doi.org/10.1038/srep10415.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539–576. http://dx.doi.org/10.1037/0033-295X.114.3.539.

Bruine de Bruin, W. (2005). Save the last dance for me: Unwanted serial position effects injury evaluations. *Acta Psychologica*, *118*(3), 245–260. http://dx.doi.org/10.1016/j.actpsy.2004.08.005.

Bruine de Bruin, W., & Bruin, W. B. D. (2006). Save the last dance II : Unwanted serial position effects in figure skating judgments. *Acta Psychologica*, *123*(3), 299–311. http://dx.doi.org/10.1016/j.actpsy.2006.01.009.

Brusovansky, M., Vanunu, Y., & Usher, M. (2019). Why we should quit while we're ahead: When do averages matter more than sums? *Decision*, *6*(1), 1–16. http://dx.doi.org/10.1037/dec0000087.supp.

Carretti, B., Cornoldi, C., & Pelegrina, S. L. (2007). Which factors influence number updating in working memory? The effects of size distance and suppression. *British Journal of Psychology*, *98*(1), 45–60. http://dx.doi.org/10.1348/000712606X104175.

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory and Cognition*, *36*(2), 429–437. http://dx.doi.org/10.3758/MC.36.2.429.

Chen, E., Simonovits, G., Krosnick, J. A., & Pasek, J. (2014). The impact of candidate name order on election outcomes in north dakota. *Electoral Studies*, *35*, 115–122. http://dx.doi.org/10.1016/j.electstud.2014.04.018.

Dougherty, M. R., & Hunter, J. E. (2003). Hypothesis generation, probability judgment, and individual differences in working memory capacity. *Acta Psychologica*, *113*(3), 263–282. http://dx.doi.org/10.1016/S0001-6918(03)00033-7.

Duncan, K. D., & Shohamy, D. (2016). Memory states influence value-based decisions. *Journal of Experimental Psychology: General*, *145*(11), 1420–1426. http://dx.doi.org/10.1037/xge0000231.

Glanzer, M. (1972). Storage mechanisms in recall. In *Psychology of learning and motivation* (pp. 129–193). http://dx.doi.org/10.1016/S0079-7421(08)60441-7.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, *4*, 351–360, Retrieved fromhttp://www.sciencedirect.com/science/article/pii/S0022537166800440papers3://publication/uuid/6B619C14-CABE-428D-BD63-C2D72F60830B.

Goswami, I., Greenberg, A. E., & Schley, D. (2021). Undersum bias. *SSRN Electronic Journal*, *23529*(2), 1–45. http://dx.doi.org/10.2139/ssrn.3954282.

Hendrick, C., & Costantini, A. F. (1970). Number averaging behavior: A primacy effect. *Psychonomic Science*, *19*(2), 121–122. http://dx.doi.org/10.3758/BF03337452.

Hilbert, L. P., Noordewier, M. K., & van Dijk, W. W. (2022). Financial scarcity increases discounting of gains and losses: Experimental evidence from a household task. *Journal of Economic Psychology*, *92*, Article 102546. http://dx.doi.org/10.1016/j.joep.2022.102546.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2014). Pillars of judgment: How memory abilities affect performance in rule-based and exemplar-based judgments. *Journal of Experimental Psychology: General*, *143*(6), 2242–2261. http://dx.doi.org/10.1037/a0037989.

Hoffmann, J. A., von Helversen, B., Weilbächer, R. A., & Rieskamp, J. (2018). Tracing the path of forgetting in rule abstraction and exemplar retrieval. *The Quarterly Journal of Experimental Psychology*, *71*, 2261–2281, Retrieved from http://journals.sagepub.com/doi/10.1177/1747021817739861.

Jones, T., & Oberauer, K. (2013). Serial-position effects for items and relations in short-term memory. *Memory*, *21*(3), 347–365. http://dx.doi.org/10.1080/09658211.2012.726629.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, *4*(6), 401–405. http://dx.doi.org/10.1111/j.1467-9280.1993.tb00589.x.

Li, C. (2010). Primacy effect or recency effect ? A long-term memory test of super bowl commercials. *Journal of Consumer Behaviour*, *44*(October 2009), 32–44. http://dx.doi.org/10.1002/cb.

Mason, A., Brown, G. D., Ward, G., & Farrell, S. (2021). Memory-based and online strategies in retrospective evaluations. (pp. 1–45). http://dx.doi.org/10.31234/osf.io/uaqs5.

McGowan, F., Denny, E., & Lunn, P. (2022). Consistent underestimation in the intuitive summation of monetary amounts. *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.4021708.

Montgomery, N. V., & Unnava, H. R. (2009). Temporal sequence effects: A memory framework. *Journal of Consumer Research*, *36*(1), 83–92. http://dx.doi.org/10.1086/595278.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482–488. http://dx.doi.org/10.1037/h0045106.

Murphy, D. H., Friedman, M. C., & Castel, A. D. (2022). Metacognitive control, serial position effects, and effective transfer to self-paced study. *Memory and Cognition*, *50*(1), 144–159. http://dx.doi.org/10.3758/s13421-021-01204-y.

Murty, V. P., FeldmanHall, O., Hunter, L. E., Phelps, E. A., & Davachi, L. (2016). Episodic memories predict adaptive value-based decision-making. *Journal of Experimental Psychology: General*, *145*(5), 548–558. http://dx.doi.org/10.1037/xge0000158.

Olschewski, S., Newell, B. R., Oberholzer, Y., & Scheibehenne, B. (2021). Valuation and estimation from experience. *Journal of Behavioral Decision Making*, *34*(5), 729–741. http://dx.doi.org/10.1002/bdm.2241.

Page, L., & Page, K. (2010). Last shall be first: A field study of biases in sequential performance evaluation on the idol series. *Journal of Economic Behaviour and Organization*, *73*(2), 186–198. http://dx.doi.org/10.1016/j.jebo.2009.08.012.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, *43*(1), 87–131. http://dx.doi.org/10.1146/annurev.ps.43.020192.000511.

Pollard, P. (1984). Intuitive judgments of proportions, means, and variances: A review. *Current Psychological Research & Reviews*, *3*(1), 5–18. http://dx.doi.org/10.1007/BF02686528.

Roediger, H. L., & Crowder, R. G. (1976). A serial position effect in recall of United States presidents. *Bulletin of the Psychonomic Society*, *8*(4), 275–278. http://dx.doi.org/10.3758/BF03335138.

Sambrook, K., Konstantinidis, E., Russell, S., & Okan, Y. (2021). The role of personal experience and prior beliefs in shaping climate change perceptions: A narrative review. *Frontiers in Psychology*, *12*, http://dx.doi.org/10.3389/fpsyg.2021.669911.

Scheibehenne, B. (2019). The psychophysics of number integration: Evidence from the lab and from the field.. *Decision*, *6*(1), 61–76. http://dx.doi.org/10.1037/dec0000089.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Slovic, P. (1995). The construction of preference. *American Psychologist*, *50*(5), 364–371.

Sprenger, A., & Dougherty, M. R. (2006). Differences between probability and frequency judgments: The role of individual differences in working memory capacity. *Organizational Behavior and Human Decision Processes*, *99*(2), 202–211. http://dx.doi.org/10.1016/j.obhdp.2005.08.002.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1–26. http://dx.doi.org/10.1016/j.cogpsych.2005.10.003, arXiv:arXiv:1011.1669v3.

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, *12*(3), 183–206. http://dx.doi.org/10.1002/(SICI)1099-0771(199909)12:3<183::AID-BDM318>3.0.CO;2-F.

Van Ittersum, K., Pennings, J. M., & Wansink, B. (2010). Trying harder and doing worse: How grocery shoppers track in-store spending. *Journal of Marketing*, *74*(2), 90–104. http://dx.doi.org/10.1509/jmkg.74.2.90.

Volstorf, J., Rieskamp, J., & Stevens, J. R. (2011). The good, the bad, and the rare: Memory for partners in social interactions. *PLoS ONE*, *6*(4), Article e18945. http://dx.doi.org/10.1371/journal.pone.0018945.

Weiss, D. J., & Anderson, N. H. (1969). Subjective averaging of length with serial presentation. *Journal of Experimental Psychology*, *82*(1), 52–63. http://dx.doi.org/10.1037/h0028028.