Not what u expect

Ortiz-Tudela, Javier; Nolden, Sophie; Pupillo, Francesco; Ehrlich, Isabelle; Schommartz, Iryna; Turan, Goezem; Shing, Yee Lee

Published in:
Journal of Experimental Psychology: General

DOI:
10.1037/xge0001367

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 8, 2160–2176
https://doi.org/10.1037/xge0001367

Not What U Expect: Effects of Prediction Errors on Item Memory

Javier Ortiz-Tudela[1], Sophie Nolden[1, 2], Francesco Pupillo[1], Isabelle Ehrlich[1], Iryna Schommartz[1, 2], Gözem Turan[1, 2], and Yee Lee Shing[1, 2, 3]

[1] Department of Psychology, Goethe University Frankfurt
[2] IDeA Center for Individual Development and Adaptive Education, Frankfurt am Main, Germany
[3] Brain Imaging Center, Goethe University Frankfurt

The characterization of the relationship between predictions and one-shot episodic encoding poses an important challenge for memory research. On the one hand, events that are compatible with our previous knowledge are thought to be remembered better than incompatible ones. On the other hand, unexpected situations, by virtue of their novelty, are known to cause enhanced learning. Several theoretical accounts try to solve this apparent paradox by conceptualizing prediction error (PE) as a continuum ranging from low PE (for expectation-matching events) to high PE (for expectation-mismatching ones). Under such a framework, the relationship between PE and memory encoding would be described by a U-shape function with higher memory performance for extreme levels of PE and lower memory for middle levels of PE. In this study, we tested the framework by using a gradual manipulation of the strength of association between scenes and objects to render different levels of PE and then tested for item memory of the (mis)matching events. In two experiments, in contrast to what was anticipated, recognition memory for object identity followed an inverted U-shape as a function of PE, with higher performance for intermediate levels of PE. Furthermore, in two additional experiments, we showed the relevance of explicit predictions at encoding to reveal such an inverted U pattern, thus providing the boundary conditions of the effect. We discussed our findings in light of existing literature relating PE and episodic memory, pointing out the potential roles of uncertainty in the environment, and the importance of the cognitive operations underlying encoding tasks.

Keywords: prediction, episodic memory, prior updating, prediction error

Supplemental materials: https://doi.org/10.1037/xge0001367.supp

The inherent regularities embedded in our environment enable the exploitation of repeating patterns to optimize information processing. Sand is usually brown, lightning often is followed by thunder, and cows are usually found in prairies. By abstracting the commonalities across experiences of daily life, our brains can predict later encounters with similar events. However, as our environment is not entirely deterministic and predictable, the use of patterns must be a dynamic process. This process entails both the exploitation of existing knowledge, to efficiently process predicted events, and the updating of the knowledge itself, in case of encountering unexpected situations. While there is accumulating evidence for predictive processing of sensory experiences (Friston, 2005, 2008; Keller & Mrsic-Flogel, 2018; Press et al., 2020; Rao & Ballard, 1999; Walsh et al., 2020), when it comes to the long-term memory consequences of encountering predicted and unpredicted events, empirical findings are mixed at best (Greve et al., 2017; Gronau & Shachar, 2015; Kafkas & Montaldi, 2018a; Ortiz-Tudela et al., 2018; Sinclair & Barense, 2018). In this study, we examined the episodic memory consequences of experiences that varied in levels of prediction violation.

This article was published Online First March 30, 2023.

Javier Ortiz-Tudela https://orcid.org/0000-0003-2844-2110

This work was supported by an ERC-Starting Grant (ERC-2018-StG-PIVOTAL-758898) awarded to Yee Lee Shing. The work of Yee Lee Shing was supported by the German Research Foundation (DFG; Project-ID 327654276, SFB 1315, "Mechanisms and Disturbances in Memory Consolidation: from to Synapses Systems") and the Hessisches Ministerium für Wissenschaft und Kunst (HMWK; project "The Adaptive Mind"). The work of Javier Ortiz-Tudela, Sophie Nolden, Francesco Pupillo was supported by the Goethe Research Academy for Early Career Researchers—Fokus Track A/B program: Promotion of Independent Grant Proposals.

Raw data and preregistrations can be found at https://osf.io/zawur/. Similarly, stimulation and analysis scripts to reproduce the results of this study can be found at: https://github.com/ortiztud/premup.

Javier Ortiz-Tudela served as lead for conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, validation, visualization, writing–original draft, and writing–review and editing and served in a supporting role for funding acquisition. Sophie Nolden served in a supporting role for writing–review and editing. Francesco Pupillo contributed equally to formal analysis and software, and served in a supporting role for methodology, visualization, and writing–review and editing. Isabelle Ehrlich served in a supporting role for validation and writing–review and editing. Iryna Schommartz served in a supporting role for conceptualization and writing–review and editing. Gözem Turan served in a supporting role for validation and writing–review and editing. Yee Lee Shing served as lead for funding acquisition, resources, and supervision and contributed equally to project administration and writing–review and editing. Sophie Nolden and Yee Lee Shing contributed to conceptualization equally.

Correspondence concerning this article should be addressed to Javier Ortiz-Tudela, Department of Psychology, Goethe University Frankfurt, Theodor-Adorno Platz, 1, 60323 Frankfurt am Main, Hessen, Germany. Email: ortiztudela@psych.uni-frankfurt.de

Previous research examining the interaction between predictions and episodic memory has revealed two seemingly incompatible sets of results. On the one hand, studies exploring prediction errors (PEs) postulate that PEs signal the need to upregulate learning as a means to gather new information that would improve future predictions. This notion stems partly from the literature that examines probabilistic/incremental learning driven by reward PE (e.g., Schultz, 1998), which has received considerable empirical support in human studies (see review in Shohamy & Adcock, 2010). In more recent years, the influence of PE has been extended and has been postulated to also modulate one-shot episodic encoding of new information (Henson & Gagnepain, 2010). Namely, situations that lead to high PEs may be strongly encoded to gain new information potentially relevant for future predictions. In contrast, situations rendering little or no PE contain mostly predictable information and thus the need for updating the internal model is reduced (Henson & Gagnepain, 2010). In other words, such PE-driven encoding hypothesis postulates that the need for learning, given high PE, is translated into enhanced episodic encoding at the time of the mismatch. In line with such postulation, recent evidence shows that events that are not well-predicted are *sometimes* better remembered than well-predicted ones (Brod et al., 2018; Greve et al., 2017; Kafkas & Montaldi, 2018b; Quent et al., 2021). For example, in Kafkas and Montaldi (2018b), participants learned associations between symbols and object categories. These associations could then be either violated or confirmed in a subsequent encoding phase. In a later recognition test, they showed that objects shown in violated trials were recollected better than those shown in non-violated trials (Kafkas & Montaldi, 2018b).

On the other hand, traditional accounts of episodic memory formation indicate that events that are congruent with one's knowledge are better remembered than those that are not (Craik & Tulving, 1975). This so-called *memory congruency effect* on memory has been replicated many times and across many different experimental setups (Bein et al., 2015; Brod & Shing, 2019; Craik & Tulving, 1975; Gronau & Shachar, 2015; Kaiser et al., 2015; Ortiz-Tudela et al., 2017; Staresina et al., 2009). For instance, in Ortiz-Tudela et al. (2017), participants performed a change detection task on real-world images in which the to-be-detected object was either semantically congruent or incongruent with the background scene. When a recognition memory test on the objects was administered, participants performed better for semantically congruent objects than for incongruent ones (Ortiz-Tudela et al., 2017). By definition, congruent events can, to some degree, be predicted from the contextual information (e.g., expecting to see food when entering a kitchen) and thus they should elicit low PE when compared with incongruent ones. As a consequence, enhanced memory of low PE events seems to stand in contrast to the PE-driven encoding hypothesis.

Assuming that predictions necessarily stem from stored knowledge, an influential theoretical model tries to resolve this apparent paradox of enhanced memory both for well-predicted events and for events giving rise to PE (Van Kesteren et al., 2012; see Press et al., 2020 for a similar discussion in the perceptual domain). This account states that two different neural mechanisms are responsible for the seemingly conflicting evidence. Within this framework, the medial temporal lobe (MTL) and medial prefrontal cortex (mPFC) are differentially involved in memory formation as a function of whether the new event was congruent or incongruent with previous knowledge. On the one hand, in the case of an encounter that is not in line with our expectations, the MTL creates a new detail-rich episodic memory of the current incongruent event. On the other hand, if an event is congruent with our previous knowledge, its pre-existing connections are strengthened by virtue of re-activation (i.e., resonance). This process, which is supported by the mPFC, makes congruent information more easily accessible at retrieval. Crucially, mPFC activation inhibits the MTL, reducing its involvement in creating the memories trace. Finally, for an encounter that is neither strongly congruent nor strikingly incongruent with our expectations, this model assumes that intermediate resonance with neocortical representations leads to weak activation of both mPFC and MTL, thus resulting in inefficient encoding and ultimately, poor memory (Van Kesteren et al., 2012). To sum up, this model predicts that the relationship between memory performance and PE, ranging from low, medium, to high PE, would be described by a U-shaped function. Namely, higher memory for extreme levels of PE than for intermediate ones.

Testing this U-shaped hypothesis requires the manipulation of PE beyond two levels (low PE/congruent vs. high PE/incongruent). However, there are still only a handful of studies with such manipulation. In one study, Greve et al. (2019) trained participants to establish associations between trial-unique pairs of objects and later re-arranged them to create congruent, unrelated, and incongruent pairs. After being exposed to the newly arranged pairs, participants' memory was tested with a recognition memory test. Results showed enhanced memory for both congruent and incongruent trials, compared to unrelated ones (Greve et al., 2019). In another study, Frank et al. (2018) manipulated pairs of stimuli to be either unrelated or to be strongly related to each other. Strongly related pairs were then met or violated, and memory for these encounters was measured. They found that pair-matching and pair-mismatching items were better retrieved than unrelated ones (Frank et al., 2018). These studies lend support to the theoretical account outlined above that considers PE as a continuum (Greve et al., 2017, 2019; Van Kesteren et al., 2012).

However, several important unresolved issues from existing findings remain. Previous studies relied on categorical experimental manipulations to render congruent, neutral, and incongruent events. However, a test of the postulated U-shape pattern along the full continuum of PE is lacking (although see Quent et al., 2021 for gradual quantitative assessment of location congruency). This would require an experimental setup that integrates prior knowledge and incoming events to render different levels of PE that are quantifiable. Indeed, even if the hypothesized U-shape would hold true, capturing a portion of the full range can lead to PE-advantages (Greve et al., 2017; Kafkas & Montaldi, 2018a), PE-impairments (Brod et al., 2013; Gronau & Shachar, 2015; Ortiz-Tudela et al., 2017), or no differences between PE levels (Ortiz-Tudela et al., 2018). Therefore, only having a full range of PE would allow a proper characterization of the postulated U-shape pattern.

A Bayesian perspective (e.g., Perfors et al., 2011) can be used as a theoretical framework to achieve such quantifiable PE manipulation. By considering knowledge as prior distributions about the most likely events to be encountered and the actual events as evidence to be contrasted with the priors, different levels of PE can be rendered by varying the strength of the prior from which the prediction is computed (Greve et al., 2019). In the present study, we adopted this approach across four experiments by developing a scene-object category learning task with varying levels of contingency; this learning task was followed by an encoding phase in which unique scene-object pairs were shown and finally by a retrieval phase with episodic memory tests.

It is important to also consider that there are other external factors that may modulate the effect of PE on episodic memory. Consolidation has been discussed as one such factor favoring better memory for events giving rise to PE. For example, Van Kesteren et al. (2012) found that schema effects for item recognition arise only 20 hr after encoding, which seems to suggest that the delay period between encoding and test can be crucial to uncover such effects (see also Sinclair & Barense, 2018). Moreover, the processes taking place at encoding, driven by the task that participants are actively carrying out, can also play a critical role. Yet, encoding tasks are often treated as cover tasks to keep participants engaged while they are exposed to certain regularities in the stimuli. In line with this idea, two recent studies showed that whether participants responded *before* the stimulus was shown (i.e., prediction) or after the stimulus had been shown (i.e., post-diction) impacted whether participants' pupils dilated in response to PE or not. More interestingly, the memory advantage that followed PE trials was only present when participants engaged into prediction but not into post-diction (Brod et al., 2018, 2020).

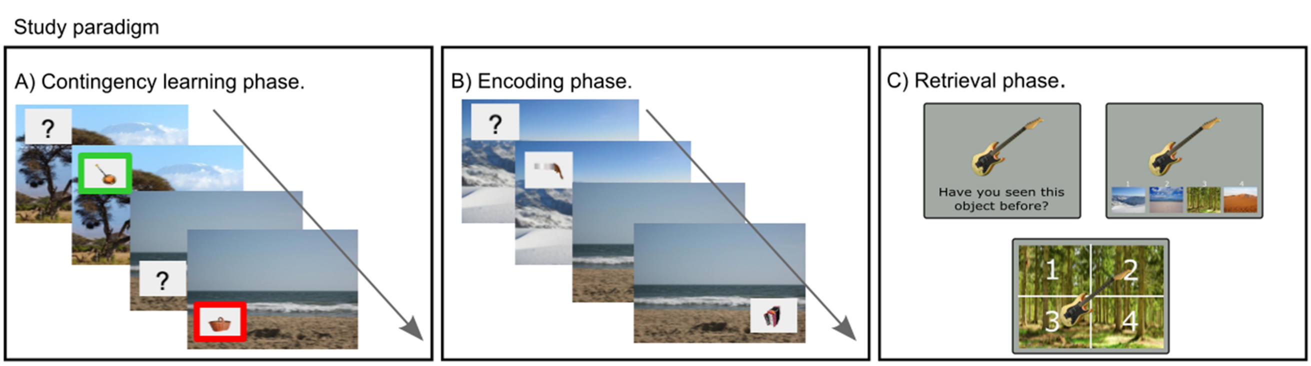
The Present Experiments

All four experiments included in this paper shared the same overall structure (Figure 1). They all started with a *contingency learning phase* in which participants learned associations between scene contexts and object categories. The strength of the associations between scene contexts and object categories (i.e., priors) was manipulated to create either biased prior distributions, in which a given object category was more likely to be shown than others (e.g., 90% vs. 10%, in a two-category scenario), or flat prior distributions, in which all object categories were equally likely to appear in a given context (e.g., 50% vs. 50%, in a two-category scenario). See the contingency levels of all experiments in Figure 2. We chose this approach, rather than relying on preexisting priors that are normatively shared among individuals (e.g., a saucepan is expected in a kitchen), in order to be able to tightly quantify and control prior strength.

After the initial learning phase, participants completed an *encoding phase* in which the established contingencies could either be met or violated (to varying degrees depending on the contingency level). To anticipate the results, performance during the contingency learning and encoding phases was very similar across experiments. During learning phases, participants started with random guesses and, in contexts with biased priors, gradually improved their performance up to the experimentally designed level; in flat prior contexts, they maintained random choices throughout the entire phase. During encoding phases, participants applied the knowledge acquired during the learning phase either by appropriately selecting the corresponding category, for contexts with biased priors, or by choosing randomly, for flat prior contexts (Figure S1 in the online supplemental materials). As this behavior was anticipated and not central to our research question, these results are not discussed further. However, this pattern of responses enabled us to infer the experienced PE when the stimulus presented (mis)matched the established contingencies. Therefore, we operationalized PE as *how unlikely it was to encounter a given object category in a given context* (1—prior strength).

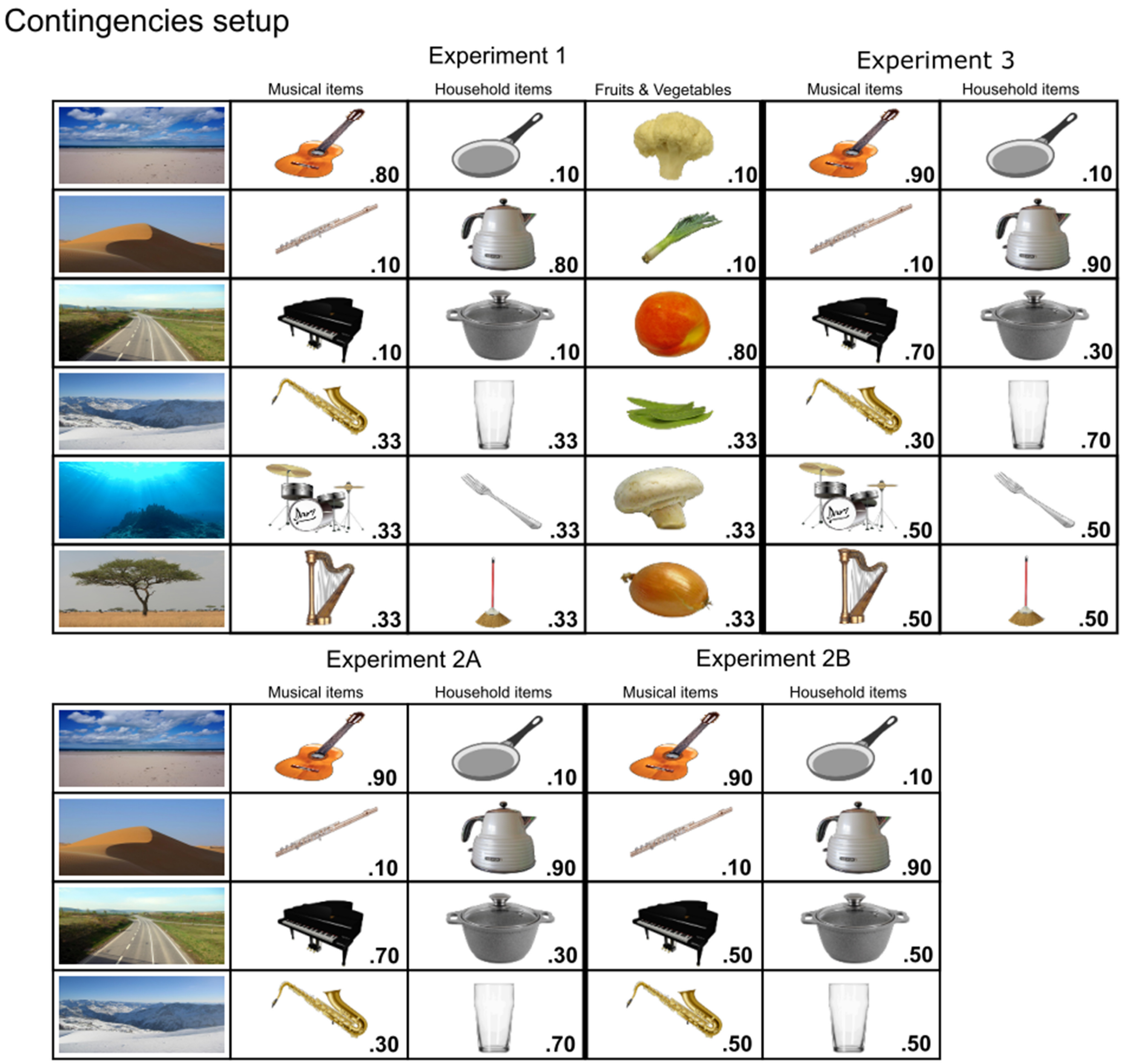
After completing the two preceding phases, participants were presented with two retrieval phases with varying delays between them.

Figure 1
Study Paradigm



Note. The same overall structure was shared across all the experiments in which participants completed three phases in the same order. (Panel A) Participants learned associations between scene contexts and object categories. In every trial, a real-world scene was shown with a question mark over a white patch. Participants' task was to predict which object category would be shown next, use the feedback provided, and improve their initial guesses over time. (Panel B) During the encoding phase, the same scenes and a new set of objects were used in either a prediction task (Experiments 1 and 3) or a categorization task (Experiments 2A and 2B). Explicit feedback on participants' performance was not shown. After each response in both the contingency learning and the encoding phases, a fixation cross was displayed for 500 ms before the next trial started; for simplification purposes, the fixation cross display is not shown here. (Panel C) In both retrieval phases (immediate and delayed; see main text for more details), participants performed an object recognition task followed by an object-scene association task (Experiments 1–3) and a location memory task (Experiments 2A, 2B, and 3). See the online article for the color version of this figure.

Figure 2
Schematic Illustration of the Contingency Setup Across Experiments



Note. Each object category was paired with a given scene category according to the numbers shown in each cell. These numbers indicate the likelihood of finding that object category in the specific context. PE was defined as a 1—*the likelihood of finding an object category in a given context*. For each experiment, the figure shows the arrangement for a given participant; across the entire sample, object categories were counterbalanced so that every object category was seen in every prior condition in every context. See the online article for the color version of this figure.

Each retrieval phase consisted of an unannounced episodic memory test that included a recognition test on the identity of the object (Experiments 1–3), an alternative forced choice (AFC) test on the object-scene association (Experiments 1–3), and another AFC on the on-screen location of the object (Experiments 2–3). The idiosyncratic details of each experiment are described in the corresponding sections. To anticipate the major differences across the experiments, Experiment 1 tested the episodic memory consequences of explicit predictions in a strongly biased versus flat prior setup (rendering three PE levels). Experiments 2A and 2B introduced an implicit prediction task during encoding in either a strongly biased versus weakly biased prior setup (four PE levels) or a strongly biased versus flat prior setup (three PE levels), in order to test for the effects of implicit predictions on episodic memory. Experiment 3 extended the results of Experiment 1 by encompassing three prior levels, namely, strongly biased, weakly biased, and flat prior (rendering five PE levels).

Based on the literature reviewed above, three testable hypotheses can be drawn regarding the relationship between PE and episodic memory: (a) knowledge-integration hypothesis: if existing knowledge facilitates the remembering of prior-matching information, then we would measure a negative monotonic relationship between PE strength and episodic memory; (b) PE-driven encoding enhancement: if PE is used as a trigger for encoding mechanisms, then we would measure a positive monotonic relationship between PE strength and episodic memory; and (c) if both knowledge-integration and PE gradually and positively improve episodic encoding and they do not compensate each other, we would observe a U-shape relationship between PE strength and episodic memory. Besides the specific directionality of the effect, if consolidation plays a critical role in revealing the effects of PE on memory, then we would find the aforementioned PE-driven effects magnified in the delayed memory phases. Finally, if prediction-related effects arise under any predictable situation, regardless of whether predictions are made explicit or not, we should observe similar PE-driven effects in Experiments 2A and 2B as in Experiments 1 and 3.

Experiment 1

The aim of Experiment 1 was to generate different levels of PE during encoding and to examine their effects on episodic memory performance. We manipulated the probability of an object category being associated with a scene category with two prior conditions, namely flat prior, where every object category is equally likely (any object category probability = .33) and strong prior, where one of the object categories are more likely than the rest (preferred object category probability = .80; remaining two object categories = .10). These two prior conditions then rendered three PE levels: low (PE = .20), intermediate (PE = .66), and high (PE = .90).

Method and Material

Participants

Thirty-two young adults (20 female; $M_{age} = 22.59$ years, $SD = 3.18$) were recruited through advertisements placed across the three campi of the Goethe University in Frankfurt. All participants had normal or corrected-to-normal vision and had no history of psychological or neurological disorders. All participants in all experiments reported here gave written informed consent prior to participation. The study was approved by the ethics committee of the Goethe University Frankfurt am Main. In exchange for participation, participants received either course credits (for Psychology majors) or an honorarium (for all other majors) of €8/hr.

Stimuli

A set of six scene images depicting real-world outdoor locations from the ECOS database (https://sites.google.com/view/ecosdatabase/) were used as context cues. The selected scene categories were beach, mountain, road, desert, savannah, and seabed. A total of 192 object images depicting real-world objects were gathered from an online search and were used as target objects. Object images included the same number of objects from three different non-overlapping categories, namely musical instruments, fruits/vegetables, and household objects. All images were subjected to Creative Commons licensing and are available at https://github.com/ortiztud/premup.

Design and Procedure

Experiment 1 was conducted over two sessions. In the first session, participants completed the learning, encoding, and first retrieval phases. The second session took place 1 week later and only included the second retrieval phase. Stimulus presentation and recording of the responses were done using MATLAB's Psychtoolbox (Brainard, 1997) in a 60 Hz monitor (resolution: 1,680 × 1,050, full HD).

Contingency Learning Phase. Participants were told that they would see objects presented within scene contexts, and that their task was to learn which type of object was more likely to belong to which scene context. They were told that it would be easier to learn from some contexts than others, but the contingencies of each context were not explicitly mentioned. Each trial started with a fixation cross at the center of the screen for 500 ms. Afterward, participants were presented with a scene image that included a rectangular white patch with a question mark. They were asked to make a prediction about the most likely object category to be encountered in such context and were given three response alternatives: musical instruments, fruits/vegetables, or household objects. Fixed category reminders were placed at the bottom of the screen and participants were asked to press one of three arrow keys (i.e., left arrow, down arrow, and right arrow) in a QWERTZ keyboard to select the correct category; the selected category was highlighted with a yellow frame. Two seconds after the scene onset, the question mark was replaced by an object, and the colored frame changed to green or to red to indicate correct or incorrect responses, respectively. The object together with feedback was shown for 1 s before the next trial began (see Figure 1 for a depiction of the paradigm). Random guesses were expected at the beginning of the phase and participants were told to use the feedback to learn the contingencies over trials.

The likelihood of encountering a given object category in a given scene context was manipulated to render different prior strengths. For one-half of the scene categories, one of the object categories was frequently presented (80% of trials), while the other two were equally (un)likely (10% of the trials each); for the other half of the scene categories, there was no preferred object category with all three being equally probable (33% of trials). Twelve different objects from the three categories were used in this phase. In order to achieve the desired probabilities, each object was repeated a different number of times depending on its category and on the context in which it was shown (see Figure 2 for the full arrangement of the object-to-scene associations and Table S1 in the online supplementary materials for a breakdown of the number of stimuli in each cell and phase). The association of each object category to each scene category was counterbalanced across participants so that across the entire sample, every object category was paired with every scene category.

Encoding Phase. From the participant's point of view, the encoding phase was almost identical to the contingency learning phase and only the following minor changes were introduced. To avoid the potential effects of explicit feedback on episodic encoding, participants no longer received feedback on their prediction. In addition, a new set of never-seen-before objects was used in this phase, and each of these objects was presented only once. To equate the number of objects in each critical cell for our analysis, we selected a fixed number of objects ($n = 20$) for each PE condition, and these were presented only once. Therefore, to achieve the desired contingencies for each scene category, we used filler objects from the same object categories and repeated them seven times (see Table S1 in the online supplementary materials); filler trials were not included in our analysis.

As in the contingency learning phase, participants' task was to predict, on every trial, which object category was the most likely to be encountered in the scene that was shown. The contingencies between object categories and scenes were the same contingencies as in the previous phase. Since priors were already built up during the first phase, at the start of the encoding phase these contingencies rendered three types of trials. Namely, trials in which participants had a strong expectation and the expectation was matched (i.e., low $\text{PE} = 1 - .80 = .20$), trials in which participants had no clear expectation, (i.e., intermediate $\text{PE} = 1 - .33 = .66$), and trials in which participants had a strong expectation and the expectation was mismatched (i.e., high $\text{PE} = 1 - .20 = .80$).

Retrieval Phase. Target objects from the encoding phase were split into two sets of equal size for the two memory sessions: immediate and 1-week delayed. Both sessions included an object recognition memory test and a scene association test. In the object recognition test, all the objects from the encoding phase together with another 192 (96 in each session) new objects were used. Trials started with a fixation cross for 500 ms, and objects were presented in isolation at the center of the screen. Participants were required to make old/new judgments and to report confidence in their responses using a 6-point Likert scale (from 1 = *low confidence* to 6 = *high confidence*). The scene association test was only performed for old objects and participants were asked to choose the scene in which the object had been presented in the encoding phase in a 6AFC format. The object was presented at the center of the screen and the six scenes used in the previous phases as scene contexts were offered as alternatives in one row at the bottom of the screen (see Figure 1). All responses in the retrieval phase were not time-constrained and the display stayed unaltered until participants made a response, and the next trial was presented.

Statistical Analysis

Across all the experiments, the effect of PE on the different memory measures was tested with generalized linear mixed-effects models using brms in R (Bürkner, 2017), based on Stan (Carpenter et al., 2017). Bayesian statistical modeling has the advantages of allowing to fit maximal varying effect structures minimizing convergence issues and has an intuitive nature, compared to frequentists methods (Nalborczyk et al., 2019). In addition, they have the benefit of incorporating prior knowledge about parameters into the model. The brms package uses the Hamiltonian Monte Carlo algorithm of the Markov Chains Monte Carlo family to draw random samples from the posterior. We followed Gelman et al.'s (2008) recommendation of using student's t distributions with $M = 0$, $df = 7$, and scale parameter of 2.5 as weekly informative priors for the population effects. As priors for the variance components, we used half Cauchy distributions (Gelman, 2006). We modeled participants as random intercepts, and the manipulated variables (PE strength and session) and their interactions as fixed effects. For each model, we run the model with 2,500 warmup iterations, 5,000 sampling iterations, and four chains. Inspection of the trace plots of the winning models for all the experiments showed convergence over the parameter estimation, with Rhat = 1 for all the models (see text in the online supplemental materials for more details on model specification).

Model comparison was performed to test for the significance of the random slopes of our manipulated variables. The final model was determined with a backward model selection approach. All the possible models were sorted in descending order according to the number of parameters included in each one. Each model, starting with the more complex one (i.e., full model) was tested against the following model that had one parameter less (i.e., reduced model) to select the model with the best out-of-sample predictive performance (McElreath, 2020). In order to compare the models, we used Bayesian leave-one-out-cross-validation (LOO-CV, Vehtari et al., 2017): smaller LOO information criterion (LOOIC) indicated better fit. We compared nested models by using Bayes factors (BFs) with marginal likelihoods from bridge sampling and stopped simplifying the models when there was strong evidence of a loss of predictive power. Marginal likelihoods from bridge sampling were also used for evaluating the evidence for the main effects, by comparing a model with the to-be-tested effect and a model without that effect. The full model always included all main effects and interactions as fixed effects and as random slopes; participants were always included as random intercepts. Once the winning model was determined, marginal likelihoods from bridge sampling were also used for evaluating the evidence for the main effects, by comparing a model with the to-be-tested effect and a model without that effect. For the sake of simplicity, only the final winning models are reported here, but see https://github.com/ortiztud/premup/blob/main/bayesian_analyses/Bayesian_analyses_all.html for a full overview of the process and for online materials to reproduce every analysis step. Finally, if the winning model rendered a significant effect of PE, linear and quadratic components of the effect were tested to statistically arbitrate among our three hypotheses. Evidence for and against linear and quadratic components was quantified by BFs estimated by the Savage–Dickey ratio (Wagenmakers et al., 2010). We also report the expected log pointwise predictive density difference (ELPDdiff) between the models for paired comparisons, and 95% credible intervals (CI) around our coefficients.

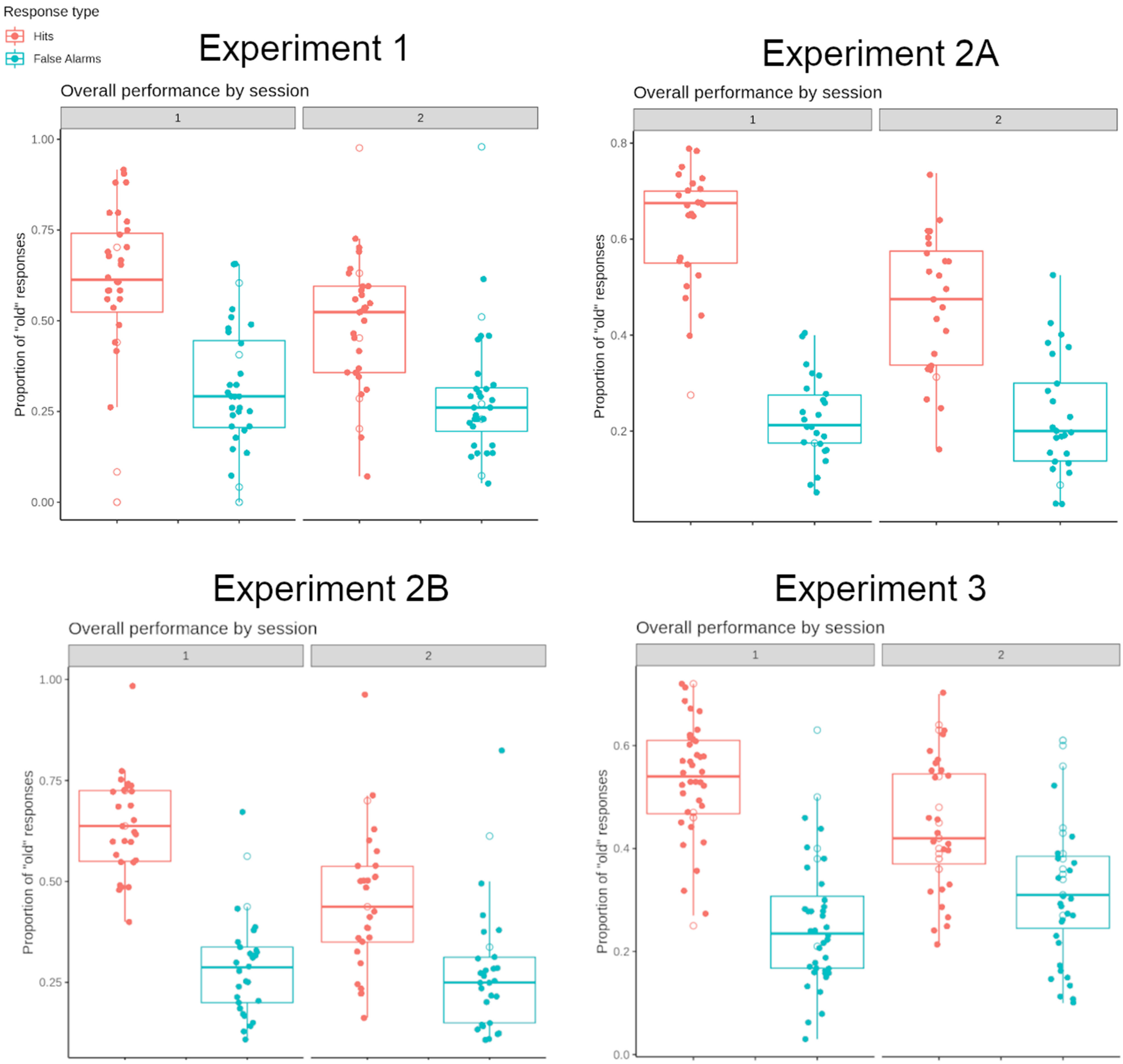
Results

Retrieval Phase

Performance on the object-scene association and the on-screen location memory tasks was not different from chance and thus, in the interest of brevity, only memory for object identity is included in the main text; see text in the online supplemental materials for a full description of the rest of the memory results.

Object Identity. An overall d' score was calculated for every participant from the average proportion of "old" responses to old (hits) and new (false alarms) trials to test for recognition memory for the objects. We obtained a fixed threshold situated above 95% of the observations of a d' prime distribution generated from 5,000 random permutations of the trial labels. Four participants whose overall d' score was below the obtained threshold and one participant with too few "old" responses (i.e., $<$20 responses per memory test) were excluded from further analysis (Figure 3, top-left). After exclusion, final overall d' was .93, $t(26) = 13.7$, $p < .05$, one-sided t test against zero.

Figure 3
Overall Pattern of Responses on the Recognition Memory Test for all Four Experiments Regardless of the PE Condition



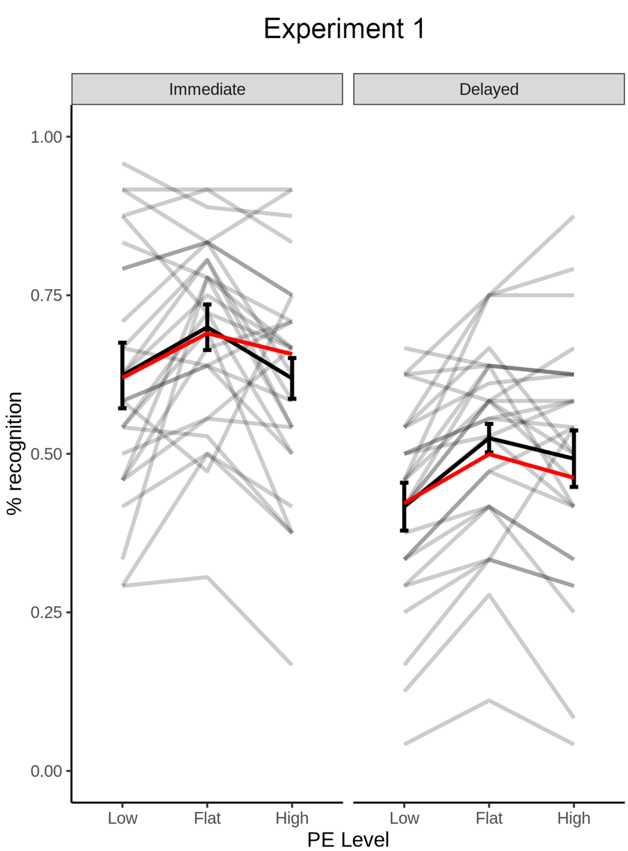
Note. Hits (right side of each plot) and False Alarms (left side of each plot) were used to compute a sensitivity score (d') for each participant. A random distribution of d' scores was generated by randomly swapping the trial labels and a performance threshold was defined as the value corresponding to the 95 percentile of the distribution. Filled dots represent participants with a d' prime score above the performance threshold and empty dots represent participants with a d' prime below the performance threshold, which were excluded from further analyses. See the online article for the color version of this figure.

Trial-level accuracy scores were further submitted to the modeling procedure and, in addition to PE, session and their interaction as fixed effects and participants as random intercept, the winning model included session as random slope, LOOIC = 5,645, SE = 47.4. The analysis of the model revealed a main effect of session ($ELPD_{diff} = 2.4$, $SE_{diff} = 2.4$, $BF_{10} > 100$), a main effect of PE ($ELPD_{diff} = 14.2$, $SE_{diff} = 5.6$, $BF_{10} > 100$), and no interaction between PE and session ($ELPD_{diff} = 0.4$, $SE_{diff} = 2.1$, $BF_{10} < 1$). We then tested the presence of a U-shape, by looking for evidence of a quadratic component. There was very strong evidence for a negative quadratic component, $\beta = -0.28$, 95% CI [−0.38, −0.18], $BF_{10} > 100$ and strong evidence for a linear component, $\beta = 0.11$, [−0.01, 0.22], $BF_{10} = 29$. In addition, we tested whether the linear and quadratic components interacted with the session and found very strong evidence for an interaction between session and a positive linear component, $\beta = 0.15$, [−0.02, 0.28], $BF_{10} = 45.73$. To break down the interaction, we looked at the immediate and delayed session separately. Results showed that while there was no evidence of a positive linear component on the immediate recognition session, $\beta = -0.01$, $[-0.18, 0.15]$, $BF_{10} < 1$, there was very strong evidence for a positive linear component in the delayed session, $\beta = -0.26$, $[0.07, 0.39]$, $BF_{10} > 100$. No evidence for the interaction between session and the negative quadratic component was found, $\beta = 0.02$, $[-0.04, 0.09]$, $BF_{10} < 1$. Note that the quadratic component was negative in sign thus signaling better memory performance for intermediate levels of PE (Figure 4).

To explore confidence ratings, these were dichotomized into low (<4) or high (>3) confidence responses. Response type (high vs. low confidence) was added as an interactive fixed effect into the winning model from above and contrasted against that same model. The new model (i.e., full) significantly improved the fit of the data, $ELPD_{diff} = 66$, $SE_{diff} = 11.4$, $BF_{10} > 100$, with response type interacting significantly with session, $ELPD_{diff} = 25.4$, $SE_{diff} = 6.9$, $BF_{10} > 100$, and with PE, $ELPD_{diff} = 3.1.4$, $SE_{diff} = 3.1$, $BF_{10} = 3.64$. The interaction with session was driven by fewer high-confidence responses in the delayed session, $\beta = 1.10$, $[0.79, 1.42]$. More interestingly, the interaction between response type and PE was driven by the quadratic component characterizing high but not low confidence responses, $\beta = 0.35$, $[0.79, 1.42]$. In other words, the quadratic relationship was only present for high-confidence responses. Finally, there was only anecdotal evidence for the three-way interaction, $ELPD_{diff} = 0.0$, $SE_{diff} = 0.1$, $BF_{10} = 1.03$.

Figure 4

Recognition Performance as a Function of PE Level in Experiment 1



Note. Solid black lines with error bars show the sample averaged responses and red lines (black lines without error bars in the printed version) show the fitted second-order polynomial model prediction. Light gray lines represent data from individual participants. Error bars show 95% CIs. The panels show data from the immediate (left) and delayed (right) memory tests. CI = confidence interval. See the online article for the color version of this figure.

Discussion

Experiment 1 was set to test the relation between PE at encoding and episodic memory performance. For this purpose, our participants learned artificial associations between scenes and objects that could vary in strength. In Bayesian terms, we built different priors for each of our encoding conditions. This manipulation of prior strength allowed us to sample three datapoints (.20, .66, and .80) along the PE continuum. Our results from the immediate memory test revealed that the relation between PE and item memory was characterized by a quadratic function. However, it is important to note that the observed component was negative in sign thus depicting an inverted U-shape pattern, that is, enhanced memory performance for intermediate levels of PE. In other words, trials in the flat prior condition were remembered better than those on either side of the strong prior condition. This outcome was not anticipated by any of the accounts previously considered. In addition, the pattern of results for the delayed session was better captured by a positive linear relationship between PE strength and memory, with higher PE levels associated with better memory than lower levels. This pattern is compatible with the PE-driven encoding enhancement accounts, according to which, PE can act as a signal to upregulate the encoding of new information (Barron et al., 2020; Henson & Gagnepain, 2010; Quent et al., 2021). We set to replicate the findings of Experiment 1, particularly the unexpected, inverted U-shape function, in Experiment 2.

Experiment 2

The pattern of results uncovered in Experiment 1 was not predicted by any of the a priori accounts. However, as the flat prior condition of Experiment 1 comprises a single data point in the PE continuum, the entire pattern could have been driven by some idiosyncratic features of that condition. Indeed, in the flat prior trials, learning of the preferred category is impossible by design and, as a consequence, participants might have approached the task with a strategy that was qualitatively different from the strong prior trials. Experiments 2A and 2B were intended as a replication and extension attempt that split efforts in two separate experiments. To test whether the obtained pattern was only driven by the flat prior condition, in Experiment 2A, the flat prior was removed and replaced with a weak prior instead. To rule out the possibility that the result of Experiment 1 was a spurious finding, Experiment 2B replicated Experiment 1 in a conceptually equivalent strong versus flat prior setup. In addition, Experiments 2A and 2B tested whether the effects obtained in Experiment 1 were exclusive to explicit predictions or whether they would also appear without the explicit requirement to make predictions. To test this idea, we postponed participants encoding response from the scene onset to the object onset. By doing so, we turned the *explicit prediction task* into a *categorization task* in which participants had to indicate the category of the presented object. Moreover, to further minimize the explicit prediction component of the encoding task, we also removed the asymmetry in the contingencies during the encoding phase (i.e., all object categories were equally likely). This manipulation had the added advantage of equating the number of trials in each cell of the critical comparisons thus avoiding the need for filler trials. If the prediction is inherent and automatic in the way our cognitive system works, once priors have been built, they should be applied regardless of whether the task requires an explicit prediction response or not. Finally, since memory for the object—scene association was very poor, we added an extra question probing a more salient feature (i.e., the on-screen location of the object at encoding).

Due to the COVID-19 pandemic, Experiments 2A and 2B had to be moved online and several implementation adjustments were necessary. Stimulus presentation and response collection were programmed in PsychoPy v2021.1.4 and hosted online in Pavlovia (https://pavlovia.org). At the beginning of each session, the experimenter met the participant in a virtual room using an online videoconferencing tool, in which the appropriateness of the testing setup was assessed with a brief set of questions about the participant's overall well-being, about the physical room in which the task would be performed and about the computer that would be used. All participants sat in a quiet room, used a laptop or a desktop computer, and were encouraged to minimize distractions as much as possible during the session. At the end of the session, the experimenter met the participant again and ask them about any unforeseen event or situation that might have come up during the completion of the task. Finally, to maximize engagement, self-administered breaks were included after every 40 trials during the contingency learning and encoding phases.

Experiments 2A and 2B were carried out across three sessions on three consecutive days. The first session consisted of the contingency learning phase and the second session included the encoding phase and the immediate memory test. Finally, the third session included the delayed memory test.

Method and Material

Participants

Twenty-six (7 female) and 29 (13 female) participants took place in Experiments 2A (M_{age}: 23.90, *SD*: 3.37) and 2B (M_{age}: 26.65, *SD*: 6.74), respectively. Participants were recruited through the Prolific platform (https://www.prolific.co/) and they all digitally signed informed consent approved by the local ethics committee.

Stimuli

Four out of the six scene categories and two out of the three object categories from Experiment 1 were used in Experiment 2. More specifically, we chose beach, desert, savannah, and mountain as scene contexts, and musical instruments and household objects, as target object categories, respectively.

Experiment 2A

Design and Procedure

The contingencies in Experiment 2A were set to .90–.10 and .70–.30, thus rendering four trial types. Namely, there were (a) trials in which participants had a strong expectation and the expectation was matched (i.e., low PE $= 1 - .90 = .10$); (b) trials in which participants had a weak expectation and that expectation was matched (i.e., medium-low PE $= 1 - .70 = .30$); (c) trials in which participants had a weak expectation and the expectation was mismatched (i.e., medium-high PE $= 1 - .30 = .70$); and (d) and trials in which participants had a strong expectation and the expectation was mismatched (i.e., high PE $= 1 - .10 = .90$). Therefore, this set increased the sampling rate from three in Experiment 1 to four data points along the PE continuum (i.e., .10, .30, .70, and .90). Finally, during the on-screen location memory question, objects were presented in one of the four quadrants of the screen (i.e., top-left, top-right, bottom-left, or bottom-right). The order of the locations was randomly selected and each of the four locations was equally likely across the entire task.

Deviations from the Registered Protocol

Prior to data collection, a registration was created, and it is available at https://osf.io/v6n2x. The first five participants complained about an initial 60–40 condition being too difficult (i.e., both object categories were perceived as equally likely even after multiple exposures). Therefore, to ensure Experiment 2A had a condition that was perceived as truly different from the flat prior condition on Experiment 1, instead of the planned .60–.40 and .80–.20 contingencies, we changed to .70–.30 and .90–.10 contingencies. The original registration was conceived as an age comparison study between children and young adults. However, due to the lack of any trend of an effect of PE on object memory at the initial stages of data collection, the children sample was not tested in the interest of resources and time. In addition to preregistered exclusion criteria, the same exclusion criteria for poor performers in the memory test that was used in Experiment 1 was also used in Experiment 2 (i.e., fixed threshold that leaves below 95% the observations of a d' distribution generated from 5,000 random permutations of the trial labels), in order to be consistent across all experiments reported here. For the same reason, instead of the preregistered analysis plan, generalized linear mixed models (see Statistical analysis section) with the same approach as in Experiment 1 were used. Such analytical approach is a more robust way of modeling the data because of the consideration of random intercepts and slopes.

Results

Retrieval Phase

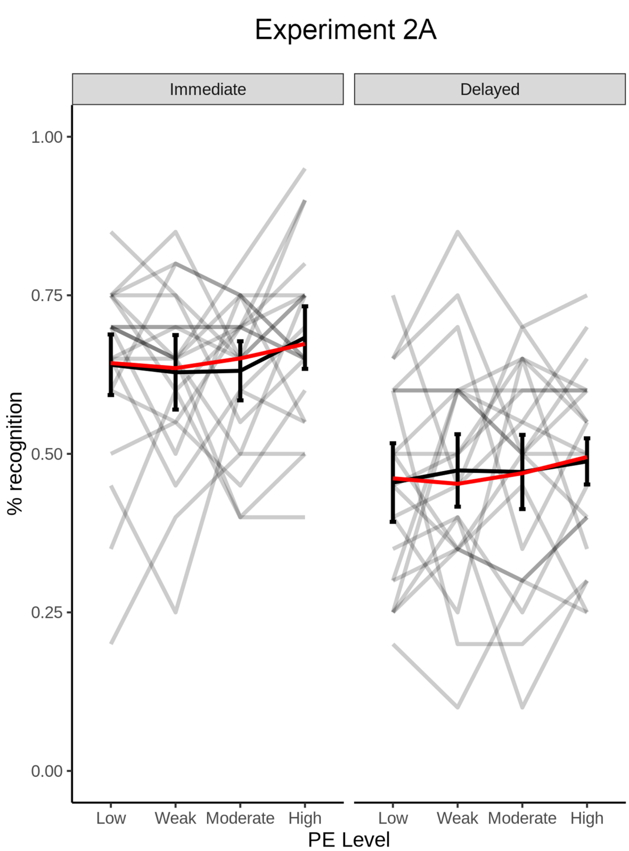
Object Identity. Overall d' was computed for every participant. Two participants were excluded by the fixed performance threshold. After exclusion, average d' was 1.13, $t(23) = 20.08$, $p < .001$ (Figure 3, top-left).

The model selection procedure revealed that, in addition to our main effects and participants as random intercepts, the winning model included session as a random slope, LOOIC = 5,286, $SE =$ 40.0. As in Experiment 1, we observed a main effect of session, $\text{ELPD}_{\text{diff}} = 2.9$, $SE_{\text{diff}} = 2.4$, $\text{BF}_{10} > 100$, with poorer memory in the delayed session. In contrast, PE had no effect in memory for the object identity, $\text{ELPD}_{\text{diff}} = 1.3$, $SE_{\text{diff}} = 1.2$, $\text{BF}_{10} < 1$, nor it interacted with session, $\text{ELPD}_{\text{diff}\,=} -1.5$, $SE_{\text{diff}} = 0.5$, $\text{BF}_{10} < 1$ (Figure 5). The evidence in favor of the model without PE, over the model including PE, was very strong, $\text{BF}_{01} = 49.40$, showing confidence toward the null effect of PE. As asymmetries in the contingencies were removed at encoding for this experiment, participants might have gradually updated their priors to match the new contingencies. In order to test for potential initial effects of PE, we repeated our analysis in the first half of the encoding trials when presumably contingencies were still intact. The results revealed strong evidence for a null effect of PE even when considering only the first half of the encoding task, ELPDdiff $= -1.3$, $SE_{\text{diff}} = 1.0$, $\text{BF}_{01} = 36.12$.

As for Experiment 1, in order to explore confidence ratings, we added dichotomized confidence responses (i.e., low and high confidence) to the winning model to find that it significantly improved the fit of the data. Again, in contrast to Experiment 1, we obtained no evidence for the interaction between session and confidence, $\text{ELPD}_{\text{diff}} = -0.2$, $SE_{\text{diff}} = 0.4$, $\text{BF}_{10} < 1$, for the interaction between confidence and PE, $\text{ELPD}_{\text{diff}} = 0.9$, $SE_{\text{diff}} = 1.4$, $\text{BF}_{10} < 1$, and only anecdotal evidence for the interaction among confidence, session, and PE, $\text{ELPD}_{\text{diff}} = -0.1$, $SE_{\text{diff}} = 0.1$, $\text{BF}_{10} = 1.01$. However, as in Experiment 1, we measured the main effect of response type, $\text{ELPD}_{\text{diff}} = 204.4$, $SE_{\text{diff}} = 17.6$, $\text{BF}_{10} > 100$, with more high than low confidence responses.

Figure 5

Recognition Performance as a Function of PE Level in Experiment 2A



Note. Solid black lines with error bars show the sample averaged responses and red lines (black lines without error bars in the printed version) show the fitted second-order polynomial model prediction. Light gray lines represent data from individual participants. Error bars show 95% CIs. The panels show data from the immediate (left) and delayed (right) memory tests. CI = confidence interval. See the online article for the color version of this figure.

Experiment 2B

Design and Procedure

Experiment 2B was identical to Experiment 2A other than for the contingencies chosen for the learning phase. While in Experiment 2A, .90–.10 and .70–.30 were used, in Experiment 2B contingencies were set at .90–.10 and .50–.50. Therefore, the strength of the priors in Experiment 2B rendered three PE levels at encoding, namely, .10, .50, and .90. Hence, Experiment 2B is a conceptual replica of the contingency conditions of Experiment 1 that also consisted of a strong and flat prior setup. Prior to data collection, a registration was created and is available at https://osf.io/jqfxh. As in Experiment 2A, we deviated from our analysis plan by adding an extra criterion for data exclusion (for details, please see the "Deviations from Registered Protocol" section).

Results

Retrieval Phase

Object Identity. Following the same procedure as in Experiment 2A, overall d' was computed for every participant, performance was filtered by a fixed threshold (obtained with a permutation test) and tested against zero with a one-side t-test. This procedure excluded two participants rendering a final average d' of 1.02, $t(26) = 20.79$, $p < .001$ (Figure 3, bottom-right).

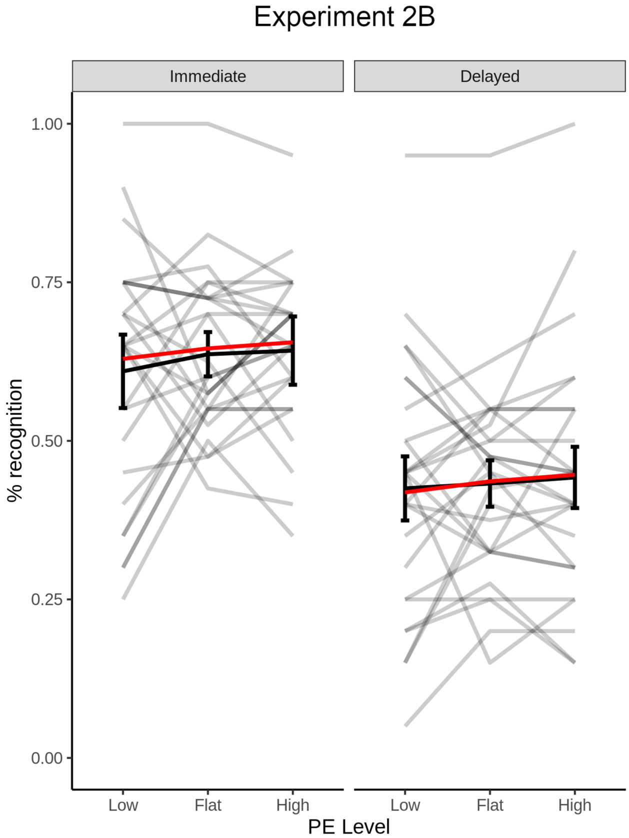
When recognition scores were submitted to the model selection procedure, the winning model included session as random slope in addition to the manipulated variables as fixed effects and participants as random intercepts, LOOIC = 2,148.6, $SE = 28.6$. Same as in Experiment 2A, we observed a main effect of session, $\text{ELPD}_{\text{diff}} = 1.6$, $SE_{\text{diff}} = 1.4$, $\text{BF}_{10} > 100$, with worse memory of the object identity on the delayed session. More importantly, and also mimicking the results of Experiment 2A, PE did not predict object recognition, $\text{ELPD}_{\text{diff}} = 2.1$, $SE_{\text{diff}} = 1.5$, $\text{BF}_{10} < 1$ (with strong evidence for a null effect of PE, B01 = 68.73), nor interacted with session, $\text{ELPD}_{\text{diff}} = 1.9$, $SE_{\text{diff}} = 0.9$, $\text{BF}_{10} < 1$ (Figure 6). Same as in Experiment 2A, we ran this analysis restricted to the first half of the trials, and still found strong evidence for the null effect of PE, B01 = 21.54.

Confidence responses were dichotomized and included into the winning model. The model with confidence responses significantly improved the fit of the data. The analysis of the model revealed a main effect of response type, $\text{ELPD}_{\text{diff}} = 204.3$, $SE_{\text{diff}} = 17.7$, $\text{BF}_{10} > 100$, with higher accuracy for high confidence than for low confidence responses. No evidence was found for the interaction between confidence and session, $\text{ELPD}_{\text{diff}} = -0.5$, $SE_{\text{diff}} = 0.4$, $\text{BF}_{10} < 1$, between confidence and PE, $\text{ELPD}_{\text{diff}} = 1.0$, $SE_{\text{diff}} = 1.4$, $\text{BF}_{10} < 1$, nor for the interaction among confidence, session, and PE, $\text{ELPD}_{\text{diff}} = -0.2$, $SE_{\text{diff}} = 0.1$, $\text{BF}_{10} < 1$.

Discussion

Neither Experiment 2A nor 2B replicated the inverted U-shape pattern obtained in Experiment 1. To consider the possible reasons for the lack of replication, there were two main aspects, one

Figure 6
Recognition Performance as a Function of PE Level in Experiment 2B



Note. Solid black lines with error bars show the sample averaged responses and red lines (black lines without error bars in the printed version) show the fitted second-order polynomial model prediction. Light gray lines represent data from individual participants. Error bars show 95% CIs. The panels show data from the immediate (left) and delayed (right) memory tests. CI = confidence interval. See the online article for the color version of this figure.

methodological and one conceptual, that differed between Experiments 1 and 2 but that were common between Experiments 2A and 2B. Namely, the online testing set-up and the task that participants performed at encoding. First, in terms of methodology, with participants doing the testing online from home, there was a range of factors that were less controlled compared to testing in the lab. The most prominent factors were the presence of distractors and participants' compliance. We tried to enhance testing quality by arranging the experimenters to meet the participants (virtually) in the beginning of the testing session to go over the task instructions and to verify the quality of the testing environment. Participants were allowed to use either a desktop or a laptop computer but were nonetheless asked to complete the task sitting in front of a desk. In addition, they were asked to avoid using their cell phones for the duration of the task and to use the bathroom before starting. Finally, the experimenters stayed available throughout the testing session and met with the participants again after the testing was completed. Through this, we ensured that participants could always ask for clarification if needed and that they felt more obliged to comply with task instruction, thereby reducing the difference between lab testing and online testing. Second, the more interesting theoretical aspect that differed between Experiments 1 and 2 is at the process level. The encoding setup of Experiments 2A and 2B minimized the need for explicit prediction by requiring a categorization (rather than a prediction) task. Indeed, recent findings on the effect of PE on pupil dilation and in relation to memory have shown that explicitly requiring a predictive response at encoding can be crucial to obtain prediction-related effects on memory performance (Brod, 2021; Brod et al., 2018). In other words, the lack of a PE-mediated effect on episodic memory in the absence of an explicit prediction task challenges the automaticity of the predictive processing mode, particularly concerning its long-term memory consequences.

Experiment 3

Experiment 1 revealed an unexpected, inverted U-shape pattern when relating PE to immediate episodic memory; Experiments 2A and 2B reduced the need for explicit predictions at encoding and did not reproduce the pattern observed in Experiment 1. However, other differences existed between Experiments 1 and 2, such as the online testing setup. The aim of Experiment 3 was to test whether the inverted U-shape pattern could be replicated under the conditions of (a) testing participants online; (b) restoring the need for prediction at encoding; and (c) increasing sampling along the PE continuum to cover five points (including the flat prior one). If under these conditions the inverted U-shape pattern could be observed again, it would suggest that the pattern is stable when explicit prediction is required, covering the full PE continuum.

Method and Material

Participants

We followed the same rationale for determining the sample size as in Experiment 2A and doubled the obtained number to end up with 40 participants in each group (see Procedure section). Eighty participants (34 female, M_{age}: 24.45; *SD*: 4.29) took part in Experiment 3 and they were randomly assigned to either an immediate memory test group or to a delayed memory test group. The session factor was manipulated between participants in this experiment to minimize interference between the increased number of objects necessary to achieve the high sampling rate of PE (see Procedure section). All participants were recruited through the Prolific platform (https://www.prolific.co/) and signed informed consent approved by the local ethics committee.

Stimuli

Stimuli from the same data sets as previous experiments were used. To achieve the high sampling along the PE continuum, we used six scene categories (as in Experiment 1) and two object categories (as in Experiment 2).

Procedure

The same overall structure of Experiment 1 was maintained while adding a third intermediate level akin to the weak prior condition in Experiment 2A. Therefore, the current experiment included three levels of prior strength (i.e., flat, weak, and strong prior) that rendered five different PE levels (i.e., low PE = 1 − .90 = .10, medium-low PE = 1 − .70 = .30, intermediate PE = 1 − .50 = .50, medium-high PE = 1 − .30 = .70, and high PE = 1 − .10 = .90). As a consequence, the contingency setup of Experiment 3 is effectively a combination of Experiments 2A and 2B with all five PE levels manipulated within participants. The increased number of prior levels, while keeping an acceptable cell size, could make it more difficult to learn and remember the contingencies and could also increase encoding and retrieval interference. Therefore, we added an extra reminder block of contingency learning right before the encoding phase. In addition, the retrieval session delay was manipulated between participants (immediate vs. 1 day after), thus reducing the number of trials needed by half. Prior to data collection, a registration was created and is available at https://osf.io/kwegs. The power calculation for this experiment mimicked that of previous ones and targeted the differences between extreme PE levels and the middle point separately for each group. A sensitivity analysis of the within-between interaction revealed that with a sample size of 80 participants, with an α level of .05 and a power of .80, we would also be able to detect effects sizes down to $f = .16$.

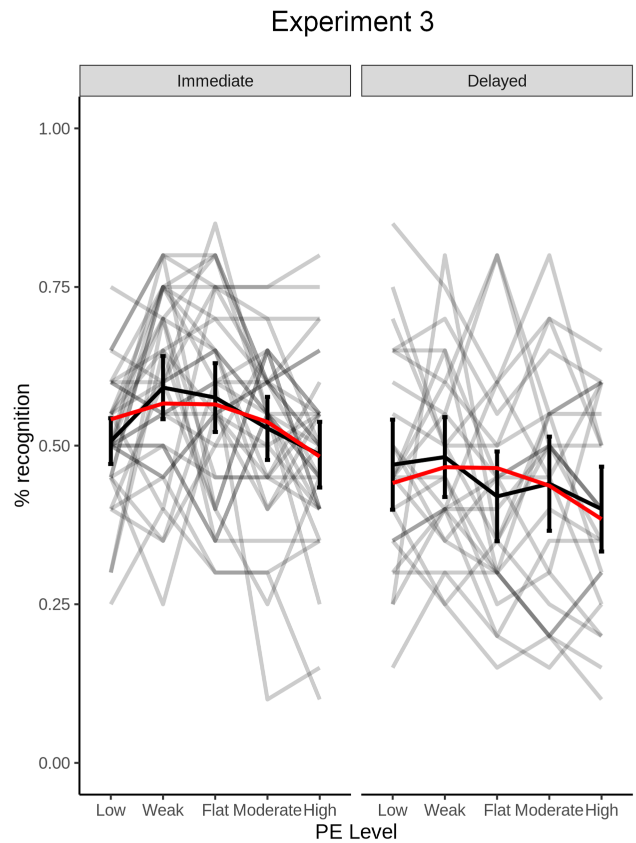
Results

Retrieval Phase

Object Identity. Overall d' was computed for every participant, five participants were filtered by the fixed threshold determined with a permutation test and the final overall d' was tested against zero with a one-side t-test. Overall d' was .90, $t(34) = 16.81$, $p < .001$ (Figure 3, bottom-left).

The model selection procedure led to a winning model that included PE, session, and their interaction as fixed effects and participants as random intercepts but no random slopes, LOOIC = 3,867.2, $SE = 24.0$. We observed a main effect of session indicating time-related memory difference, $\text{ELPD}_{\text{diff}} = 1.2$, $SE_{\text{diff}} = 1.6$, $\text{BF}_{10} = 19.54$, and no interaction with PE level, $\text{ELPD}_{\text{diff}} = -0.2$, $SE_{\text{diff}} = 2.4$, $\text{BF}_{10} < 1$. Interestingly, we observed a main effect of PE, $\text{ELPD}_{\text{diff}} = 9.3$, $SE_{\text{diff}} = 5.0$, $\text{BF}_{10} > 100$, in the same direction as that found in Experiment 1. This effect was characterized by both a quadratic, β = −0.18, 95% CI [−0.35, −0.01], $\text{BF}_{10} =$ 55.5, and a linear component, β = −0.40, [−0.58, −0.23], $\text{BF}_{10} > 100$, revealing higher memory performance for intermediate levels of PE when compared to the two extreme levels and an overall tendency of lower memory for strong PE levels (Figure 7). The linear effect was also qualified by strong evidence for an interaction with session, β = −0.27, [−0.60, 0.05], $\text{BF}_{10} = 18$. Analysis of the linear trend separately for the immediate versus delayed session revealed a stronger negative linear trend in the delayed recognition task, compared to the immediate one, immediate: β = −0.24, [−0.42, −0.07], $\text{BF}_{10} = 85.21$; delayed: β = −0.55, [−0.78, −0.33], $\text{BF}_{10} > 100$. Finally, there was no evidence for an interaction between the negative quadratic trend and session, β = 0.19, [−0.07, 0.47], $\text{BF}_{10} < 1$. Note that the direction of the linear component was opposite to one reported above: while in Experiment 1, we observed a positive linear relationship, in Experiment 3, this relationship was negative in sign with worst memory performance for high PE conditions. As this pattern was not expected and does not

Figure 7
Recognition Performance as a Function of PE Level in Experiment 3



Note. Solid black lines with error bars show the sample averaged responses and red lines (black lines without error bars in the printed version) show the fitted second-order polynomial model prediction. Light gray lines represent data from individual participants. Error bars show 95% CIs. The panels show data from the immediate (left) and delayed (right) memory tests. CI = confidence interval. See the online article for the color version of this figure.

replicate across conceptually similar experiments, we refrain from interpreting it further.

As in the previous experiments, confidence responses were dichotomized and added as a fixed effect in the winning model from above. The model with confidence responses improved the fit of the data, by showing a main effect of confidence, $\text{ELPD}_{\text{diff}} =$ 321.9, $SE_{\text{diff}} = 23.8$, $\text{BF}_{10} > 100$. There was no evidence for the interactions between confidence ratings and session, $\text{ELPD}_{\text{diff}} =$ 0.2, $SE_{\text{diff}} = 1.2$, $\text{BF}_{10} < 1$, between confidence ratings and PE, $\text{ELPD}_{\text{diff}} = 1.6$, $SE_{\text{diff}} = 2.7$, $\text{BF}_{10} < 1$, and among confidence ratings, session, and PE, $\text{ELPD}_{\text{diff}} = -0.2$, $SE_{\text{diff}} = 1.6$, $\text{BF}_{10} < 1$.

Discussion

The results of Experiment 3 replicated those of Experiment 1 by showing that intermediate levels of PE render enhanced memory performance when compared to the extreme levels at both ends. In addition, it extends the inverted U-shape pattern to five PE levels, thus ruling out the alternative explanation that the increased memory for intermediate levels of PE in Experiment 1 was unique for the flat prior condition where there is no information to be learned (see "General Discussion" section). Finally, the significant PE effects on memory after restoring the explicit prediction task challenges the idea that this type of effects on memory arise automatically and reflect a default mode of processing.

General Discussion

The conflicting postulations on whether memory is enhanced when events match our expectations or whether it is the unexpected events that are well remembered pose a theoretical challenge. The present series of experiments aimed to investigate how different quantifiable levels of PE influence immediate and delayed episodic memory performance. To create such different levels, we developed a novel paradigm that combines incremental learning with different tests of episodic memory, most importantly item memory. This setup allowed a gradual manipulation of prior strength that rendered different levels of PE varying within participants. When item memory for the events that generated the (mis)matching situations was tested with a recognition memory test immediately after encoding, we observed an inverted U-shape pattern, namely better memory for the intermediate PE levels compared to the more extreme points in the PE spectrum. This finding was found in two independent experiments, both comprising a prediction task that required participants to form an explicit prediction of the category of the upcoming stimulus. Two additional experiments provided the boundary conditions under which the inverted U-shape pattern disappears. Experiments 2A and 2B, with a setup almost identical to Experiment 3 but with no explicit prediction, rendered no effects of PE on memory.

One possible explanation for the higher memory performance for flat prior trials in Experiment 1 concerns the equal distribution of probabilities across object categories: When there was no prediction to be made (i.e., flat contexts), participants might voluntarily detract attention away from the prediction task and toward the object as an isolated element and, as a consequence, unintentionally boosted encoding of the object. Since the enhancement in memory performance in Experiment 1 was specific to the objects presented in contexts where the contingencies were equally likely for every object category, the inverted U-shape pattern might have been driven by participants adopting a qualitatively different attentional setting for that condition. However, the results of Experiment 3 seem to suggest that this was not the case. The inverted U-shape in Experiment 3 was not restricted to the flat prior condition, but rather it extended to other intermediate PE levels (i.e., weak prior conditions). Finally, when testing after overnight consolidation (i.e., the delayed recognition memory test), the relationship between PE and memory offered a different picture. When memory was tested 1 week after encoding (Experiment 1), memory for trials in the low PE extreme of the continuum decayed more strongly than intermediate or high levels, leading to a positive linear relationship with PE. This pattern of result was not replicated when memory was tested only 24 hr after encoding (Experiment 3), which could indicate the relevance of an extended period of time to allow memory decay to reveal PE-driven memory enhancements. However, since this pattern rests on a single measure (delayed memory for Experiment 1), plus it was not anticipated by any a priori account, and our a priori power calculations were not optimized for this particular comparison, the results must be interpreted carefully.

It is also important to note that, in order to maintain the asymmetric contingencies at encoding (Experiments 1 and 3) without exponentially increasing the number of to-be-remembered objects and thus the interference at test, we used filler objects which were not trial-unique. These filler objects could generate context-item contingencies, which could interfere with the context category established in the learning phase, thus rendering prediction (errors) at different levels. However, we would argue that this is very unlikely to have a meaningful impact on encoding. First, the maximum number of repetitions for a given filler object was very small in relation to the total number of trials (7 out of 312 for Experiment 1 and 16 out of 330 for Experiment 3). Second, the presentation of these filler objects was fully randomized within- and between-context, so that the presentation of the trial-unique objects was intermixed with the filler objects. In practice, this implies that the total number of repetitions of filler objects before encountering a given trial-unique object was even lower than the maximum stated above. Finally, during the learning phase, even in the easiest condition (Experiment 1; strong priors), it took participants at least 30 trials to learn the much easier contingency between a context and an object category (Figure S1 in the online supplemental materials). Thus, it is very unlikely that participants were able to learn these context-item contingencies fast enough to have interfered with the context-category ones. Nevertheless, future studies with a different strategy to control for the total number of unique objects could provide further insights into this.

U-Shaped Pattern: Upright or Inverted?

The inverted U-shaped pattern in Experiments 1 and 3 was not anticipated by the accounts reviewed in the introduction and challenges rather simple assumptions on the relationship between PE and episodic memory. Indeed, a knowledge-integration account would predict that events experienced under low PE situations would be better remembered than those that, by virtue of being poorly predicted, are more difficult to reconcile with existing knowledge (Craik & Tulving, 1975). Conversely, accounts that posit that PE boosts encoding of the mismatching information would predict a positive monotonic relationship between the level of PE and memory (Henson & Gagnepain, 2010). Finally, the dual, noncompensatory mechanisms account would predict better memory for both extremes of the continuum than for intermediate levels (Van Kesteren et al., 2012). The present set of results fits neither of these accounts. What follows in the next few paragraphs is thus an a posteriori consideration of the results in the light of other potential accounts.

The results of the current study could be interpreted under a recent framework which proposes that, during circumstances of uncertainty, increased attention, exploration, and information seeking for evidence can together contribute to enhanced memory performance (Gruber & Ranganath, 2019). In our task, the scene contexts that led to intermediate levels of PE were, by definition, more uncertain environments and the proposed information-seeking behavior might have been triggered to a larger degree than when facing more certain contexts. As uncertainty increased toward the center of the PE continuum, our participants might have upregulated evidence acquisition in an attempt to reduce uncertainty. In line with this framework is another (not mutually exclusive) account put forward recently (Sherman & Turk-Browne, 2020). They argued that having a strong prediction requires an active hippocampal representation that serves as the source of prediction. This predictive representation in the hippocampus biases the memory network toward a retrieval state, which prevents it from encoding the predicted event (i.e., the target object in our task). Since our memory capacities are limited, it would be adaptive to downweigh storing new information when an internal model that produces reliable predictions in the corresponding context already exists (see also Henson & Gagnepain, 2010 for a similar rationale). Accordingly, more extreme PE values in our task resulted from more certain contexts, in which a reliable predictive model was available (e.g., contingency of .90–.10 in Experiment 3); intermediate values of PE, however, came from more uncertain contexts in which the existing model was insufficient to produce accurate predictions (e.g., contingency of .50/.50 or .70/.30 in Experiment 3, Stanek et al., 2019). Therefore, the reliability of the predictive model in our task (increasing as we move away from the center of the PE continuum in both directions) might have downweighed the encoding of new information by biasing the system toward a retrieval state.

Both of the accounts reviewed above highlight the role that uncertainty in the environment might play in revealing prediction-related effects. Previous studies that report an upright U-shape relationship between PE and memory have not explicitly manipulated uncertainty at encoding (Quent et al., 2021) or have tested memory of the predictive cues rather than the (un)predicted events (Greve et al., 2019). Our encoding task, on the other hand, has a probabilistic nature (akin to statistical learning) in which the degree of uncertainty rests at the core of the manipulation. Under these circumstances, particularly when learning/updating is promoted (Experiments 1 and 3), encoding might have been boosted by uncertainty in the environment (Gruber & Ranganath, 2019), with this being particularly strong for intermediate levels of PE where a reliable predictive model is not available (Sherman & Turk-Browne, 2020).

Another alternative explanation that could reconcile the inverted U-shape pattern for item memory reported here with the upright U-shape pattern reported elsewhere (Greve et al., 2019; Quent et al., 2021), predominantly for associative memory, involves a trade-off between encoding of the object and encoding of the association. Namely, extreme levels of PE might boost associative encoding by diverting attention away from the objects and toward the association between the object and the context, thus rendering two opposite U-shape patterns for associative and item memory, respectively. In a similar way, a diversion of attention away from encoding of the object itself could lead to enhanced encoding of incidental details of the event (e.g., the object's location on the screen). Indeed, previous reports on the effects of prediction on episodic memory and hippocampal-based accounts of PE-driven memory enhancements (Greve et al., 2019; Kafkas & Montaldi, 2018b; Van Kesteren et al., 2012) postulate that high PE situations should lead to increased distinctiveness of the resulting memory traces. Despite the appeal of these accounts, they remain still purely speculative as none of the measures of distinctiveness (i.e., object-scene association and location memory) taken in the present experiment rendered above chance performance. Future studies which combine off-ceiling item memory with off-floor associative memory would be needed to shed light on this issue.

Lastly, the PE-related effects reported here contrast with a previous study on the effects of explicit (spatial) predictions on item memory (Ortiz-Tudela et al., 2018). This study reported no effects of prediction in the encoding of item identity. We argue, though, that both sets of results can be reconciled when considering the match between the level at which predictions are made (i.e., which stimulus feature is anticipated) and the level at which memory is tested (i.e., which stimulus features are probed). Ortiz-Tudela et al. (2018) used spatial predictions which were entirely orthogonal to the item identity; similarly, in all the experiments reported here, predictions were established at the category level (i.e., which object category is more likely to be shown in this context?) while memory was tested at the item level (i.e., have you seen this particular object before?). In order to be able to measure prediction (mis)matches effects on memory, predictions must be at the same level as the memory probes (an idea already outlined in Henson and Gagnepain, 2010; see also Ptok et al., 2019, for compatible evidence from the cognitive control literature). Interestingly, Ortiz-Tudela et al. (2018) also included a neutral condition which is conceptually similar to the flat prior condition in the present experiments. Yet, no encoding advantage for this condition was observed. This contrasting pattern suggests that the inverted-U shape arises specifically when updating the internal models (in order to improve predictions) is called for. This was not the case in Ortiz-Tudela et al. (2018) and neither in Experiments 2A and 2B in the present series (see more on this below).

Predictions and Predictive Effects on Memory: Automatic or Task Dependent?

The PE-driven effects reported here are restricted to encoding that entailed an explicit prediction. Although predictive processes are often assumed to be automatic in nature and put forward as a universal principle of the brain (Friston, 2008), our results showed that their effects on long-term memory might depend on whether predictions are explicitly made. Interestingly, a set of recent studies directly targeted this question (Brod et al., 2018, 2020, 2022). In these studies, participants were asked to make numerical predictions while pupil dilation measures were recorded. The authors showed that when the correct number was not in line with participants predictions, pupil dilation increased compared to when the correct number was expected. Critically, this PE effect was not found when participants had to retrospectively judge whether they would have predicted the outcome or not (cf. our encoding task in Experiments 2A and 2B). Moreover, when they correlated pupillary surprise responses with participants' update in knowledge, they found a positive relationship only for the condition in which a prediction task was required (Brod et al., 2018; see also Lohnas et al., 2018) for a recent account challenging the automaticity of related encoding and retrieval memory processes).

But what makes a prediction special and how is it different from a post-diction? We speculate on this distinction based on the similarities between our encoding tasks and those in Brod et al. (2018; see also Brod et al., 2022) for a recent proposal on the increased subjective value of highly surprising outcomes). More specifically, with almost identical setups, prediction, and post-diction mainly differ by the fact that prediction *prompts* the pre-activation of a representation that can be contrasted directly against the incoming stimulus while post-diction does not (since this contrasting can only happen after the stimulus onset; Brod, 2021). In other words, the pre-activation of a specific representation based on prior knowledge or memory is the necessary condition for PE to arise. Interestingly, this pre-activation can be present in tasks that are not explicitly predictive (Bein et al., 2020; Kafkas & Montaldi, 2018b). For instance, a recent study by Bein et al. (2020) explored the relation between PE and memory in the context of a statistical learning task, using pairs of objects that, unbeknownst to participants, occurred in sequences. Their results revealed that when a well-learned object pair was violated, encoding of the violating object was enhanced in comparison to objects with no prediction or violation. Importantly, this enhancement was only present with an encoding task that emphasized the inter-item relational processing of the object pairs (i.e., whether object at trial N was smaller or larger than object at trial $N-1$) but not with a task that prompted within-item feature evaluation (i.e., whether object at trial N is bigger or smaller than a shoebox). Their inter-item encoding task is likely to have prompted participants to activate the upcoming item in the presence of the previous one. Conversely, the within-item encoding task, which does not require any relational processing of the items, would not have entailed such pre-activations and hence showed no subsequent PE or PE effects on memory. Finally, it is important to note that there is at least one study that shows PE memory effects with within-item tasks (Kafkas & Montaldi, 2018b). However, in this study, the encoding task was preceded by an explicit prediction phase that smoothly faded out into the encoding task. This smooth transition might have nevertheless led participants' predictive processing mode to be carried over to the encoding task.

In sum, we argue that the effects of PE on episodic memory are not as straightforward as previously postulated (see also Ortiz-Tudela et al., 2018). While predictive processing may be default for some processes such as perception or action (Clark, 2013; de Lange et al., 2018; Miall & Wolpert, 1996; Wolpert et al., 1995), our results, together with the findings reviewed above, suggest that the extent to which it renders long-term memory effects relies on the pre-activation of particular (feature) representations. Critically, situations that require an explicit prediction from one event to the next (Experiments 1 and 3 and also Brod et al., 2018, 2020, 2022), which prompt event-to-event relational encoding (Bein et al., 2020) or which involve the continuous unfolding of events over time (Quent et al., 2021; Sinclair & Barense, 2019) are more likely to ensure this dynamic.

Finally, it is important to highlight the key role that uncertainty might play in moderating the relationship between PE and memory. In the light of the data presented here, we argue that the way PEs are processed depends on contextual demands and task goals. In contexts where prior updating is promoted (e.g., when participants need to ascertain the correct upcoming stimulus, as in Experiments 1 and 3), intermediate levels of PE signal the existence of a suboptimal predictive internal model (i.e., matched and mismatched predictions stem from guesses). In these situations, which resemble the initial stages of a statistical learning setup, acquiring new information can be beneficial. Conversely, when the existing model consistently provides high and low PEs (i.e., frequent matched responses in spite of the sporadic mismatched ones), the acquisition of new information is likely to be downweighed as updating such model might not be deemed beneficial (Gruber & Ranganath, 2019; Sherman & Turk-Browne, 2020). In contrast, contexts in which the environment is assumed to be stable and no updating of the prior is required (Ortiz-Tudela et al., 2018; Quent et al., 2021), intermediate PE levels do not convey informative value about the environment and neither benefit from the integration within pre-existing schemas nor from the distinctiveness that follows of a novelty signal (Van Kesteren et al., 2012).

Conclusion

The present set of experiments explored the relation between PE and episodic memory performance, showing that uncertainty in the environment (i.e., prior strength) is likely an important moderator in this relation, and that making explicit predictions can play a critical role in determining whether prediction-related effects on memory would occur. In our study, the different levels of PE were obtained by experimentally manipulating the strength of the prior from which predictions were drawn. Such a gradual manipulation brings in a new approach that is arguably closer to the conception of PE from the incremental learning literature. It has great potential to be utilized for systematic investigation of the interplay between PE and episodic memory. However, there are at least two other ways of experimentally generating different levels of PE, namely (a) by varying the precision of the evidence and (b) by varying the distance between the evidence and the prior (Greve et al., 2019). Each one of these approaches could have a different impact on the interplay between PE and episodic memory. Future studies with different approaches to quantify PE would further contribute to clarify this issue. In addition, a computational modeling approach would allow for the estimations of trial-level PE and uncertainty values, which both could be tested for their relations to memory. Such an approach would also allow the joint consideration of highly relevant metrics such as learning rates. Our results underscore that the relationship between PE and episodic memory is still far from fully characterized and that other moderating factors such as the strength of prior, the automaticity of the prediction, and potentially also the consolidation period, need to be taken into account.

Context

This project stems from the realization that the current debate on whether unpredicted situations render better episodic memories than well-predicted ones needed a quantitative description of prediction error and a paradigm that rested on minimal assumptions. To accomplish that, we worked together to develop a paradigm that incorporates explicit predictions, their (mis)matches, an episodic memory test, and a quantifiable way of measuring PE. The paradigm was initially conceived in a winter retreat in Riezlern and was developed jointly in the LISCO lab over the course of the following years. The unexpected pattern observed, and the novel approach taken have both theoretical and practical implications for the study of the relation between prediction and memory.

References

Barron, H. C., Auksztulewicz, R., & Friston, K. (2020). Prediction and memory: A predictive coding account. *Progress in Neurobiology*, *192*, Article 101821. https://doi.org/10.1016/j.pneurobio.2020.101821

Bein, O., Duncan, K., & Davachi, L. (2020). Mnemonic prediction errors bias hippocampal states. *Nature Communications*, *11*(1), Article 3451. https://doi.org/10.1038/s41467-020-17287-1

Bein, O., Livneh, N., Reggev, N., Gilead, M., Goshen-Gottstein, Y., & Maril, A. (2015). Delineating the effect of semantic congruency on episodic memory: The role of integration and relatedness. *PLoS One*, *10*(2), Article e0115624. https://doi.org/10.1371/journal.pone.0115624

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brod, G. (2021). Predicting as a learning strategy. *Psychonomic Bulletin & Review*, *28*(6), 1839–1847. https://doi.org/10.3758/s13423-021-01904-1

Brod, G., Breitwieser, J., Hasselhorn, M., & Bunge, S. A. (2020). Being proven wrong elicits learning in children—but only in those with higher executive function skills. *Developmental Science*, *23*(3), Article e12916. https://doi.org/10.1111/desc.12916

Brod, G., Greve, A., Jolles, D., Theobald, M., & Galeano-Keiner, E. M. (2022). Explicitly predicting outcomes enhances learning of expectancy-violating information. *Psychonomic Bulletin & Review 29*(6), 2192–2201. https://doi.org/10.3758/S13423-022-02124-X

Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction*, *55*, 22–31. https://doi.org/10.1016/j.learninstruc.2018.01.013

Brod, G., & Shing, Y. L. (2019). A boon and a bane: Comparing the effects of prior knowledge on memory across the Lifespan. *Developmental Psychology*, *55*(6), 1326–1337. https://doi.org/10.1037/dev0000712

Brod, G., Werkle-Bergner, M., & Shing, Y. L. (2013). The influence of prior knowledge on memory: A developmental cognitive neuroscience perspective. *Frontiers in Behavioral Neuroscience*, *7*, Article 139. https://doi.org/10.3389/fnbeh.2013.00139

Bürkner, P. C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Frank, D., Montaldi, D., Wittmann, B., & Talmi, D. (2018). Beneficial and detrimental effects of schema incongruence on memory for contextual events. *Learning and Memory*, *25*(8), 352–360. https://doi.org/10.1101/lm.047738.118

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4*(11), Article e1000211. https://doi.org/10.1371/journal.pcbi.1000211

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534. https://doi.org/10.1214/06-BA117A

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, *94*, 149–165. https://doi.org/10.1016/j.jml.2016.11.001

Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *Journal of Experimental Psychology: General*, *148*(2), 325–341. https://doi.org/10.1037/xge0000498

Gronau, N., & Shachar, M. (2015). Contextual consistency facilitates long-term memory of perceptual detail in barely seen images. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 1095–1111. https://doi.org/10.1037/xhp0000071

Gruber, M. J., & Ranganath, C. (2019). How curiosity enhances hippocampus-dependent memory: The prediction, appraisal, curiosity, and exploration (PACE) framework. *Trends in Cognitive Sciences*, *23*(12), 1014–1025. https://doi.org/10.1016/j.tics.2019.10.003

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, *20*(11), 1315–1326. https://doi.org/10.1002/hipo.20857

Kafkas, A., & Montaldi, D. (2018a). How do memory systems detect and respond to novelty? *Neuroscience Letters*, *680*, 60–68. https://doi.org/10.1016/j.neulet.2018.01.053

Kafkas, A., & Montaldi, D. (2018b). Expectation affects learning and modulates memory experience at retrieval. *Cognition*, *180*, 123–134. https://doi.org/10.1016/j.cognition.2018.07.010

Kaiser, D., Stein, T., & Peelen, M. V. (2015). Real-world spatial regularities affect visual working memory for objects. *Psychonomic Bulletin & Review*, *22*(6), 1784–1790. https://doi.org/10.3758/s13423-015-0833-4

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: A canonical cortical computation. *Neuron*, *100*(2), 424–435. https://doi.org/10.1016/j.neuron.2018.10.003

Lohnas, L. J., Duncan, K., Doyle, W. K., Thesen, T., Devinsky, O., & Davachi, L. (2018). Time-resolved neural reinstatement and pattern separation during memory decisions in human hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(31), E7418–E7427. https://doi.org/10.1073/pnas.1717088115

McElreath, R. (2020). Statistical rethinking: A Bayesian course with examples in R and Stan. *Statistical Rethinking*, *47*(4), 464–475. https://doi.org/10.1201/9780429029608

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, *9*(8), 1265–1279. https://doi.org/10.1016/S0893-6080(96)00035-4

Nalborczyk, L., Batailler, C., Loevenbruck, H., Vilain, A., & Bürkner, P. C. (2019). An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1225–1242. https://doi.org/10.1044/2018_JSLHR-S-18-0006

Ortiz-Tudela, J., Milliken, B., Botta, F., LaPointe, M. R. P., & Lupiañez, J. (2017). A cow on the prairie vs. a cow on the street: Long-term consequences of semantic conflict on episodic encoding. *Psychological Research*, *81*(6), 1264–1275. https://doi.org/10.1007/s00426-016-0805-y

Ortiz-Tudela, J., Milliken, B., Jiménez, L., & Lupiáñez, J. (2018). Attentional influences on memory formation: A tale of a not-so-simple story. *Memory and Cognition*, *46*(4), 544–557. https://doi.org/10.3758/s13421-017-0784-2

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321. https://doi.org/10.1016/j.cognition.2010.11.015

Press, C., Kok, P., & Yon, D. (2020). The perceptual prediction paradox. *Trends in Cognitive Sciences*, *24*(1), 13–24. https://doi.org/10.1016/j.tics.2019.11.003

Ptok, M. J., Thomson, S. J., Humphreys, K. R., & Watter, S. (2019). Congruency encoding effects on recognition memory: A stage-specific account of desirable difficulty. *Frontiers in Psychology*, *10*, Article 858. https://doi.org/10.3389/fpsyg.2019.00858

Quent, J. A., Greve, A., & Henson, R. (2021). Shape of U: The relationship between object-location memory and expectedness. *Preprint*, *53*(9), 1689–1699. https://doi.org/10.31234/osf.io/xq37j

Rao, R. P. N., & Ballard, D. H. (1999). Hierarchical predictive coding model hierarchical predictive coding of natural images. *Nature Neuroscience*, *2*, 79–87. https://doi.org/10.1038/4580

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*(1), 1–27. https://doi.org/10.1152/jn.1998.80.1.1

Sherman, B. E., & Turk-Browne, N. B. (2020). Statistical prediction of the future impairs episodic encoding of the present. *Proceedings of the*

National Academy of Sciences of the United States of America, *117*(37), 22760–22770. https://doi.org/10.1073/pnas.2013291117

Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in Cognitive Sciences*, *14*(10), 464–472. https://doi.org/10.1016/j.tics.2010.08.002

Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: Prediction error influences episodic memory reconsolidation. *Learning and Memory*, *25*(8), 369–381. https://doi.org/10.1101/lm.046912.117

Sinclair, A. H., & Barense, M. D. (2019). Prediction error and memory reactivation: How incomplete reminders drive reconsolidation. *Trends in Neurosciences*, *42*(10), 727–739. https://doi.org/10.1016/j.tins.2019.08.007

Stanek, J. K., Dickerson, K. C., Chiew, K. S., Clement, N. J., & Adcock, R. A. (2019). Expected reward value and reward uncertainty have temporally dissociable effects on memory formation. *Journal of Cognitive Neuroscience*, *31*(10), 1443–1454. https://doi.org/10.1162/jocn_a_01411

Staresina, B. P., Gray, J. C., & Davachi, L. (2009). Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. *Cerebral Cortex*, *19*(5), 1198–1207. https://doi.org/10.1093/cercor/bhn165

Van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. https://doi.org/10.1111/nyas.14321

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880–1882. https://doi.org/10.1126/science.7569931

Received December 16, 2021
Revision received November 30, 2022
Accepted December 21, 2022 ■