bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Statistical learning is not error-driven

İlayda Nazlı, Ambra Ferrari, Christoph Huber-Huber, Floris P. de Lange

Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

Author Note

This work was supported by a personal scholarship provided to İ.N. by the Ministry of National Education of the Republic of Turkey, and by a personal grant provided to F.P.d.L. by the European Union (ERC Consolidator Grant 101000942, "Surprise").

Author contributions: İ.N., A.F. and F.P.d.L. designed the research. İ.N. collected data. İ.N. and C.H.H. analyzed data. İ.N. wrote the first draft. İ.N., A.F., C.H.H and F.P.d.L. wrote the paper.

Correspondence concerning this article should be addressed to İlayda Nazlı, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands. E-mail: ilayda.nazli@donders.ru.nl

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Prediction errors have a prominent role in many forms of learning. For example, in reinforcement learning agents learn by updating the association between states and outcomes as a function of the prediction error elicited by the event. An empirical hallmark of such error-driven learning is Kamin blocking, whereby the association between a stimulus and outcome is only learnt when the outcome is not already fully predicted by another stimulus. It remains debated however to which extent error-driven computations underlie learning of automatically formed associations as in statistical learning. Here we asked whether the automatic and incidental learning of the statistical structure of the environment is error-driven, like reinforcement learning, or instead does not rely on prediction errors for learning associations. We addressed this issue in a series of Kamin blocking studies. In three consecutive experiments, we observed robust incidental statistical learning of temporal associations among pairs of images, but no evidence of blocking. Our results suggest that statistical learning is not error-driven but may rather follow the principles of basic Hebbian associative learning.

Keywords: statistical learning, Kamin blocking, prediction errors, incidental learning

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Statistical learning is not error-driven

Learning is an essential feat of animal cognition. It allows us to build and refine our internal models of the world, so that we predict and flexibly adapt to our dynamic environment. A key feature of learning is the ability to form associations between events that take place in a systematic relationship across space or time (Gershman, 2017). For example, in a typical classical conditioning experiment (Pavlov, 1927), a dog automatically salivates (i.e., unconditioned response) in response to food (i.e., outcome or unconditioned stimulus). During conditioning, the sound of a bell (i.e., cue or conditioned stimulus) is repeatedly paired with the food. Once conditioning is accomplished, the bell itself elicits salivation (i.e., conditioned response).

Cue competition is a crucial phenomenon in associative learning. It refers to the observation that learning which cues predict an outcome not only depends on the presence of the cues before the outcome. Rather, cues compete with each other to gain predictive power over the outcome, and this moderates the learning process (Boddez et. al., 2014; De Houwer et. al., 2005; Luque et. al., 2018; Schmidt & De Houwer, 2019). Cue competition is exemplified by the Kamin blocking effect (Kamin, 1969). In a typical blocking paradigm (see Table 1), observers first learn the association between cue A and outcome X, and later they are trained with the association between cues A + B and outcome X. As a result of blocking, observers do not learn the association between cue B and outcome X, because X is already completely predicted by cue A. In other words, the previously learned A-X association blocks learning the association between cue B and outcome X.

Blocking cannot be explained by simple contiguity-dependent Hebbian associative learning (Hebb, 1949). Thereby, it suggests that the simple temporal co-occurrence of different stimuli is not sufficient for learning to occur. Instead, the model developed by Rescorla and Wagner (1972) provides a viable explanation for blocking. According to the Rescorla-Wagner model, changes in associative strength are determined by the amount of discrepancy between the expected and the observed outcome, i.e. the prediction error. In the blocking procedure, the previously learned A→X

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

association prevents the formation of an associative link between the second cue B and the outcome X, because the cue A already minimizes the prediction error during the exposure to the AB→X compound stimulus. In typical blocking experiments, associations are learned either when the outcome is a reward (Aggarwal et. al., 2020; Aggarwal & Wickens, 2020; Sharpe et.al., 2017; Steinberg et. al., 2013) or when performance-related feedback is provided (Blanco et. al., 2014; Kruschke & Blair, 2000; Le Pelley et. al., 2005, 2007; Luque et. al., 2018, Mitchell et. al., 2006). This provides support that reinforcement learning (i.e., learning associations between events in a self-supervised manner, via trial and error) relies on an error-driven learning algorithm (Gershman & Daw, 2017).

Another powerful form of learning is known as statistical learning, often defined as the automatic and incidental extraction of regularities from the environment (Batterink et al., 2019; Frost et al., 2019; Saffran et. al., 1996; Sherman et al., 2020; Turk-Browne et al., 2010). In the context of statistical learning, we have limited information about how the learning process itself occurs. Several studies are suggestive of the fact that statistical learning may indeed similarly rely on prediction errors. In rats, dopaminergic activity in the ventral tegmental area is important for the formation of an association between two non-rewarding stimuli (Keiflin et al., 2019; Sharpe et al., 2017). In humans, statistical learning involves the ventral striatum (Klein-Flügge et al., 2019), which has been hypothesized to signal prediction errors (Klein-Flügge et al., 2019; O'Doherty et. al, 2004; McClure et. al., 2003).

However, other researchers, using variants of Kamin's blocking paradigm, did not find clear-cut evidence for error-driven statistical learning. Beeslay and Shanks (2012) did not observe any blocking in contextual cueing experiments, where participants incidentally learnt the spatial relationship among distracters and targets in a visual search task. This paradigm however deviates from classical blocking paradigms, which rely on a *temporal* prediction between a cue and a future outcome (Aggarwal et. al., 2020; Aggarwal & Wickens, 2020; Blanco et. al., 2014; De Houwer &

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Beckers, 2003; De Houwer et. al., 2005; Kruschke & Blair, 2000; Le Pelley et. al., 2005, 2007; Luque et. al., 2018, Mitchell et. al., 2006; Sharpe et.al., 2017; Steinberg et. al., 2013; Vandorpe et. al., 2005). Two subsequent experiments (Moris et al., 2014; Schmidt and De Houwer, 2019) observed blocking of temporal associations only for material that was intentionally learnt, but not for incidentally learnt stimulus associations. Such learning conditions substantially deviates from a typical statistical learning scenario, where observers automatically extract regularities without intention nor awareness (Batterink et al., 2019; Frost et al., 2019; Sherman et al., 2020; Turk-Browne et al., 2010). Overall, it is therefore still unclear whether statistical learning require prediction errors.

We addressed this unresolved question in three consecutive experiments, in order to understand whether statistical learning is error-driven. On every trial, we presented participants with two consecutively presented stimuli. Unbeknownst to participants, we manipulated the conditional probabilities between successively presented leading and trailing stimuli, such that each trailing image could be predicted on the basis of its preceding, leading image. After learning, we evaluated statistical learning by presenting participants with expected and unexpected image pairs. Successful earning was indexed by faster reaction times to expected relative to unexpected trailing stimuli (Hunt & Aslin, 2001; Richter & de Lange, 2019; Turk-Browne et. al., 2005).

Experiment 1

Method

Preregistration and data availability

All experiments were preregistered on the Open Science Framework. Deviations from preregistration are mentioned as such and justified in the corresponding sections below.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Participants

The experiment was performed online using the Gorilla platform (Anwyl-Irvine et al., 2020), and participants were recruited through the Prolific platform (https://www.prolific.co/). 148 participants performed the experiment. 47 of them were excluded before they finished the experiment based on a priori exclusion criteria (see section 'Exclusion and inclusion criteria'), and one participant was excluded from the final data analysis due to excessively slow responses (RTs above 3 times the mean absolute deviation [MAD] from the group mean). As a result, one hundred participants (37 females; mean age 24.49, range 18-40 years) were included in the data analysis. This final number of included participants was preregistered based on previous research (Richter & de Lange, 2019; Schmidt & De Houwer, 2019) considering that online data would be noisier and, therefore, a larger number of participants would be required to maintain the same statistical power. The pre-selected sample size yielded 84% power to detect a small sized (Cohen's d = 0.3) effect ($\alpha = 0.05$).

All participants had normal or corrected to normal vision, normal hearing and no history of neurological or psychiatric conditions. They provided written informed consent and received financial reimbursement (8 euro per hour) for their participation in the experiment. The study followed the guidelines for ethical treatment of research participants by CMO 2014/288 region Arnhem-Nijmegen, The Netherlands.

Exclusion and inclusion criteria

The online experiment was terminated if the percentage of correct responses during object categorization was below 80% (threshold was defined based on a preliminary pilot study) in any training or test phase (see 'Experimental design' and Figure 1a) or if the percentage of correct responses in attention check trials was below 80% in any of the experimental phases (see section 'Experimental design').

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Prior to the main data analysis, we discarded trials with no responses, wrong responses, or anticipated responses (i.e., response time < 200 ms). We also rejected trial outliers (response times exceeding 3 MAD from mean RT of each participant) and subject outliers (participants whose RTs exceeded 3 MAD from the group mean). For the accuracy analysis of the pair recognition task, we rejected trial outliers in terms of response speed (response times exceeding 3 MAD from mean RT of each participant).

Experimental design

In each experimental trial, participants were exposed to two images presented in the center of the screen in quick succession: a leading stimulus was followed by a trailing stimulus. For each participant, there were 4 leading stimuli (2 geometric shapes and 2 everyday objects) and 4 trailing stimuli (all objects). Everyday objects were randomly chosen from a pool of 64 stimuli derived from Brady et al. (2008) per participant, thereby eliminating potential effects induced by individual image features at the group level. In each stimulus set, 50% of objects were electronic (consisting of electronic components and/or requiring electricity to function) and 50% were non-electronic. The expectation manipulation consisted of a repeated pairing of images in which the leading image predicted the identity of the trailing image, thus over time making the trailing image expected given the leading image. Importantly, each trailing image was only (un)expected depending on which leading image it was preceded. Thus, each trailing image served both as an expected and unexpected image depending on the leading image. In addition, trial order was pseudo-randomized, with the pairs distributed equally over time. In sum, any difference between expected and unexpected occurrences cannot be explained in terms of familiarity, adaptation or trial history. Throughout the experiment, participants needed to categorize the trailing object as electronic or non-electronic as fast as possible. This task was aimed at assessing any implicit reaction time (RT) benefits due to incidental learning of the temporal statistical regularities: upon learning, leading images could be used to predict the correct categorization response before the trailing image

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

appeared. In addition, there were attention check trials where participants were simply asked to press a specific key based on a message on screen (e.g., "Press left-arrow key"). The aim of these trials (7% of all trials per participant) was to monitor participants' vigilance (see 'Exclusion and inclusion criteria'). A fixation bull's-eye was presented in the center of the screen throughout the experiment.

The blocking paradigm comprised two consecutive training phases, followed by one test phase (see Figure 1a). During the two training phases, leading stimuli were perfectly predictive of their respective trailing stimuli (i.e. P(trailing | leading = 1)). Participants were not informed about this deterministic association, nor were they instructed to learn this association at the beginning of the experiment. Therefore, the pair associations were could only be learned incidentally. In training phase 1, the leading stimulus was either a shape or an object, and it was always followed by the same trailing object. In training phase 2, a novel leading stimulus (blocked [B] leading stimulus) was presented along with the leading stimulus presented in training phase 1 (antedating [A] leading stimulus). If the antedating leading stimulus was an object, then the blocked leading stimulus was a shape or vice versa. In addition, novel leading (shape + object) and trailing (object) stimulus pairs were presented as a control. In the test phase, the leading stimulus of each condition (antedating [A] / blocked [B] / control [C]) was presented alone, followed by either the expected stimulus (based on the training phases), or an unexpected trailing stimulus. Expected and unexpected stimulus pairs were presented equally often to prevent any learning at this final test stage. In the test phase, control (C) trials were compared to blocked (B) trials to assess blocking while controlling for the amount of exposure. Also, the control trials in the test phase showed whether new associations had been learned during training phase 2.

Data was collected during one single session per participant. Firstly, participants familiarized themselves with all trailing objects. In each trial, an object image was presented for 3500 ms in the center of the screen, and participants had 1500 ms to categorize the image as electronic or non-

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

electronic (via a keyboard key press, keys counterbalanced across participants). Then, written feedback indicated the true category and the name of the object for 2000 ms (8 pairs × 2 trials / pairs = 16 trials in total). Afterwards, participants performed the experiment (i.e., training phase 1, training phase 2 and test phase). In each trial, the leading and trailing stimuli were presented for 500 ms successively with no inter-stimulus interval, followed by a 1500 ms inter-trial interval. Participants categorized the trailing object as electronic or non-electronic as fast as possible (via keyboard key press, keys counterbalanced across participants). Training phase 1 and training phase 2 started with a short practice period (practice training phase 1: 4 pairs × 4 trials / pairs = 16 trials in total; practice training phase 2: 8 pairs × 4 trials / pairs = 32 trials in total). After each practice, participants completed the training phases (training phase 1: 4 pairs × 26 trials / pairs = 104 trials in total; training phase 2: 8 object pairs × 26 trials / object pair = 208 trials in total). In addition, attention check trials (see above) were pseudo-randomly interspersed throughout the training phases without repetitions in successive trials. Afterwards, participants completed the test phase (12 pairs × 24 trials / pairs = 288 trials in total). Crucially, for each leading stimulus, both expected and unexpected trailing objects belonged to the same category (electronic or non-electronic). This ensured that differences in RTs during object categorization would not arise by mere response adjustments costs, but instead reflected perceptual surprise to unexpected trailing objects.

Finally, at the end of the experiment participants performed a pair recognition task to probe their explicit knowledge of the statistical regularities. Before starting the recognition task, participants were informed about the presence of statistical regularities among leading and trailing images in the previous experimental phases (i.e., training phases 1 and 2), and they were asked to indicate whether the trailing object was likely or unlikely given the leading stimulus according to what they saw during these previous phases. Participants familiarized themselves with the procedure via a brief practice (12 pairs × 2 trials / pairs = 24 trials in total) before completing the recognition task (12 pairs × 8 trials / pairs = 96 trials in total).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data analysis

We analyzed the RT data in the test phase in order to test for incidental learning of predictable stimulus transitions: upon learning, participants were hypothesized to react faster to expected relative to unexpected trailing stimuli (Richter et al., 2018, Richter & de Lange, 2019). Furthermore, we analyzed the accuracy data in the pair recognition test to assess participants' explicit knowledge about learnt statistical regularities. For both analyses, we used a Bayesian mixed effect model approach. The Bayesian framework allows a three-way distinction between evidence for an effect, evidence for no effect, and absence of evidence (Dienes, 2016; Keysers et al., 2020). This three-way distinction is important in the present study because it allowed us to draw conclusions from the initial experiment, consider alternative explanations, and run follow-up experiments to test these alternative explanations. An additional reason for this approach was the violation of the normality assumption for repeated measures ANOVAs of response times. Data were analyzed using the *brm* function of the BRMS package (Bürkner, 2017) in R. In the Supplementary information, we additionally provide classic frequentist analyses (i.e., ANCOVA of the reaction time data of the test phase and one-way ANOVA of the accuracy data of the pair recognition test) for comparability with previous studies and to verify that our conclusions do not depend on the analytical framework employed. Furthermore, in supplementary tables we provide post-hoc Bayesian mixed effect models that follow significant interaction effects.

Analysis of RT data in test phase. Firstly, we modeled the behavioral data of the antedating condition, where one leading stimulus was followed by one trailing stimulus. This served as a sanity check to verify the baseline assumption that participants were able to learn the temporal association between the leading and trailing stimuli. The model of the antedating (A) condition included reaction time as dependent variable and Expectation (unexpected / expected) as a fixed factor. To model the overall effect of time on task, we included Exposure as a continuous numeric predictor. Exposure was scaled between -1 and 1 to be numerically in the same range as the other factors,

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

which aids model convergence. For the interpretation of the results, the model coefficient for Exposure represents the increase in RT from the first to the last exposure. Finally, we included the interaction between Exposure and Expectation in the model, to probe extinction of the learnt associations. Namely, during the test phase participants are exposed equally often to expected and unexpected stimulus pairs, potentially resulting in extinction of the RT advantage for expected stimuli over time. The model included a full random effect structure (i.e., a random intercept and slopes for all within-participant effects).

Secondly, we determined whether there was blocking by jointly modeling the blocked (B) and control (C) conditions. The model of blocked and control conditions included reaction time as a dependent variable and Expectation (unexpected / expected), Condition (control / blocked) and Exposure as fixed independent variables. We included the interaction between Expectation and Condition to test for the blocking effect. The contrasts of the factors Expectation and Condition were coded as successive difference contrasts. Exposure was a continuous predictor scaled between -1 and 1, as in the antedating condition analysis. Again, we also modeled extinction (Expectation × Exposure interaction) and its interaction with Condition to probe for potential differences in extinction between conditions. The models were constructed using weakly informative priors centered at zero. The response time data was modelled using the exgaussian family and four chains with 25,000 iterations each (12,500 warm up) per chain and inspected for chain convergence. Coefficients were accepted as statistically significant if the associated 95% posterior confidence intervals were non-overlapping with zero. To measure the amount of evidence for and against an effect (evidence of absence), we calculated Bayes factors (BF) for each fixed effect parameter against the null hypothesis of this parameter being zero with the *hypothesis* function in BRMS.

Analysis of RT data split by stimulus type in test phase. We conducted a follow-up analysis that tested for the effect of the type of leading stimulus (shape / object). We reasoned that leading object stimuli may have attracted more attention than leading shape stimuli, given that they were

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

visually more salient than the surrounding grey shapes, and their category was task-relevant, as the task required object categorization on the trailing image. Given that associative learning depends on attention (Kruschke, 2001; Pacton & Perruchet, 2008), it was therefore conceivable that leading objects, rather than shapes, developed a stronger temporal association with trailing objects. We fit the model of antedating condition and the model of blocked and control conditions as described above, but with the inclusion of leading Stimulus Type (shape / object) as additional fixed factor. The model included a full random effect structure (i.e., a random intercept and slopes for all within-participant effects). If the posterior confidence intervals of the interaction effects between Expectation and leading Stimulus Type did not overlap with zero, we run separate models for shapes and objects respectively, in order to test for the blocking effect for each stimulus type. The models were constructed using weakly informative priors centered at zero. All other analysis settings were as specified above.

Analyses of accuracy data in pair recognition test. Firstly, we determined whether accuracy was above chance level within each condition (antedating / blocked / control). Hence, we created three separate binomial mixed-effects models with response error as dependent variable. Secondly, we determined whether there was a blocking effect in the explicit knowledge of implicitly learned associations. To do so, we created a binomial mixed-effects model with response error as binary dependent variable and Condition (blocked / control) as fixed factor. The models included a full random effect structure (i.e., a random intercept and slopes for the within-participant effects). The models were constructed using weakly informative priors centered at zero. All accuracy models were fit using Bernoulli family and four chains with 25,000 iterations each (12,500 warm up) per chain and inspected for chain convergence. With respect to significance and amount of evidence we used the same criteria as for the RT data.

Results

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Analysis of RT data in test phase. First, we compared the reaction times of expected and unexpected trials in the antedating condition to test whether repeated exposure to leading-trailing pairs led to learning their temporal association (see Table 2). We observed faster reaction times in expected (493 ms) than unexpected (508 ms) trials (b = 11.23, CI = [6.80, 15.59], $BF_{10} > 1000$, see Figure 1b), indicating successful learning of stimulus transition probabilities and the consequent behavioral benefit of expectation in terms of response speed. In addition, we tested whether this behavioral benefit remained stable during the test phase or dwindled, as would be expected by extinction. In line with the latter, we observed an interaction effect between Expectation and Exposure (b = -9.28, CI = [-15.26, -3.38], $BF_{10} = 9.01$), indicating that learning showed rapid extinction (expectation effect for run 1: 22 ms, run 2: 9 ms, run 3: 6 ms; see Figure 1c).

Next, we moved to our main question and tested for the presence of blocking (see Table 3 and Figure 1d). The reaction time difference between unexpected and expected trials was not different between control (11 ms) and blocked (12 ms) conditions (b = 1.85, CI = [-3.95, 7.51], $BF_{10} = 0.05$, see Figure 1b). With a $BF_{10} < 0.10$, this pattern of results presents strong evidence for the absence of blocking. There was also no difference in how the reaction time benefit for expected items behaved over time (b = -2.29, CI = [-11.17, 6.13], $BF_{10} = 0.01$; expectation effect in blocked condition for run 1: 13 ms, run 2: 4 ms, run 3: 12 ms; expectation effect in control condition for run 1: 18 ms, run 2: 10 ms, run 3: 7 ms; see Figure 1c).

Analyses of RT data split by stimulus type in test phase. In a follow-up analysis, we tested whether the type of leading stimulus (shape / object) affected statistical learning. In the antedating condition (see Table 4), the reaction time difference between unexpected and expected trials was larger for leading object (20 ms) compared to leading shape (9 ms) trials according to the posterior CI, with the BF being inconclusive (b = -10.00, CI = [-18.57, -1.48], $BF_{10} = 1.21$), which indicated that object-object associations were somewhat stronger than shape-object associations. While the difference in RT was larger for object-object associations than shape-object associations, separate

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

follow-up models showed that the reaction time difference was significant with strong BF evidence when the leading stimulus was an object (b = 15.19, CI = [7.98, 22.46], BF_{10} = 175.97, see Table S1 and Figure 2a-e), and it was still significant but with an inconclusive BF when it was a shape (b = 5.44, CI = [0.83, 10.05], BF_{10} = 0.61, Table S2 and see Figure 2b-f).

Across blocked and control conditions (see Table 5), the reaction time difference between unexpected and expected trials was also larger when the leading stimulus was an object (18 ms for B, 27 ms for C) compared to a shape (0 ms for B, 1 ms for C) (b = 18.40, CI = [11.52, 25.41], BF_{10} > 1000). Separate follow-up models showed that reaction times were faster in expected trials than in unexpected trials when the leading stimulus was an object (RT difference = 18 ms in blocked condition and 27 ms in control condition; b = 18.73, CI = [12.83, 24.5], BF_{10} > 1000, see Table S3 and Figure 2a-e). This was not the case when the leading stimulus was a shape (RT difference = 0 ms in blocked condition and 1 ms in control condition; b = 0.11, CI = [-3.27, 3.44], BF_{10} = 0.03, Table S4 and see Figure 2b-f). Overall, the data suggest that shape – object associations could be learnt, but to a lesser extent than object – object associations. In particular, shape – object associations could be learnt only if a leading shape in isolation was followed by a trailing object (i.e., in the antedating condition), but not when the leading shape was concurrently paired with a leading object (in a compound stimulus) and then followed by the trailing object (i.e., in the blocked and control conditions). This pattern of results fits our prediction that leading objects attract more attention than shapes, given that they were visually more salient, and their category was task-relevant. As associative learning depends on attention (Kruschke, 2001; Pacton & Perruchet, 2008), this may have hampered associative learning between leading shapes and trailing objects. In other words, we found cue competition among the leading shape and object in the forms of overshadowing (Boddez et. al., 2014; Pavlov, 1927; Schmidt & De Houwer, 2019), with the leading shape being overshadowed by the leading object. Finally, there was evidence for the absence of an interaction between Expectation, Condition and leading Stimulus Type (b = 4.09, CI = [-6.18,

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

15.80], $BF_{10} = 0.10$), indicating that the absence of blocking did not depend on leading Stimulus Type.

Analyses of accuracy data in pair recognition test. Participants were able to indicate whether the trailing object was likely or unlikely given the leading stimulus above chance level in the antedating (b = 0.32, CI = [0.23, 0.42], $BF_{10} > 1000$)**,** blocked (b = 0.16, CI = [0.09, 0.24], $BF_{10} = 185.67$) and control (b = 0.12, CI = [0.04, 0.19], $BF_{10} = 4.90$) conditions. Response errors did not differ between the blocked and control conditions (b = -0.05, CI = [-0.15, 0.05], $BF_{10} = 0.08$), indicating no blocking for the explicit knowledge of incidentally learned associations.

Experiment 2

Experiment 1 showed that the type of leading stimulus critically influenced statistical learning. Antedating and control leading shapes got less strongly associated with the trailing object than antedating and control leading objects. Moreover, blocked and control leading shapes could not compete with the concurrent leading objects for associative strength because they attracted less attention. This imbalance between shapes and objects may provide an alternative explanation for the lack of blocking that we observed. Therefore, in Experiment 2 we made one modification to our paradigm and only presented objects as leading and trailing stimuli to remove any potential difference in attention between different leading stimuli, which might finally result in a blocking effect.

Method

Participants

The experiment was performed online by using the Gorilla platform (Anwyl-Irvine et al., 2020), and participants were recruited through the Prolific platform (https://www.prolific.co/). 81 participants performed the experiment. 27 of them were excluded before they finished the experiment based on a priori exclusion criteria (see section ‘Exclusion and inclusion criteria’

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

above). Four extra participants were excluded from the final data analysis: two showed accuracy below 50% chance level in test phase; two showed overall excessively slow responses (RTs above 3 MAD from the group mean). As a result, fifty participants (16 females; mean age 23.90, range 18-34 years) were included in the data analysis, as preregistered. This final number of included participants was derived from the following a priori power calculation: we aimed for 90% power to detect the effect size of Cohen's d = 0.468 derived in the antedating leading object condition of Experiment 1 ($\alpha = 0.05$).

All participants had normal or corrected to normal vision, normal hearing and no history of neurological or psychiatric conditions. They provided written informed consent and received financial reimbursement (8 euros per hour) for their participation in the experiment. The study followed the guidelines for ethical treatment of research participants by CMO 2014/288 region Arnhem-Nijmegen, The Netherlands.

Experimental design

The design and procedure of Experiment 2 was identical in all respects to Experiment 1, apart from the type of leading stimuli and their location (see Figure 3a). Both leading and trailing stimuli were everyday objects. Leading and trailing objects were randomly presented on the left or right side of the central fixation point. Stimuli position (left / right) was counterbalanced with respect to Expectation (expected / unexpected) and Condition (antedating / blocked / control). In other words, leading and trailing objects appeared equally often on the left or right side of the central fixation point across trials. As a result, the expectation manipulation did not depend on spatial position. In addition, both hemi-fields were equally task-relevant, which fostered participants' attention to both sides.

Data analysis

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The data analysis of Experiment 2 was identical in all respects to Experiment 1, except for omitting the factor Stimulus Type because this experiment featured only object stimuli.

Results

Analyses of RT data in test phase. First, we compared the reaction times of expected and unexpected trials in the antedating condition (see Table 6). We observed that reaction times for expected (503 ms) and unexpected (510 ms) trials, although showing a qualitative pattern similar to Experiment 1, were not significantly different from each other (b = 4.95, CI = [-0.07, 9.96], BF_{10} = 0.33, see Figure 3b). Therefore, unlike Experiment 1, our data do not provide robust support for learning of the conditional probabilities in condition A (please note that we found a significant result via a classic frequentist approach; see 'Analyses of RT data in test phase using ANCOVA' in the Supplementary information). There was however some statistical support for extinction, as the reaction time difference between expected and unexpected trials tended to decrease as the exposure increased, however with an inconclusive BF (b = -8.17, CI = [-15.39, -0.91], BF_{10} = 0.77) (expectation effect for run 1: 17 ms, run 2: 6 ms, run 3: 0 ms; see Figure 3c).

Next, we moved to our main question and compared reaction time differences between expected and unexpected stimulus pairs between B and C (see Table 7 and Figure 3d). The reaction time difference between unexpected and expected trials was not statistically different between control (8 ms) and blocked (1 ms) conditions (b = 3.34, CI = [-3.11, 9.85], BF_{10} = 0.08, see Figure 3b). Moreover, extinction was not different between B and C (b = 0.37, CI = [-9.60, 10.22], BF_{10} = 0.10; expectation effect in blocked condition for run 1: 6 ms, run 2: -2, run 3: 0 ms; expectation effect in control condition for run 1: 11 ms, run 2: 4 ms, run 3: 5 ms; see Figure 3c).

Analysis of accuracy data in pair recognition test. Participants were not able to indicate above chance level whether the trailing object was likely or unlikely given the leading object in the antedating (b = 0, CI = [-0.15, 0.14], BF_{10} = 0.70), blocked (b = -0.05, CI = [-0.17, 0.07], BF_{10} = 0.09) or control (b = 0, CI = [-0.13, 0.14], BF_{10} = 0.07) conditions. Response errors did not differ

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between the blocked and control conditions ($b = 0.06$, CI = [-0.01, 0.21], $BF_{10} = 0.10$), indicating no blocking for the explicit knowledge of incidentally learned associations.

Experiment 3

Although Experiment 2 did not show any blocking effect, the data remained inconclusive: without a robust expectation effect in the antedating condition, which is a prerequisite for a valid blocking procedure (Rescorla & Wagner, 1972), we could not clearly establish whether participants were able to learn any temporal associations between the leading and trailing stimuli. In other words, it could be that learning was overall too weak in order for blocking to occur. Again, attention to the stimuli could likely have been a modulatory factor. It is well-known that attention to the stimuli is a prerequisite for statistical learning (Richter & de Lange, 2019; Turk-Browne et. al., 2005). In Experiment 2, the leading images were not task-relevant and they were easier to ignore (they appeared in the periphery) than Experiment 1 (where they appeared in the center of the screen, at fixation). Therefore, we created a slight modification in Experiment 3. We made the leading stimulus task-relevant with the intention to draw more attention to it under the hypothesis that this would enhance learning of the association and allow us to examine blocking with larger sensitivity.

Method

Participants

The experiment was performed online by using the Gorilla platform (Anwyl-Irvine et al., 2020), and participants were recruited through the Prolific platform (https://www.prolific.co/). 92 participants performed the experiment. 42 of them were excluded before they finished the experiment based on a priori exclusion criteria (see section ‘Exclusion and inclusion criteria’ above). As a result, fifty participants (18 females; mean age 25.80, range 18-40 years) were included in the data analysis. This final number of included participants was based on the same power analysis used for Experiment 2.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

All participants had normal or corrected to normal vision, normal hearing and no history of neurological or psychiatric conditions. They provided written informed consent and received financial reimbursement (8 euro per hour) for their participation in the experiment. The study followed the guidelines for ethical treatment of research participants by CMO 2014/288 region Arnhem-Nijmegen, The Netherlands.

Experimental design

The design and procedure of Experiment 3 was identical in all respects to Experiment 2, apart from the addition of an oddball detection task involving the leading stimuli in the training phases: participants were required to press a specific button as soon as they saw an animate leading stimulus (see Figure 4a). The aim of the animate detection task was to ensure that participants also paid attention to the leading stimuli, such that the association would be better learnt. For each participant, 4 animate leading stimuli (i.e., 2 for antedating leading stimulus and 2 for blocked leading stimulus) were randomly chosen from a pool of 8 stimuli derived from Brady et al. (2008). In addition, given that we observed fast extinction in Experiments 1 and 2, the number of trials in the test phase was decreased to 192 trials (i.e., 16 pair repetitions).

Data analysis

The data analysis of Experiment 3 was identical in all respects to Experiment 2, apart from the following: we adjusted the priors of the main effect of Expectation and Exposure and the prior of their interaction based on the posteriors of Experiment 2. Each prior was centered according to the median of the respective posterior estimate, and its standard deviation equated to the posterior estimate error times two to make the priors weakly informative. Note that specifying the priors in this way turns the results of Experiment 2 into the combined evidence from Experiments 1 *and* 2. Crucially, the pattern of results from Experiment 2 was exactly the same when priors were centered at zero.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Analyses of RT data in test phase. Firstly, we compared the reaction times of expected and unexpected trials in the antedating condition (see Table 8). We observed faster reaction times in expected (460 ms) than in unexpected (477 ms) trials (b = 10.81, CI = [5.04, 16.16], $BF_{10} > 214.11$, see Figure 4b), indicating successful learning of conditional probabilities and the consequent behavioral benefit of expectation in terms of response speed. In addition, we evaluated how this learning effect changed across exposure. Again, we observed an interaction effect between expectation and exposure (b = -9.01, CI = [-16.83, -1.18], $BF_{10} = 3.65$), indicating that learning showed rapid extinction (expectation effect for run 1: 26 ms, run 2: 11 ms; see Figure 4c).

Next, we modeled the blocked and control conditions to test whether we found blocking (see Table 9 and Figure 4d). There was a weak evidence for an interaction effect between expectation and condition (b = -9.48, CI = [-18.26, -0.45], $BF_{10} = 0.53$, see Figure 4b), with the BF being smaller than one, however, pointing rather at the absence of an interaction. We performed separate analyses for the blocked and control conditions to test for the presence of an expectation effect in each condition respectively. The reaction times in expected (481 ms) and unexpected (489) trials were not different from each other in the control condition (b = 4.36, CI = [-0.73, 9.51], $BF_{10} =$ 1.16, see Table S5). On the other hand, reaction times were clearly faster in expected (469 ms) than in unexpected (488 ms) trials of the blocked condition (b = 10.11, CI = [4.82, 15.16], $BF_{10} =$ 277.17, see Table S6). Interestingly, this is exactly the opposite pattern of what would be expected under blocking, and rather supports better learning of the associations among blocked stimuli than control stimuli. Extinction was not different between B and C conditions (b = -1.63, CI = [-14.19, 11.00], $BF_{10} = 0.14$; expectation effect in blocked condition for run 1: 13 ms, run 2: 18 ms; expectation effect in control condition for run 1: 6 ms, run 2: 3 ms; see Figure 4c).

Analysis of accuracy data in pair recognition test. Participants were able to indicate whether the trailing object was likely or unlikely given the leading object in the antedating (b = 0.39, CI =

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[0.26, 0.51], $BF_{10} > 1000$), blocked (b = 0.29, CI = [0.17, 0.42], $BF_{10} = 349.97$) and control (b = 0.39, CI = [0.24, 0.54], $BF_{10} > 1000$) conditions. Response errors did not differ between the blocked and control conditions (b = -0.1, CI = [-0.08, 0.29], $BF_{10} = 0.17$), indicating the absence of blocking effect for the explicit knowledge of incidentally learned associations.

Discussion

Statistical learning allows us to detect and learn structure in the environment, with direct benefits for directing our limited processing resources more efficiently to optimize behavior. This results, for example, in more efficient behavioral processing (Fiser & Aslin, 2001, 2002; Hunt & Aslin, 2001; Saffran et. al., 1996, 1999) and more efficient neural processing (Batterink & Paller, 2017; Henin et. al., 2021; Richter et. al., 2018; Richter & de Lange, 2019; Turk-Browne et. al., 2009) for predictable than unpredictable events. While the benefits of statistical learning are obvious, the mechanisms of statistical learning itself are less clear. In this study, we used a Kamin blocking paradigm (Kamin, 1969) to determine whether statistical learning is error-driven. We find no evidence of blocking during statistical learning, suggesting that statistical learning does not critically depend on prediction error.

Selective attention clearly mediated the effectiveness of our blocking procedure. Experiment 1 showed cue competition among the two concurrently presented leading stimuli, the shape and the object, in the forms of overshadowing (Boddez et. al., 2014; Pavlov, 1927; Schmidt & De Houwer, 2019). Specifically, the leading shape was overshadowed by the leading object. Originally, overshadowing was conceived as a direct consequence of error-driven learning (Rescorla & Wagner, 1972; Schmidt & De Houwer, 2019). However, it is becoming increasingly clear that perceptual saliency and feature relevance, which both strongly modulate attention, is at the core of overshadowing and of cue competition phenomena more generally (Endo & Takeda, 2004; Lau et. al., 2020; Luque et. al., 2018; Mackintosh, 1976; Pavlov, 1927; but see Murphy & Dunsmoor, 2017). Top-down selective attention is clearly implicated too, as dual task settings diminish the

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

blocking effect (De Houwer & Beckers, 2003; Vandorpe et. al., 2005). Experiment 2 further underscored the key modulatory role of attention in learning: reduced attention to our leading stimuli hampered statistical learning in the antedating condition. This echoes earlier findings showing that attention to signals containing regularities is critical for instantiating the behavioral (Turk Browne et. al., 2005; Zhao et. al., 2013) and neural (Richter & de Lange, 2019) consequences of statistical learning. Therefore, in Experiment 3 we controlled for any possible effects of attention by directing participants' attention to both leading and trailing images. Intriguingly, Experiment 3 showed strong learning of the associations for the blocked (B) stimulus condition; in fact, learning was even stronger for B stimuli compared to control (C) condition, a phenomenon which is sometimes referred to as 'augmentation' (Batson & Batsell, 2000; Beesley & Shanks, 2012; Vadillo & Matute, 2010). This pattern of results is opposite to the predictions of Kamin blocking and suggests that prediction error is not essential for statistical learning.

We speculate that selective attention may provide a parsimonious explanation for the observed augmented learning in the blocked condition. Several recent studies show that attentional allocation may proceed in order to maximize learning. For example, observers preferentially attend to stimuli that are not completely predictable or unpredictable (Gottlieb et al., 2013; Kidd et al., 2012; Poli et al., 2020). In other words, their attention is drawn to stimuli that offer maximum information gain (though see Mather, 2013 for a discussion on the effects of familiarity on attention). In our experiment, the association between the antedating leading object (A) and the trailing object was learnt during the first training phase. Therefore, participants' attention may have shifted to the novel blocked (B) leading image during the second training phase, enhancing learning of the association between the blocked leading image and the trailing image. On the other hand, in the control (C) condition, two novel leading objects were presented in the second training phase. In line with overshadowing, these two leading objects may have competed for associative strength

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with the trailing object and hence their individual predictive power was reduced (Rescorla & Wagner, 1972).

Considering the existing literature more broadly, there is mounting evidence for the absence of blocking in associative learning (Maes, 2016; but see Soto, 2018). Across three consecutive experiments, while progressively ruling out potential alternative explanations, we provide converging evidence specifically in statistical learning. We observed that participants learned the temporal association between antedating leading stimuli and trailing stimuli. However, such learning did not prevent participants from creating new subsequent associations in the blocked condition. This result supports the conclusion that incidental and automatic learning of simple temporal transitions between adjacent regularities does not depend on the use of prediction errors; instead, it may be a direct function of the amount of exposure. Moreover, it seems that the independence from prediction errors enables learning of additional contingencies (absence of a blocking effect) which might otherwise not be learned (blocked). At the computational level, such learning mechanism is compatible with chunking models of statistical learning (PARSER; Perruchet & Vinter, 1998; Perruchet, 2019), which may be implemented via fast Hebbian learning (Hebb, 1949) in functionally specific areas (Conway, 2020; Reber, 2013). This is in line with evidence of pair coding in the inferior temporal cortex of macaques during incidental statistical learning of adjacent visual object regularities (Meyer & Olson, 2011).

However, not all instances of statistical learning may follow this simple exposure-driven principle. In particular, learning more complex regularities may require error-driven mechanisms. Interestingly, observers are more aware of non-adjacent than adjacent regularities, even though the former ones are more complex (Romberg & Saffran, 2013). Furthermore, unimodal (e.g. visual-visual) regularities are learned quickly and automatically, whereas crossmodal (e.g. audio-visual) regularities cannot be learned through simple incidental exposure, but may instead require active intentional learning (Walk & Conway, 2016). These results have recently led to the suggestion that

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

different neuro-cognitive mechanisms of statistical learning may be at work depending on information complexity (Conway, 2020). Non-adjacent statistical structure, links between stimuli of different nature (i.e. crossmodal stimuli) or associations that depend on specific contexts cannot be formed via simple chunking mechanisms that rely on exposure-driven strengthening of synaptic connections within a specific area (Reber, 2013). Instead, transient midbrain activity may act as the teaching signal that functionally couples task-relevant brain areas, for example those responsible for processing stimuli across different sensory modalities (den Ouden et al., 2009; 2010). Finally, explicit and intentional associative learning in the form of causal inference likely is error-driven (De Houwer & Beckers, 2003; De Houwer et. al., 2005). Here, observers first learn that event A is the cause of outcome X. Then, in a subsequent phase where they observe B together with A, both of which are followed by X, they do not interpret B as a possible cause of X. Crucially, task instructions influence this process: when A is not described as the cause of outcome X, but simply as a likely preceding event, the blocking effect is significantly reduced (De Houwer & Beckers, 2003). Thus, the effortful evaluation of causal associations is required for the blocking effect to occur in such instances (Vandorpe et. al., 2005). To sum up, the present study shows a clear absence of Kamin blocking during incidental statistical learning of adjacent regularities. Thereby, it supports the conclusion that observers can attune themselves to simple environmental regularities by mere exposure, without the use of prediction errors. This suggests that incidental statistical learning may be implemented by a qualitatively different learning algorithm than intentional learning of rules and regularities.

Acknowledgements

The authors thank ...

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Aggarwal, M., Akamine, Y., Liu, A. W., & Wickens, J. R. (2020). The nucleus accumbens and inhibition in the ventral tegmental area play a causal role in the Kamin blocking effect. *European Journal of Neuroscience*, *52*(3), 3087-3109.

Aggarwal, M., & Wickens, J. (2020). The Kamin Blocking Effect in Sign and Goal Trackers. *bioRxiv*.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods*, *52*(1), 388-407.

Batson, J. D., & Batsell, W. R. (2000). Augmentation, not blocking, in an A+/AX+ flavor-conditioning procedure. *Psychonomic Bulletin & Review*, *7*(3), 466-471.

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31-45.

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the neural bases of implicit and statistical learning. *Topics in cognitive science*, *11*(3), 482-503.

Beesley, T., & Shanks, D. R. (2012). Investigating cue competition in contextual cuing of visual search. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(3), 709–725.

Blanco, F., Baeyens, F., & Beckers, T. (2014). Blocking in human causal learning is affected by outcome assumptions manipulated through causal structure. *Learning & behavior*, *42*(2), 185-199.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325-14329.

Boddez, Y., Haesen, K., Baeyens, F., & Beckers, T. (2014). Selectivity in associative learning: a

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cognitive stage framework for blocking and cue competition phenomena. *Frontiers in psychology*, *5*, 1305.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software*, *80*(1), 1-28.

Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience & Biobehavioral Reviews*, *112*, 279-299.

De Houwer, J., & Beckers, T. (2003). Secondary task difficulty modulates forward blocking in human contingency learning. *The Quarterly Journal of Experimental Psychology Section B,* 56(4b), 345-357.

De Houwer, J., Vandorpe, S., & Beckers, T. (2005). Evidence for the role of higher order reasoning processes in cue competition and other learning phenomena. *Learning & Behavior*, *33*(2), 239-249.

den Ouden, H. E. M., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience,* 30(9), 3210–3219.

den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cerebral Cortex,* 19(5), 1175–1185.

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78-89.

Endo, N., & Takeda, Y. (2004). Selective learning of spatial configuration and object identity in visual search. *Perception & Psychophysics*, *66*(2), 293-302.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological science*, *12*(6), 499-504.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fiser, J. Z., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 28, 458–467.

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128.

Gershman, S. J. (2017). Context-dependent learning and causal structure. *Psychonomic Bulletin & Review*, *24*(2), 557-565.

Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: an integrative framework. *Annual review of psychology*, *68*, 101-128.

Gottlieb, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences,* 17(11), 585-593.

Hebb, D. O. (1949). The Organization of Behaviour (Hoboken, NJ).

Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., ... & Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances*, *7*(8), eabc4530.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*(4), 658.

Kamin, L. J. (1969). Predictability, Surprise, Attention, and Conditioning. In B. A. Campbell, & R. M. Church (Eds.), Punishment Aversive Behavior (pp. 279-296). New York: Appleton-Century-Crofts.

Keiflin, R., Pribut, H. J., Shah, N. B., & Janak, P. H. (2019). Ventral tegmental dopamine neurons participate in reward identity predictions. *Current Biology*, *29*(1), 93-103.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Keysers, C., Gazzola, V., & Wagenmakers, E. J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, *23*(7), 788-799.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one*, *7*(5), e36399.

Klein-Flügge, M. C., Wittmann, M. K., Shpektor, A., Jensen, D. E. A., & Rushworth, M. F. S. (2019). Multiple associative structures created by reinforcement and incidental statistical learning mechanisms. *Nature Communications*, *10*(1), 1–15.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of mathematical psychology*, *45*(6), 812-863.

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, *7*(4), 636-645.

Lau, J. S. H., Casale, M. B., & Pashler, H. (2020). Mitigating cue competition effects in human category learning. *Quarterly Journal of Experimental Psychology*, *73*(7), 983-1003.

Le Pelley, M. E., Oakeshott, S. M., & McLaren, I. P. (2005). Blocking and unblocking in human causal learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(1), 56.

Le Pelley, M. E., Beesley, T., & Suret, M. B. (2007). Blocking of human causal learning involves learned changes in stimulus processing. *The Quarterly Journal of Experimental Psychology*, *60*(11), 1468-1476.

Luque, D., Vadillo, M. A., Gutiérrez-Cobo, M. J., & Le Pelley, M. E. (2018). The blocking effect in associative learning involves learned biases in rapid attentional capture. *Quarterly Journal of Experimental Psychology*, *71*(2), 522-544.

Mackintosh, N. J. (1976). Overshadowing and stimulus intensity. *Animal learning & behavior*, *4*(2), 186-192.

Maes, E., Boddez, Y., Alfei, J. M., Krypotos, A. M., D'Hooge, R., De Houwer, J., & Beckers, T.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(2016). The elusive nature of the blocking effect: 15 failures to replicate. *Journal of Experimental Psychology: General*, *145*(9), e49.

Mather, E. (2013). Novelty, attention, and challenges for developmental psychology. *Frontiers in psychology,* 4, 491.

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*(2), 339-346.

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences*, *108*(48), 19401-19406.

Mitchell, C. J., Lovibond, P. F., Minard, E., & Lavis, Y. (2006). Forward blocking in human learning sometimes reflects the failure to encode a cue–outcome relationship. *Quarterly Journal of Experimental Psychology*, *59*(5), 830-844.

Morís, J., Cobos, P. L., Luque, D., & López, F. J. (2014). Associative repetition priming as a measure of human contingency learning: Evidence of forward and backward blocking. *Journal of Experimental Psychology: General*, *143*(1), 77–93.

Murphy, G. L., & Dunsmoor, J. E. (2017). Do salient features overshadow learning of other features in category learning?. *Journal of Experimental Psychology: Animal Learning and Cognition*, *43*(3), 219.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*(5669), 452-454.

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 80.

Pavlov, I.P., 1927. Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. Oxford Univ. Press, Oxford, England.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Perruchet, P. (2019). What Mechanisms Underlie Implicit Statistical Learning? Transitional Probabilities Versus Chunks in Language Learning. *Topics in Cognitive Science,* 11(3), 520–535.

Perruchet, P., & Vinter, A. (1998). PARSER: A Model for Word Segmentation. *Journal of Memory and Language,* 39(2), 246–263.

Poli, F., Serino, G., Mars, R. B., & Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science Advances,* 6(39), eabb5053.

Reber, P. J. (2013). The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia,* 51(10), 2026–2042.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A. H., & Prokasy, W. F. (Eds.), *Classical conditioning II: Current research and theory*, 64–99. New York: Appleton-Century-Crofts.

Richter, D., & de Lange, F. P. (2019). Statistical learning attenuates visual activity only for attended stimuli. *Elife*, *8*, e47869.

Richter, D., Ekman, M., & de Lange, F. P. (2018). Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *Journal of Neuroscience*, *38*(34), 7452-7461.

Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science,* 37(7), 1290–1320.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27-52.

Schmidt, J. R., & De Houwer, J. (2019). Cue competition and incidental learning: No blocking or

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

overshadowing in the colour-word contingency learning procedure without instructions to learn. *Collabra: Psychology*, *5*(1), 1–16.

Sharpe, M. J., Chang, C. Y., Liu, M. A., Batchelor, H. M., Mueller, L. E., Jones, J. L., … Schoenbaum, G. (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nature Neuroscience*, *20*(5), 735–742.

Sherman, B. E., Graves, K. N., & Turk-Browne, N. B. (2020). The prevalence and importance of statistical learning in human cognition and behavior. *Current opinion in behavioral sciences*, *32*, 15-20.

Soto, F. A. (2018). Contemporary associative learning theory predicts failures to obtain blocking: Comment on Maes et al. (2016).

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature neuroscience*, *16*(7), 966-973.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of cognitive neuroscience*, *21*(10), 1934-1945.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, *30*(33), 11177-11187.

Vadillo, M. A., & Matute, H. (2010). Augmentation in contingency learning under time pressure. *British Journal of Psychology*, *101*(3), 579-589.

Vandorpe, S., De Houwer, J., & Beckers, T. (2005). Further evidence for the role of inferential reasoning in forward blocking. *Memory & Cognition,* 33(6), 1047-1056.

Walk, A. M., & Conway, C. M. (2016). Cross-domain statistical–sequential dependencies are

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

difficult to learn. *Frontiers in psychology*, *7*, 250.

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological science*, *24*(5), 667-677.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1

General experimental design (Kamin blocking paradigm).

Training phase 1	*Training phase 2*	*Test phase*
A → X	AB → X	A → X
	CD → Y	B → X
		D → Y

Table 2

Fixed effects of the model of antedating condition on reaction times in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	*BF_{10}*
Intercept	502.44	8.42	485.21 – 518.66	
Expectation	11.23	2.25	6.80 – 15.59	>1000
Exposure	-15.14	3.51	-22.08 – -8.19	>1000
Expectation × Exposure	-9.28	3.01	-15.26 – -3.38	9.01

Table 3

Fixed effects the model of blocked and control conditions on reaction times in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	*BF_{10}*
Intercept	494.45	8.34	478.02 – 510.93	
Expectation	10.88	1.6	7.76 – 13.98	>1000
Condition	4.30	1.95	0.38 – 8.10	0.27
Exposure	-19.10	3.08	-25.19 – -13.08	>1000
Expectation × Condition	1.85	2.91	-3.95 – 7.51	0.05
Expectation × Exposure	-7.19	2.24	-11.61 – -2.87	8.61

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Condition × Exposure	-3.00	2.26	-7.49 – 1.40	0.11
Expectation × Condition × Exposure	-2.29	4.48	-11.17 – 6.13	0.10

Table 4

Fixed effects the model of antedating condition on reaction times split by stimulus type in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	*BF_{10}*
Intercept	502.27	8.42	485.68 – 518.72	
Expectation	10.37	2.18	6.15 – 14.62	114.90
Leading stimulus type	56.95	5.6	46.13 – 67.99	>1000
Exposure	-15.35	3.52	-22.36 – -8.31	>1000
Expectation × Leading stimulus type	-10.00	4.37	-18.57 – -1.48	1.21
Expectation × Exposure	-7.26	2.61	-12.36 – -2.18	2.06
Leading stimulus type × Exposure	10.55	3.81	3.02 – 18.15	3.19
Expectation × Leading stimulus type × Exposure	1.12	5.38	-9.32 – 11.81	0.07

Table 5

Fixed effects the model of blocked and control conditions on reaction times split by stimulus type in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	*BF_{10}*
Intercept	494.13	8.29	477.45 – 510.54	

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Expectation	9.30	1.65	6.05 – 12.49	>1000
Condition	4.58	1.96	0.71 – 8.47	0.53
Leading stimulus type	-79.57	6.11	-91.88 – -67.48	>1000
Exposure	-19.54	3.03	-25.49 – -13.66	>1000
Expectation × Condition	1.97	2.60	-3.13 – 7.09	0.50
Expectation × Leading stimulus type	18.40	3.62	11.52 – 25.41	>1000
Condition × Leading stimulus type	-6.78	3.88	-14.36 – 0.89	0.24
Expectation × Exposure	-5.79	1.90	-9.53 – -2.06	3.63
Condition × Exposure	-3.45	1.98	-7.35 – 0.37	0.14
Leading stimulus type × Exposure	-8.64	2.93	-14.29 – -2.90	2.28
Expectation × Condition × Leading stimulus type	4.90	5.55	-6.18 – 15.80	0.10
Expectation × Condition × Exposure	-3.96	3.77	-11.36 – 3.41	0.08
Expectation × Leading stimulus type × Exposure	-17.54	3.70	-24.82 – -10.32	>1000
Condition × Leading stimulus type × Exposure	-0.21	3.74	-7.41 – 7.13	0.05
Expectation × Condition x Leading stimulus type × Exposure	14.37	7.56	-0.59 – 28.99	0.46

Table 6

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fixed effects the model of antedating condition on reaction times in Experiment 2. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	BF_{10}
Intercept	512.83	18.97	475.52 – 549.47	
Expectation	4.95	2.51	-0.07 – 9.96	0.33
Exposure	-18.40	4.29	-26.74 – -10.02	259.48
Expectation × Exposure	-8.17	3.70	-15.39 – -0.91	0.77

Table 7

Fixed effects the model of blocked and control conditions on reaction times in Experiment 2. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	BF_{10}
Intercept	515.30	19.42	476.89 – 533.55	
Expectation	3.82	1.61	0.64 – 6.90	0.49
Condition	-1.68	2.33	-6.23 – 2.94	0.04
Exposure	-21.29	4.38	-29.92 – 12.70	>1000
Expectation × Condition	3.34	3.28	-3.11 – 9.85	0.08
Expectation × Exposure	-3.47	2.51	-8.35 – 1.42	0.13
Condition × Exposure	1.12	2.58	-3.86 – 6.15	0.06
Expectation × Condition × Exposure	0.37	5.08	-9.60 – 10.22	0.10

Table 8

Fixed effects the model of antedating condition on reaction times in Experiment 3. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	BF_{10}
Intercept	474.88	10.37	454.81 – 495.21	

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Expectation	10.81	2.83	5.04 – 16.16	214.11
Exposure	-23.39	3.70	-30.63 – -16.08	>1000
Expectation × Exposure	-9.01	4.00	-16.83 – -1.18	3.65

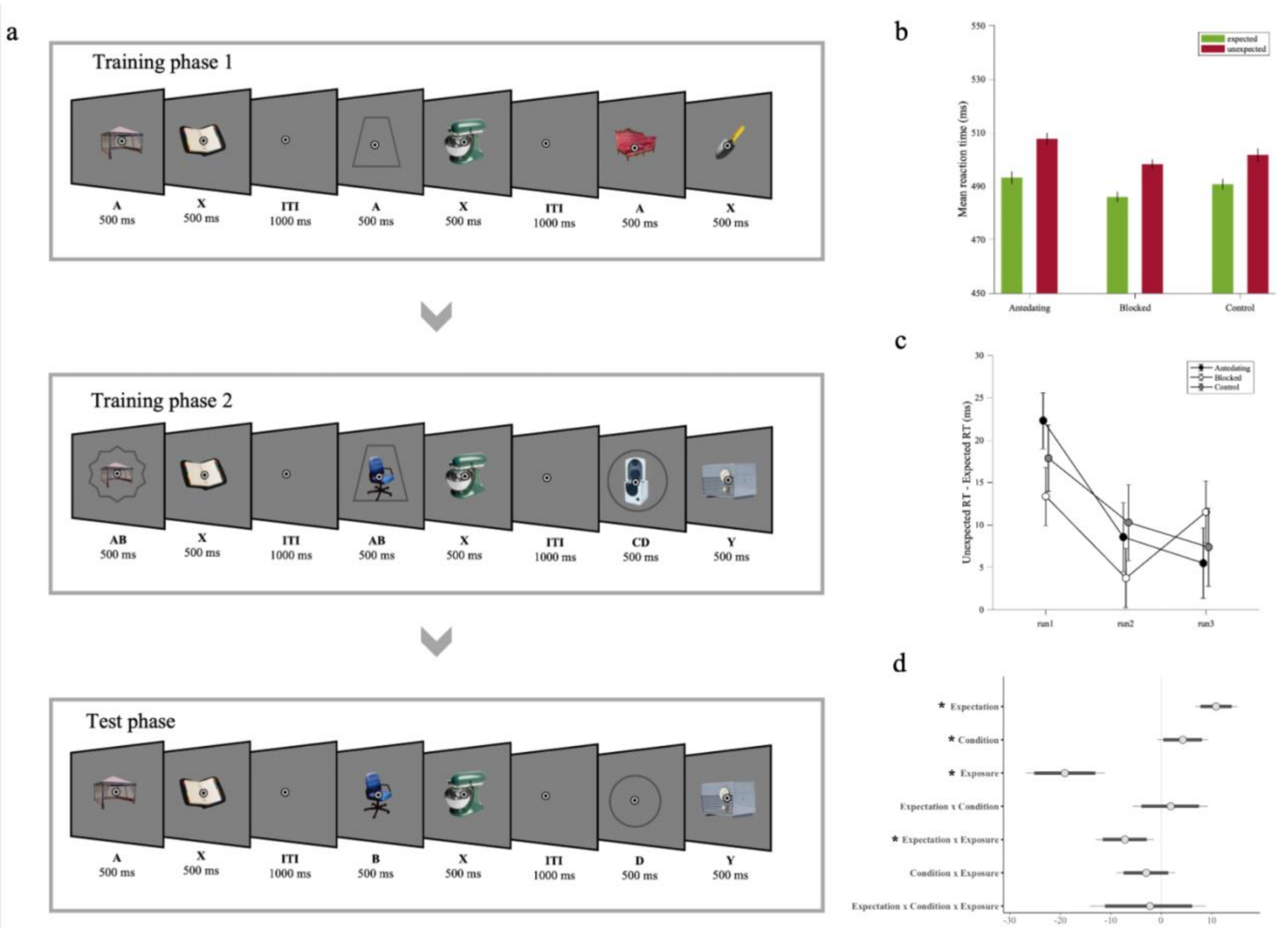
Table 9

Fixed effects the model of blocked and control conditions on reaction times in Experiment 3. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	BF_{10}
Intercept	487.75	9.52	469.42 – 506.90	
Expectation	7.92	2.18	3.57 – 12.23	98.81
Condition	5.87	3.71	-1.38 – 13.26	0.18
Exposure	-29.01	4.09	-36.93 – -20.88	>1000
Expectation × Condition	-9.48	4.49	-18.26 – -0.45	0.54
Expectation × Exposure	-1.05	2.92	-6.78 – 4.67	0.46
Condition × Exposure	-3.33	3.20	-9.63 – 2.97	0.11
Expectation × Condition × Exposure	-1.63	6.45	-14.19 – 11.00	0.14

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1



Experimental procedure and results of Experiment 1

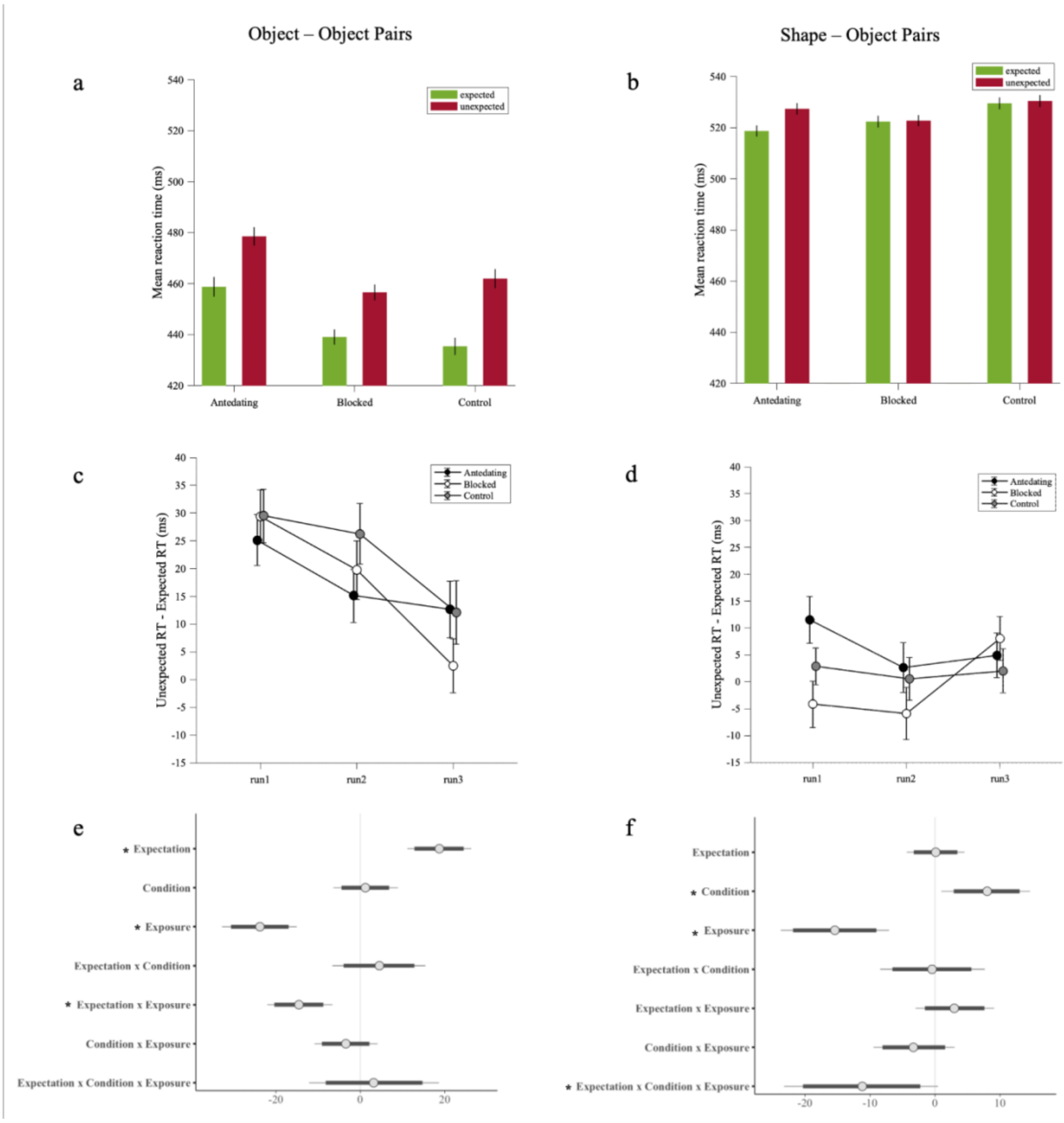
Note. (a) Experiment 1 comprised two training phases (training phase 1 and training phase 2) and a test phase. On every trial throughout the experiment, participants saw a pair of consecutively presented stimuli, i.e., a leading image followed by a trailing image. In training phase 1, the antedating leading stimulus (i.e., A), which could be either a shape or object, was followed by a specific trailing object. In training phase 2, a novel blocked leading stimulus (i.e., B) was presented in compound, along with the antedating (A) leading stimulus (i.e., AB), and followed by the same trailing object from the antedating stimulus in training phase 1. In addition, we introduced novel control compound leading (i.e., CD) and trailing (i.e., Y) stimuli. In the test phase, antedating, blocked or control leading stimuli were followed by the associated (expected) or not associated (unexpected) trailing object. Throughout the experiment, participants performed a categorization task on the trailing object. They reported, as fast as possible, whether the trailing object was electronic or non-electronic. (b) Across participants' mean reaction times as a function of Expectation (expected / unexpected) and Condition (antedating / blocked / control). Participants responded faster to expected than unexpected trailing objects in each condition. There was no difference between blocked and control conditions. (c) Across participants' mean reaction time difference between expected and unexpected trials as a function of time. Please note that we split data into successive runs for visualization purposes only; data analysis was performed with number of trials as a continuous fixed factor (Exposure). Associations were rapidly extinguished during the

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

test phase. Extinction was not different between conditions. (d) Posterior coefficient estimates of effects of the model jointly analyzing blocked and control conditions with error bars representing 95% confidence intervals. Estimates indicate significant results when they do not overlap with zero.

Figure 2

Results of Experiment 1 as a function of Stimulus Type



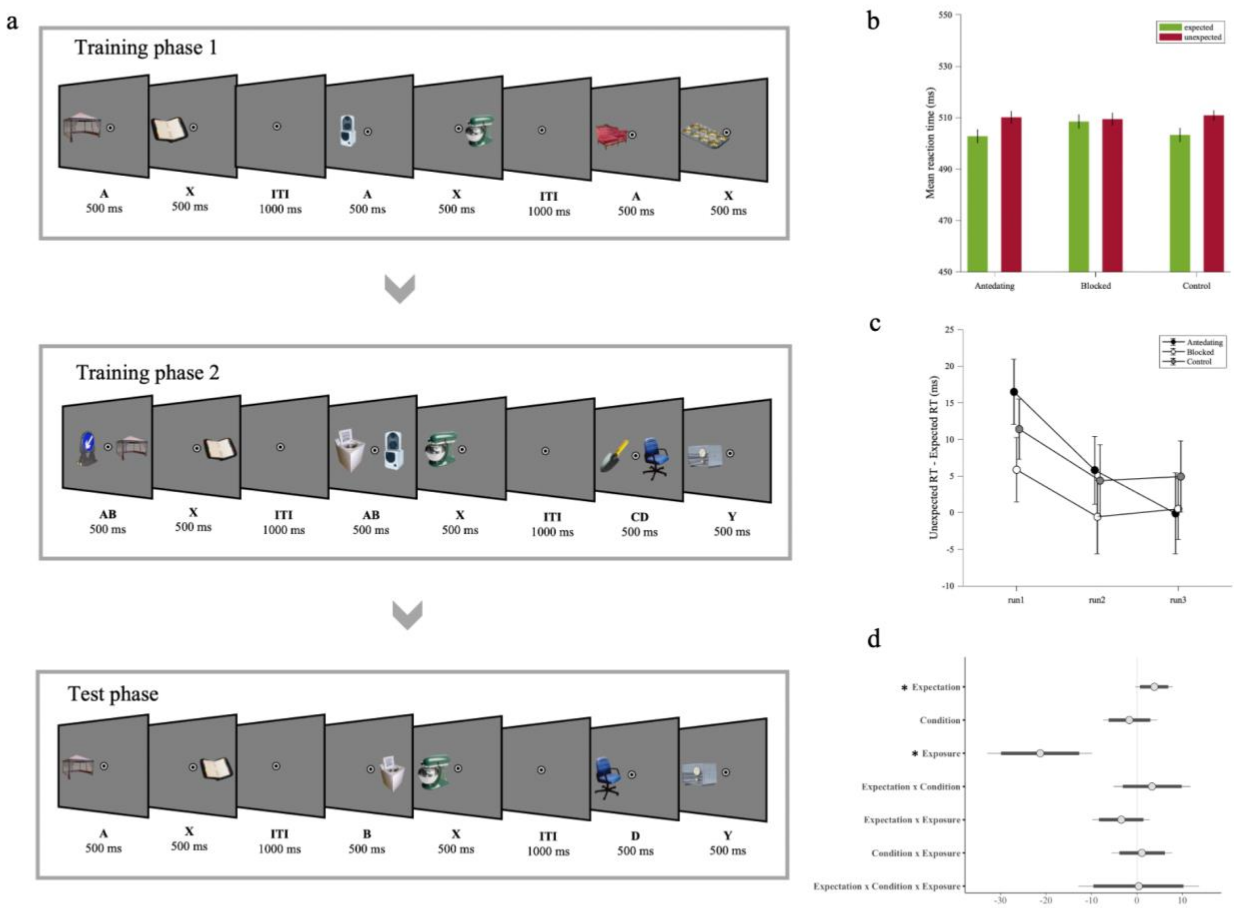
Note. (a-b) Across participants' mean reaction times as a function of Expectation (expected / unexpected) and Condition (antedating / blocked / control) in leading objects (a) and leading shapes (b). The difference between expected and unexpected reaction times was larger for stimulus pairs with leading objects, compared to leading shapes. (c-d) Across participants' mean reaction time difference between expected and unexpected trials as a function of time in leading objects (c) and leading shapes (d). The decrease in reaction time difference between expected and unexpected trials over exposure showed rapid extinction in learning only in leading objects. (e-f) Posterior coefficient

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

estimates of effects of the model jointly analyzing blocked and control conditions with error bars representing 95% confidence intervals in leading objects (e) and leading shapes (f). Estimates indicate significant results when they do not overlap with zero.

Figure 3

Experimental procedure and results of Experiment 2

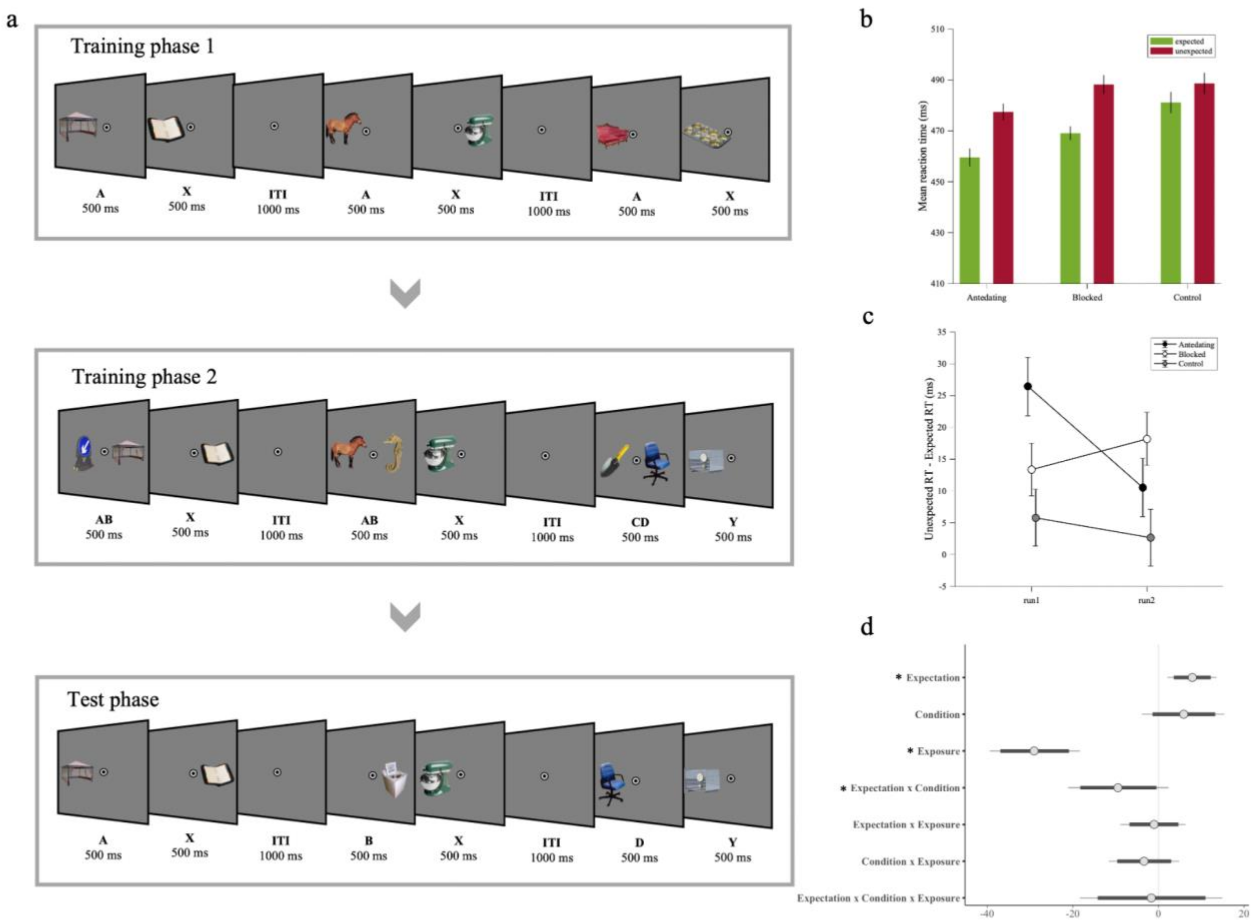


Note. (a) The design and procedure of experiment 2 was identical in all respects to experiment 1, apart from the fact that the leading stimulus was an object presented in the left or right side of the fixation point, and it was followed by the trailing object presented in the left or right side of the fixation point. (b) Across participants' mean reaction times as a function of Expectation (expected / unexpected) and Condition (antedating / blocked / control). Reaction times were faster to expected than unexpected trailing objects in blocked and control conditions. There was no difference between blocked and control condition in terms of reaction time difference between expected and unexpected trials, providing evidence for the absence of blocking effect. (c) Across participants' mean reaction time difference between expected and unexpected trials as a function of time. The decrease in reaction time difference between expected and unexpected trials over exposure showed rapid extinction in learning antedating condition. (d) Posterior coefficient estimates of effects of the model jointly analyzing blocked and control conditions with error bars representing 95% confidence intervals. Estimates indicate significant results when they do not overlap with zero.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4

Experimental procedure and results of Experiment 3



Note. (a) The design and procedure of experiment 3 was identical in all respects to experiment 2, apart from the addition of an oddball detection task on the leading stimuli in the training phases: participants reported as soon as they saw an animate leading stimulus. (b) Across participants' mean reaction times as a function of Expectation (expected / unexpected) and Condition (antedating / blocked / control). Reaction times were faster to expected than unexpected trailing objects in each condition. The reaction time difference between expected and unexpected trials was greater in blocked than control trials, providing evidence for the absence of blocking effect and the augmentation of learning. (c) Across participants' mean reaction time difference between expected and unexpected trials as a function of time. The decrease in reaction time difference between expected and unexpected trials over exposure showed rapid extinction in learning antedating condition. (d) Posterior coefficient estimates of effects of the model jointly analyzing blocked and control conditions with error bars representing 95% confidence intervals. Estimates indicate significant results when they do not overlap with zero.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary information

Supplementary text

In addition to the Bayesian analysis reported in the main text, we conducted classic frequentist analyses using R (i.e., ANCOVA of reaction time data in the test phase and one-way ANOVA of accuracy data in the pair recognition test) for comparability with previous studies and to verify that our conclusions do not depend on the analytical framework employed.

Analyses of RT data in test phase using ANCOVA

In line with the Bayesian analysis, we first conducted a one-way ANCOVA to determine whether there was a significant difference between reaction times of expected and unexpected trials in the antedating condition, while controlling for the amount of exposure. Secondly, we performed a 2 (Expectation: expected, unexpected) × 2 (Condition: control, blocked) ANCOVA to determine whether there was a significant difference between reaction times of expected and unexpected trials in the control and blocked conditions, while controlling for the amount of exposure. To determine the effect of Stimulus Type in Experiment 1, we conducted a 2 (Expectation: expected, unexpected) × 2 (Stimulus Type: object, shape) ANCOVA and a 2 (Expectation: expected, unexpected) × 2 (Condition: control, blocked) × 2 (Stimulus Type: object, shape) ANCOVA. In line with the primary analysis, the contrasts of these factors were coded as successive difference contrasts.

In Experiment 1, the main effect of Expectation was significant in the antedating condition after controlling for Exposure ($F(1, 2382) = 31014.445$, $p < 0.001$, partial $\eta^2 = 0.93$). The main effects of Expectation ($F(1, 4721) = 0.470$, $p = 0.49$, partial $\eta^2 = 9.95e\text{-}5$) and Condition ($F(1, 4721) = 0.012$, $p = 0.91$, partial $\eta^2 = 2.46e\text{-}6$) were not significant across control and blocked trials The Expectation × Condition interaction was not significant ($F(1, 4721) = 0.858$, $p = 0.35$, partial $\eta^2 = 1.82e\text{-}4$).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In the analysis split by stimulus type in Experiment 1, in the antedating condition, the main effect of expectation was significant after controlling for exposure ($F(1, 4502) = 45911.333$, $p < 0.001$, partial $\eta^2 = 0.91$), but the main effect of leading stimulus type ($F(1, 4502) = 0.001$, $p = 0.98$, partial $\eta^2 = 1.40e{-7}$) and the interaction effect between expectation and leading stimulus type ($F(1, 4502) = 0.001$, $p = 0.98$, partial $\eta^2 = 1.60e{-7}$) were not significant. In blocked and control conditions, the main effect of expectation ($F(1, 8348) = 0.211$ $p = 0.65$, partial $\eta^2 = 2.53e{-5}$), leading stimulus type ($F(1, 8348) = 0.176$, $p = 0.68$, partial $\eta^2 = 2.11e{-5}$) and condition ($F(1, 8348) = 0$, $p = 0.99$, partial $\eta^2 = 1.95e{-8}$) and the interaction between expectation and leading stimulus type ($F(1, 8348) = 0.005$, $p = 0.95$, partial $\eta^2 = 5.62e{-7}$), and the interaction between expectation and condition ($F(1, 8348) = 0.689$, $p = 0.41$, partial $\eta^2 = 8.25e{-5}$), and the interaction between leading stimulus type and condition ($F(1, 8348) = 0.864$, $p = 0.35$, partial $\eta^2 = 1.03e{-4}$) and the interaction between expectation, leading stimulus type and condition ($F(1, 8348) = 0.531$, $p = 0.47$, partial $\eta^2 = 6.36e{-5}$) were not significant. Overall, in Experiment 1, the ANCOVA analysis confirmed the results of the Bayesian mixed effect model analysis reported in the main text: in the antedating condition, we found successful learning of repeated stimulus pairs and the consequent behavioral benefit of expectation in terms of response speed; crucially, we found no blocking effect for incidentally learned stimulus pairs.

In Experiment 2, the main effect of Expectation was significant in the antedating condition after controlling for Exposure ($F(1, 1177) = 7357.152$, $p < 0.001$, partial $\eta^2 = 0.86$). The main effects of Expectation ($F(1, 2314) = 0.002$, $p = 0.96$, partial $\eta^2 = 7.18e{-7}$) and Condition ($F(1, 2314) = 0.375$, $p = 0.54$, partial $\eta^2 = 1.62e{-4}$) were not significant across control and blocked trials. The Expectation × Condition interaction was not significant ($F(1, 2314) = 0.05$, $p = 0.94$, partial $\eta^2 = 2.16e{-6}$). Overall, in Experiment 2, the ANCOVA analysis showed successful learning of repeated stimulus pairs in the antedating condition; crucially, we again found no blocking effect for incidentally learned stimulus pairs.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In Experiment 3, the main effect of Expectation was significant in the antedating condition after controlling for Exposure ($F(1, 792) = 106e+4$, $p < 0.001$, partial $\eta^2 = 0.93$). Across control and blocked trials, the main effect of Expectation ($F(1, 1577) = 4.329$, $p = 0.04$, partial $\eta^2 = 2.74e-3$) was significant, but the main effect of Condition ($F(1, 1577) = 0.621$, $p = 0.43$, partial $\eta^2 = 3.93e-4$) was not significant. The Expectation × Condition interaction was not significant ($F(1, 1577) = 0.202$, $p = 0.65$, partial $\eta^2 = 1.28e-4$). Overall, in Experiment 3, the ANCOVA analysis confirmed the results of the Bayesian mixed effect model analysis: in the antedating condition, we found successful learning of repeated stimulus pairs; crucially, we found no blocking effect for incidentally learned stimulus pairs.

Analyses of accuracy data in pair recognition test using ANOVA and t-test

In line with the Bayesian analysis, we first conducted a one-sample t-test to determine whether the level of accuracy was above chance level in each condition. Secondly, we performed a one-way (Condition: control – blocked) ANOVA to test for the blocking effect.

In Experiment 1, the level of accuracy was above chance level in the antedating ($t(99) = 6.862$, $p < 0.001$, Cohen's $d = 0.68$), blocked ($t(99) = 4.117$, $p < 0.001$, Cohen's $d = 0.41$) and control ($t(99) = 3.164$, $p < 0.01$, Cohen's $d = 0.32$) conditions. Secondly, the one-way ANOVA showed that the main effect of Condition ($F(1, 198) = 0.69$, $p = 0.41$, partial $\eta^2 = 3.47e-3$) was not significant. Overall, in Experiment 1, the ANOVA analysis confirmed the results of the Bayesian mixed effect model analysis reported in the main text: we found clear explicit knowledge of incidentally learned associations in each condition and no blocking effect for such explicit knowledge.

In Experiment 2, the level of accuracy was below chance level in the antedating ($t(49) = -0.035$, $p = 0.97$, Cohen's $d = -5.08e3$), blocked ($t(49) = -0.812$, $p = 0.42$, Cohen's $d = -0.11$) and control ($t(49) = 0.076$, $p = 0.94$, Cohen's $d = 0.01$) conditions. Secondly, the one-way ANOVA showed that the main effect of Condition ($F(1, 98) = 0.2374$, $p = 0.54$, partial $\eta^2 = 3.80e-3$) was not

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

significant. Overall, in Experiment 2, the ANOVA analysis confirmed the results of the Bayesian mixed effect model analysis reported in the main text: we found no explicit knowledge of incidentally learned associations in each condition and, consequently, no blocking effect.

In Experiment 3, the level of accuracy was above chance level in the antedating ($t(49) = 6.368$, $p < 0.001$, Cohen's $d = 0.90$), blocked ($t(49) = 4.599$, $p < 0.001$, Cohen's $d = 0.65$) and control ($t(49) = 5.481$, $p < 0.001$, Cohen's $d = 0.78$) conditions. Furthermore, the main effect of Condition ($F(1, 98) = 1.012$, $p = 0.31$, partial $\eta^2 = 0.01$) was not significant, indicating the absence of blocking effect for the explicit knowledge of incidentally learned associations. Overall, in Experiment 3, the ANOVA analysis confirmed the results of the Bayesian mixed effect model analysis reported in the main text: we found clear explicit knowledge of incidentally learned associations in each condition and no blocking effect for such explicit knowledge.

Supplementary tables

Table S1

Fixed effects the post-hoc model of antedating condition on reaction times of leading objects in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	*BF_{10}*
Intercept	475.08	9.53	456.05 – 493.87	
Expectation	15.19	3.66	7.98 – 22.46	175.97
Exposure	-20.33	4.36	-28.85 – -11.90	>1000
Expectation × Exposure	-8.22	3.94	-16.06 – -0.75	0.78

Table S2

Fixed effects the post-hoc model of antedating condition on reaction times of leading shapes in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	*BF10*
Intercept	529.39	7.58	514.68 – 544.25	
Expectation	5.44	2.36	0.83 – 10.05	0.61
Exposure	-9.94	3.43	-16.74 – -3.24	6.47
Expectation × Exposure	-6.51	3.50	-13.21 – 0.30	0.36

Table S3

Fixed effects the post-hoc model of blocked and control conditions on reaction times of leading objects in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	*BF10*
Intercept	456.23	9.63	437.31 – 475.28	
Expectation	18.73	2.95	12.83 – 24.50	>1000
Condition	1.22	2.91	-4.46 – 6.83	0.04
Exposure	-23.80	3.46	-30.64 – -16.99	>1000
Expectation × Condition	4.51	4.24	-3.96 – 12.81	0.11
Expectation v Exposure	-14.54	2.96	-20.38 – -8.76	>1000
Condition × Exposure	3.21	5.94	-8.19 – 14.73	0.12
Expectation × Condition × Exposure	-3.43	2.86	-9.11 – 2.16	0.14

Table S4

Fixed effects the post-hoc model of blocked and control conditions on reaction times of leading shapes in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	*BF10*
Intercept	530.92	7.75	515.88 – 546.07	

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479428; this version posted February 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Expectation	0.11	1.69	-3.27 – 3.44	0.03
Condition	7.98	2.59	2.88 – 13.02	2.47
Exposure	-15.42	3.26	-21.85 – -9.02	>1000
Expectation × Condition	-0.46	3.06	-6.56 – 5.59	0.04
Expectation × Exposure	2.96	2.34	-1.56 – 7.60	0.10
Condition × Exposure	-3.31	2.45	-8.08 – 1.55	0.13
Expectation × Condition × Exposure	-11.22	4.64	-20.32 – -2.27	1.55

Table S5

Fixed effects the post-hoc model of control condition on reaction times in Experiment 3. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	BF_{10}
Intercept	491.09	9.73	472.16 – 510.23	
Expectation	4.36	2.59	-0.73 – 9.51	1.16
Exposure	-30.03	4.33	-38.34 – -21.45	>1000
Expectation × Exposure	-2.02	3.71	-9.26 – 5.21	0.62

Table S6

Fixed effects the post-hoc model of blocked condition on reaction times in Experiment 3. Estimate, estimation error, lower/upper limit of 95% profile confidence intervals, bayes factor.

Predictors	*Estimate*	*Est. Error*	*CI (95%)*	BF_{10}
Intercept	485.56	9.70	466.60 – 504.68	
Expectation	10.11	2.65	4.82 – 15.16	277.17
Exposure	-27.38	3.94	-35.05 – -19.58	>1000
Expectation × Exposure	-0.95	3.68	-8.26 – 6.25	0.57