

## RESEARCH ARTICLE

## Forward and backward blocking in statistical learning

Ilayda Nazlı<sup>1\*</sup>, Ambra Ferrari<sup>1,2</sup>, Christoph Huber-Huber<sup>1,3</sup>, Floris P. de Lange<sup>1</sup>

**1** Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands, **2** Max Plank Institute for Psycholinguistics, Nijmegen, The Netherlands, **3** Center for Mind/Brain Sciences (CIMeC), University of Trento, Rovereto, Italy

\* [ilayda.nazli@donders.ru.nl](mailto:ilayda.nazli@donders.ru.nl)

## 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. One paradigm often used to study error-driven learning is blocking. In forward blocking, participants are first presented with stimulus A, followed by outcome X ( $A \rightarrow X$ ). In the second phase, A and B are presented together, followed by X ( $AB \rightarrow X$ ). Here,  $A \rightarrow X$  blocks the formation of  $B \rightarrow X$ , given that X is already fully predicted by A. In backward blocking, the order of phases is reversed. Here, the association between B and X that is formed during the first learning phase of  $AB \rightarrow X$  is weakened when participants learn exclusively  $A \rightarrow X$  in the second phase. The present study asked the question whether forward and backward blocking occur during visual statistical learning, i.e., the incidental learning of the statistical structure of the environment. In a series of studies, using both forward and backward blocking, we observed statistical learning of temporal associations among pairs of images. While we found no forward blocking, we observed backward blocking, thereby suggesting a retrospective reevaluation process in statistical learning and supporting a functional similarity between statistical learning and reinforcement learning.

## OPEN ACCESS

**Citation:** Nazlı İ, Ferrari A, Huber-Huber C, de Lange FP (2024) Forward and backward blocking in statistical learning. PLoS ONE 19(8): e0306797. <https://doi.org/10.1371/journal.pone.0306797>

**Editor:** Fernando Blanco, University of Granada: Universidad de Granada, SPAIN

**Received:** November 24, 2023

**Accepted:** June 24, 2024

**Published:** August 5, 2024

**Copyright:** © 2024 Nazlı et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The minimal underlying data are freely available on the Donders Repository (<https://doi.org/10.34973/pwza-qh43>). To access the data, ensure you are logged in by clicking the “Login” button at the top left corner of the page. You can optionally log in using your ORCID account. Once registered and logged in, you can request access by confirming your compliance with all regulations imposed by the ethics committee and acknowledging that you will use the data responsibly and credit it when publicly presenting any results derived from it.

## Introduction

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 [1]. For example, in a typical classical conditioning experiment [2], 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 category of phenomena 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 [3–6].

**Funding:** This work was supported by a personal scholarship provided to İ.N. by the Ministry of National Education of the Republic of Türkiye, and by a personal grant provided to F.P.d.L. by the European Union (ERC Consolidator Grant 101000942, “Surprise”). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

One key example of cue competition is Kamin blocking, also known as forward blocking [7]. In a typical forward blocking paradigm (see Table 1), observers first learn the association between cue A and outcome X ( $A \rightarrow X$ ), and later they are trained with the association between cues A + B and outcome X ( $AB \rightarrow X$ ). As a result of forward blocking, observers learn the association between cue B and outcome X less strongly, 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. Forward blocking cannot be explained by simple contiguity-dependent Hebbian associative learning [8]. 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 [9] provides an explanation for blocking (though see [10] for a modification of the traditional model). 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 forward blocking procedure, the previously learned  $A \rightarrow X$  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  $A \rightarrow X$  pairs in the first training phase.

A similar, but distinct form of cue competition is backward blocking, which is an example of retrospective revaluation: a change in the associative strength occurs because the association between the companion cue (i.e., the cue that is previously associated with the target cue and outcome) and the outcome is reevaluated. In the backward blocking paradigm [11], observers are first trained with  $AB \rightarrow X$  association, and subsequently with  $A \rightarrow X$  association. In spite of the reversed order of training phases compared to forward blocking, backward blocking leads to a similar outcome as forward blocking: a lack of association between blocked cue B and outcome X. Here, in the first training phase, both A-X and B-X associations are formed equally (i.e., depending on the saliency of cues). However, in the second training phase, as observers are trained with  $A \rightarrow X$  association, the associative strength between cue A and outcome X becomes stronger, which in turn weakens the association between cue B and outcome X. While this form of retrospective revaluation cannot be explained by the traditional Rescorla-Wagner model, as this model assumes that the relevant cue must be present in order to change the associative strength [9,12,13], backward blocking can be successfully modeled by a slightly revised version of the traditional model. For example, backward blocking can be explained by a Rescorla-Wagner learning model that assigns non-zero salience to non-presented blocked stimuli whose memories or representations are retrieved by competing stimuli that had previously been paired with those blocked stimuli [14] or by a Bayesian generalization of the Rescorla-Wagner model, the Kalman filter [12,15], where the weights of all possible cues are updated simultaneously, and the sum of all possible weights equals to 1.

In typical blocking experiments, associations are learned either when the outcome is a reward [16–19] or when performance-related feedback is provided [5,20–25]. This provides

**Table 1. General experimental design.**

	<i>Training phase 1</i>	<i>Training phase 2</i>	<i>Test phase</i>
	A → X	AB → X	A → X
Forward blocking		CD → Y	B → X D → Y
Backward blocking	AB → X CD → Y	A → X	A → X B → X D → Y

Note. Letters denote conditions (i.e., A for Antedating, B for Blocked, C-D for Control).

<https://doi.org/10.1371/journal.pone.0306797.t001>

support that reinforcement learning (i.e., learning associations between events via trial and error) relies on an error-driven learning algorithm [26]. Another powerful form of learning is known as statistical learning, often defined as the incidental extraction of regularities from the environment without intention [27–31]. In the context of statistical learning, we have limited information about how the learning process itself occurs. Several studies suggest 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 [18,32]. In humans, statistical learning involves the ventral striatum [33], which has been hypothesized to signal prediction errors [33–35]. However, other researchers, using variants of forward blocking, did not find clear-cut evidence for error-driven statistical learning. Beesley and Shanks [36] did not observe any forward blocking in contextual cueing experiments, where participants incidentally learnt the spatial relationship among distractors and targets in a visual search task. This procedure however deviates from classic forward blocking paradigms, which rely on a *temporal* prediction between a cue and a future outcome [4,5,16,17,19–23,25,37,38]. Two subsequent experiments [6,39] observed forward blocking of temporal associations only for material that was intentionally learnt, but not for incidentally learnt stimulus associations. Such learning conditions substantially deviate from a typical statistical learning scenario, where observers extract regularities without intention [27,28,30,31]. While few studies investigated forward blocking in incidental learning [6,20,36,39], less is known about backward blocking in incidental learning. Importantly, there is evidence of retrospective revaluation (of which backward blocking is an instance) not only in adults and children [40–43] but also in 8-month-old infants [44,45], who clearly did not follow any explicit task instructions. This suggests that backward blocking may be present even in incidental learning, where observers attune themselves to statistical regularities by simple passive exposure.

According to the attentional account model [46], forward blocking occurs as a result of less attention devoted to the blocked cue. Namely, in the first training phase observers learn that cue A predicts outcome X and cue A therefore becomes a relevant and attended cue. Therefore, in the second training phase where novel cue B is presented together with cue A, more attention is paid to the relevant and predictive cue A, compared to the novel predictive cue B. The lower attention to cue B leads to the failure to associate cue B with outcome X. This attentional explanation would however not be able to explain backward blocking, given that equal attention is paid to cues A and B in the first phase. Hence, the use of backward blocking enables us to examine more closely the attentional account of blocking and, crucially, test for retrospective revaluation in statistical learning. Therefore, we set out to examine forward and backward blocking during statistical learning in a series of experiments. In some statistical learning experiments, participants are exposed to a continuous stream of stimuli containing statistical regularities [27,29,47–50]. Other studies have instead presented two successive stimuli on each trial, with conditional probabilities controlling their pairing [51,52]. In terms of neural processing, both continuous streams [53] and pairs [54] show identical modulations of sensory responses after statistical learning, suggesting that both paradigms elicit similar learning processes. We opted for pairs of stimuli in order to connect our study to the classic forward and backward blocking paradigms [7,11]. On every trial, we presented participants with two consecutive visual object stimuli and asked them to categorize the trailing object as either electronic or non-electronic. Unbeknownst to participants, we manipulated the conditional probabilities between the 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 and measuring

their reaction time for categorization judgments of the trailing image. Successful learning was indexed by faster reaction times to expected relative to unexpected trailing stimuli [49,52,55].

## Experiment 1

### Method

**Preregistration and data availability.** All experiments were preregistered on the Open Science Framework (<https://osf.io/r243e> for Experiment 1; <https://osf.io/7kmtv> for Experiment 2). All data and code used for the analyses are freely available on the Donders Repository (<https://doi.org/10.34973/pwza-qh43>). Deviations from the preregistration are mentioned as such and justified in the corresponding sections below.

**Participants.** The experiment was performed online by using the Gorilla platform [56], and participants were recruited through the Prolific platform (<https://www.prolific.co/>). 92 participants performed the experiment. 42 of them were excluded based on a priori exclusion criteria (see section ‘Exclusion and inclusion criteria’ below) before they started the second training phase (i.e. before the relevant data for the analysis was collected). Importantly, our selection criteria applied to a very simple task where the general population is expected to score at ceiling [51,52,57]. Thus, our criteria (i.e. accuracy below 80%) allowed us to exclude outliers who clearly underperformed either because they did not read the instructions carefully, or did not understand the requirements of the task, or did not pay enough attention to stimuli; accordingly, it is common in online experiments that approximately half of the participants shows careless and inattentive behavior [58,59]. Consequently, we carried out our analyses on a subset of the population who showed high motivation and adequate attention to the stimuli, as required to support statistical learning (Richter & de Lange, 2019). 50 participants (18 females; mean age 25.80, range 18–40 years) were included in the final data analysis. In Supplementary Forward Blocking Experiment 1 with 100 participants (see Supplementary information 2), we found successful learning of stimulus transition probabilities ( $b = 11.23$ ,  $CI = [6.80, 15.59]$ , Cohen’s  $d_z = 0.54$ ). From this observation, we concluded that 50 participants was an adequate sample size for Experiment 1.

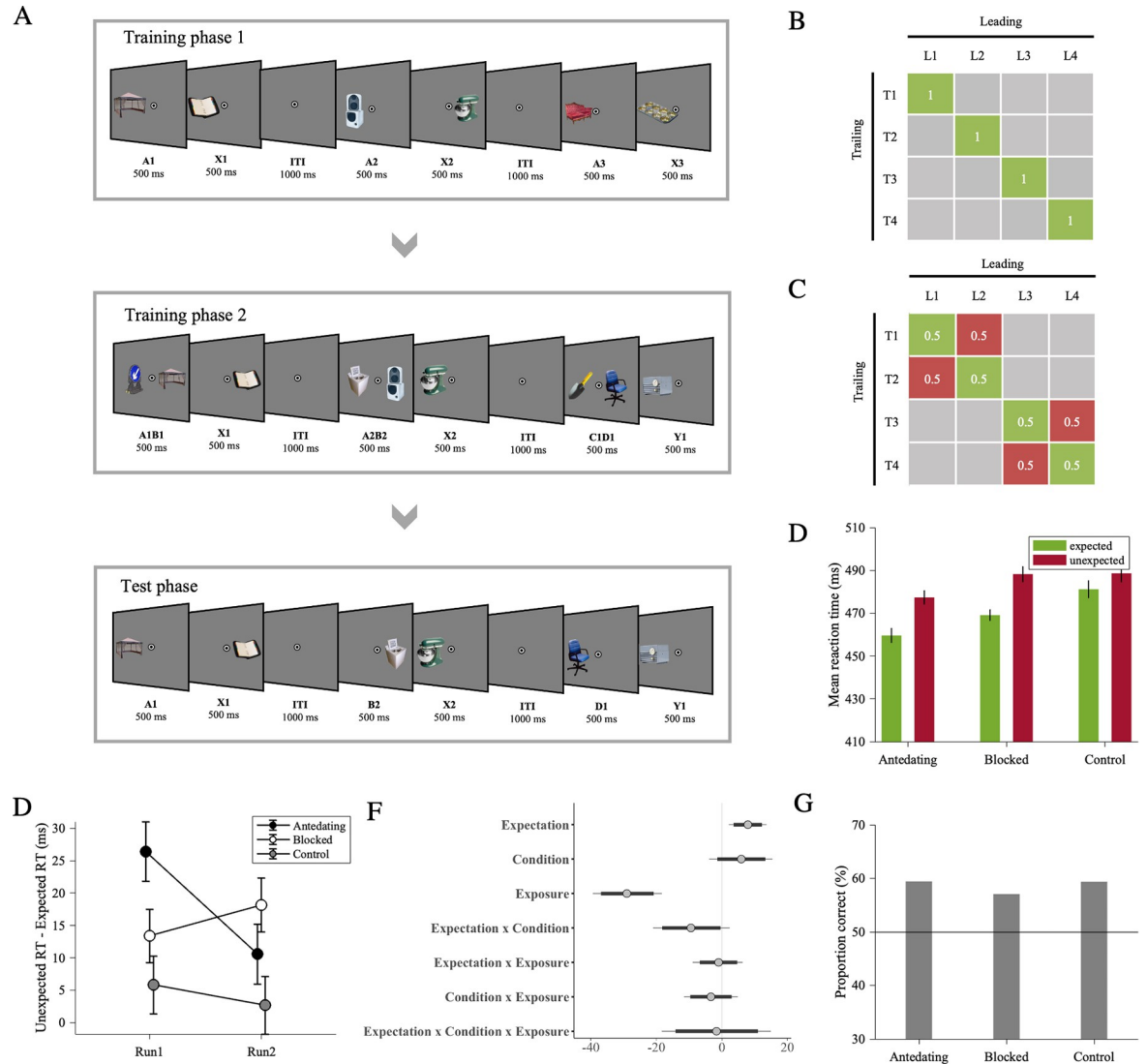
All participants had normal or corrected to normal vision, normal hearing and no history of neurological or psychiatric conditions. They provided written informed consent and received financial reimbursement (8 euro per hour) for their participation in the experiment. The study followed the guidelines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

**Experimental design.** In each experimental trial, participants were exposed to two images presented on the left or right side of the central fixation point in quick succession: a leading stimulus was followed by a trailing stimulus. For each participant, there were 4 leading objects and 4 trailing stimuli 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 objects in which the leading object predicted the identity of the trailing object, thus over time making the trailing object expected given the leading object. Importantly, each trailing object was only (un)expected depending on which leading object it was preceded by. Thus, each trailing object served both as an expected and unexpected object depending on the leading object at test phase. 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. In addition, object 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. Also, both hemi-fields were equally task-relevant, which fostered participants' attention to both sides. 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 object could be used to predict the correct categorization response before the trailing object appeared. In addition to the main object categorization task, there was an oddball detection task involving the leading stimuli in the training phases (16% of all trials per participant): participants were required to press a specific button as soon as they saw an animate leading stimulus. 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 [60]. Finally, 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 Fig 1A). During the two training phases, leading objects were perfectly predictive of their respective trailing objects (i.e.  $P(\text{trailing} \mid \text{leading}) = 1$ ); see Fig 1B). 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 likely learned incidentally. Note that the participants may, however, still develop explicit knowledge of the associations over the course of the experiment, which we tested in a final recognition task. In training phase 1, the leading object (A) was always followed by the same trailing object (X). In training phase 2, a novel leading object (blocked [B] leading object) was presented along with the leading object presented in training phase 1 (antedating [A] leading stimulus), hence creating a compound stimulus (AB). This was followed by the same trailing object (X) as in training phase 1. In addition, two novel leading (object + object [CD]) and a trailing (object [Y]) objects were presented as a control condition. In the test phase, the leading stimulus of each condition (antedating [A] / blocked [B] / control [D]) was presented alone, followed by either the expected trailing object (based on the training phases), or an unexpected trailing object. Expected and unexpected object pairs were presented equally often to prevent any learning at this final test stage (see Fig 1C). In the test phase, control (D) trials were compared to blocked (B) trials to assess blocking while controlling for the amount of exposure. It should be noted that the amount of exposure to trailing object X and trailing object Y are not the same, given that trailing object Y was only introduced in the second learning phase. This difference is an inevitable feature in classic blocking paradigms. If we would have presented trailing object Y in isolation in an additional experimental phase or if we would have paired Y with another leading stimulus in training phase 1, this could have elicited latent inhibition (i.e., difficulty in learning associations as a result of pre-exposure, [61]). Thus, we opted for the classic blocking paradigm. Furthermore, the control trials in the test phase allowed us to assess whether new associations were learned during training phase 2.

Data was collected during one single session per participant. Firstly, participants familiarized themselves with all trailing objects (both X and Y). In each trial, an object image was presented for 3500 ms, and participants had 1500 ms to categorize the object image as electronic



**Fig 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 object followed by a trailing object. In training phase 1, the antedating leading object (i.e., A) was followed by a specific trailing object. In training phase 2, a novel blocked leading object (i.e., B) was presented in compound, along with the antedating (A) leading object (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) objects. In the test phase, antedating, blocked or control leading stimuli were followed by the associated (expected) or not associated (unexpected) trailing object. There were four different object pairs for AB→X and CD→Y. 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) Statistical regularities depicted as image transition matrix with stimuli pairs in training phase 1 and training phase 2. Ls represent leading stimuli, and Ts represent trailing stimuli. There were 16 different leading objects and 8 different trailing objects coming from four different AB→X and CD→Y pairs. (C) Statistical regularities depicted as image transition matrix with stimuli pairs in test phase. Green cells represent expected pairs, and red cells represent unexpected pairs. (D) 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. (E) 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). The decrease in reaction time difference between expected and unexpected trials over exposure showed rapid extinction in learning antedating condition. (F) 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. (G) Across participants' proportion correct responses in pair recognition test. Participants showed slightly above chance-level performance in all conditions indicating whether the trailing object was likely or unlikely given the leading object.

<https://doi.org/10.1371/journal.pone.0306797.g001>

or non-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  $\times$  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 objects 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  $\times$  4 trials / pairs = 16 trials in total; practice training phase 2: 8 pairs  $\times$  4 trials / pairs = 32 trials in total). After each practice, participants completed the training phases (training phase 1: 4 object pairs  $\times$  30 trials = 120 trials in total; training phase 2: 8 object pairs  $\times$  30 trials = 240 trials in total). In addition, animate detection and 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  $\times$  16 trials = 192 trials in total). Crucially, for each leading object, 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  $\times$  2 trials / pairs = 24 trials in total) before completing the recognition task (12 pairs  $\times$  8 trials / pairs = 96 trials in total).

**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 Fig 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’).

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).

**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). We did not statistically analyze the accuracy data in the test phase, given that the categorization task was not challenging, and performance was near ceiling levels (97% in Experiment 1 and 97% in Experiment 2). 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. Data were analyzed using the *brm* function of the BRMS package [62] in R. Furthermore, supplementary tables (see S1 File) 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, 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 were 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) to account for individual variance.

Secondly, we determined whether there was blocking by jointly modeling the blocked (B) and control (D) 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  $\times$  Exposure interaction) and its interaction with Condition to probe for potential differences in extinction between conditions. We adjusted the priors of the main effect of Expectation and Exposure and the prior of their interaction based on the posteriors of pilot experiments. 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 less informative. The prior for the Condition effect and its interaction with Expectation, i.e., blocking effect, was centered at zero. Note that specifying the priors in this way turns the estimates of Expectation and Exposure effects of Experiment 1 into the combined evidence from pilot experiments *and* Experiment 1. Crucially, the pattern of results from Experiment 1 was exactly the same when not only the priors for the Condition effect but also for Expectation and Exposure were centered at zero. Further details and the complete model parametrization can be found in the R codes provided on the Donders Repository. The response time data was modelled using the ex-gaussian family and four chains with 25,000 iterations each (12,500 warm up) per chain and inspected for chain convergence. We report posterior fixed effects model coefficients. Coefficients were accepted as convincing statistical evidence, analogously to statistically significant in a frequentist framework, if the associated 95% posterior credible intervals were non-overlapping with zero.

*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. If accuracy was above chance level within each condition, we then 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

**Table 2. Posterior fixed effects of the model of antedating condition on reaction times in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.**

Predictors	Estimate	Est. Error	CI (95%)
Intercept	474.88	10.37	454.81–495.21
Expectation	10.81	2.83	5.04–16.16
Exposure	-23.39	3.70	-30.63 --16.08
Expectation × Exposure	-9.01	4.00	-16.83 --1.18

<https://doi.org/10.1371/journal.pone.0306797.t002>

and inspected for chain convergence. With respect to significance and amount of evidence we used the same criteria as for the RT data.

## Results

### Analyses of RT data in test phase

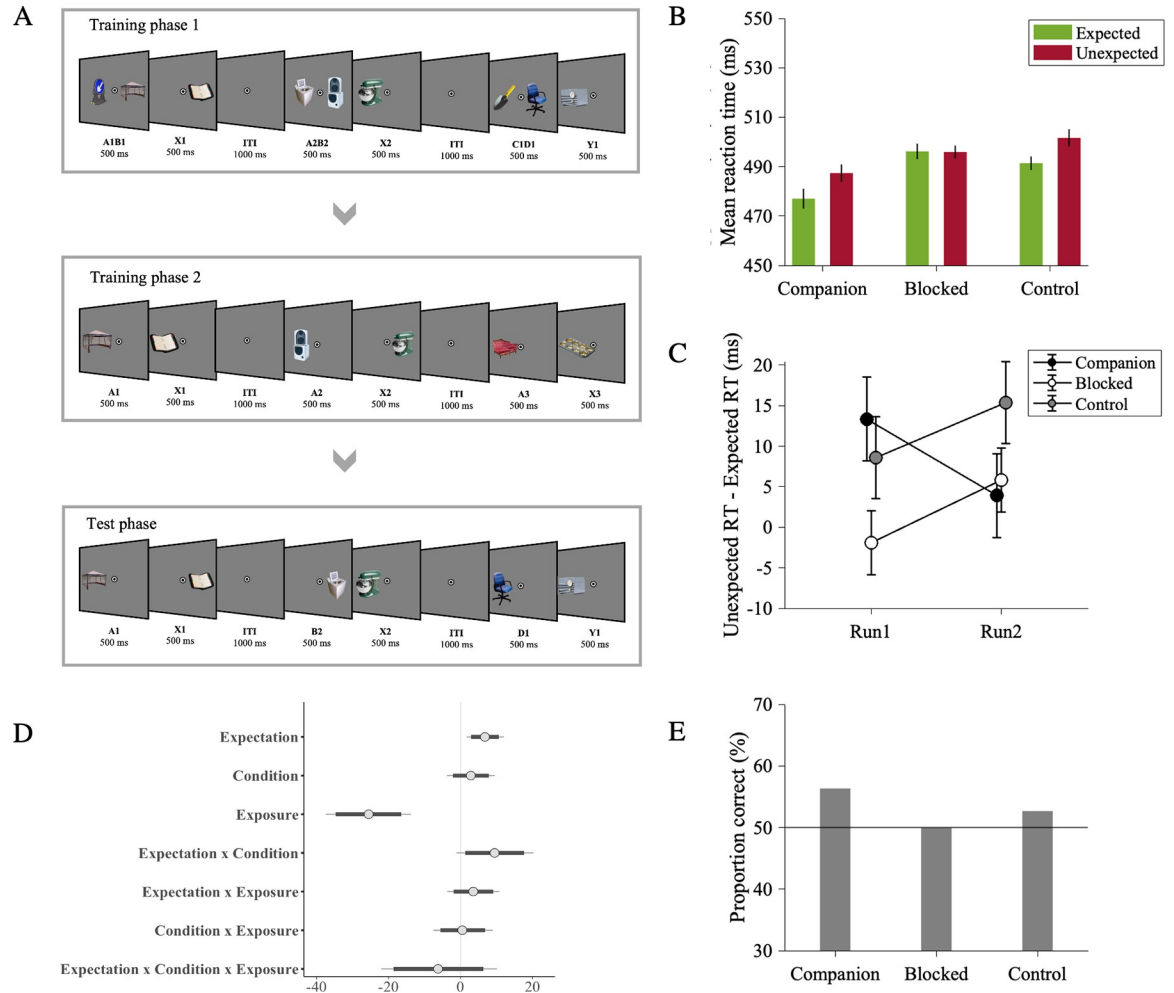
Firstly, we compared the reaction times of expected and unexpected trials in the antedating condition (see Table 2). We observed faster reaction times in expected (460 ms) than in unexpected (477 ms) trials ( $b = 10.81$ ,  $CI = [5.04, 16.16]$ , Cohen's  $d_z = 0.61$ , see Fig 1D), indicating successful 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 ( $b = -9.01$ ,  $CI = [-16.83, -1.18]$ ), indicating that learning showed rapid extinction (expectation effect for run 1: 26 ms, run 2: 11 ms; see Fig 1E).

Next, we modeled the blocked and control conditions to test whether we found blocking (see Table 3 and Fig 1F). There was an interaction effect between expectation and condition ( $b = -9.48$ ,  $CI = [-18.26, -0.45]$ , Cohen's  $d_z = -0.26$ , see Fig 1B). 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 ( $b = 4.36$ ,  $CI = [-0.73, 9.51]$ , Cohen's  $d_z = 0.20$ , see S1 Table in S1 File). On the other hand, reaction times were clearly faster in expected (469 ms) than in unexpected (488 ms) trials of the blocked condition ( $b = 10.11$ ,  $CI = [4.82, 15.16]$ , Cohen's  $d_z = 0.65$ , see S2 Table in S1 File). 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 blocked and control conditions ( $b = -1.63$ ,  $CI = [-14.19, 11.00]$ ); 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 Fig 1C).

**Table 3. Posterior fixed effects of the model of blocked and control conditions on reaction times in Experiment 1. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.**

Predictors	Estimate	Est. Error	CI (95%)
Intercept	487.75	9.52	469.42–506.90
Expectation	7.92	2.18	3.57–12.23
Condition	5.87	3.71	-1.38–13.26
Exposure	-29.01	4.09	-36.93 --20.88
Expectation × Condition	-9.48	4.49	-18.26 --0.45
Expectation × Exposure	-1.05	2.92	-6.78–4.67
Condition × Exposure	-3.33	3.20	-9.63–2.97
Expectation × Condition × Exposure	-1.63	6.45	-14.19–11.00

<https://doi.org/10.1371/journal.pone.0306797.t003>



**Fig 2. 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 order of training phases. (B) Across participants' mean reaction times as a function of Expectation (expected / unexpected) and Condition (companion / blocked / control). Reaction times were faster to expected than unexpected trailing objects in companion and control conditions but not in blocked condition, providing evidence for the presence of backward blocking in statistical learning. (C) Across participants' mean reaction time difference between expected and unexpected trials as a function of time. There was no extinction in learning in any 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. (E) Across participants' proportion correct responses in pair recognition test. Participants showed slightly above chance-level performance in companion condition indicating whether the trailing object was likely or unlikely given the leading object.

<https://doi.org/10.1371/journal.pone.0306797.g002>

### Analysis of accuracy data in pair recognition test

Participants showed slightly above-chance level performance in indicating whether the trailing object was likely or unlikely given the leading object in the antedating (proportion correct = 59%;  $b = 0.39$ ,  $CI = [0.26, 0.51]$ ), blocked (proportion correct = 57%;  $b = 0.29$ ,  $CI = [0.17, 0.42]$ ) and control (proportion correct = 59%;  $b = 0.39$ ,  $CI = [0.24, 0.54]$ ) conditions (see Fig 1G). Response errors did not differ between the blocked and control conditions ( $b = -0.1$ ,  $CI = [-0.08, 0.29]$ ), indicating the absence of blocking effect for the explicit knowledge of incidentally learned associations.

## Experiment 2

In Experiment 1, we observed a stronger reaction time benefit for  $B \rightarrow X$  compared to control, indicating successful learning and the absence of forward blocking. We speculated that this pattern of results may be explained by the following process: upon learning the  $A \rightarrow X$  association in the first training phase, attention may have shifted to the novel (and therefore potentially more salient) leading image B during the second training phase, thereby enhancing the learning of the  $B \rightarrow X$  association. Importantly, this attentional mechanism is not at play in the related, but distinct paradigm of backward blocking [11]. Here, the order of training phases is reversed compared to forward blocking. Observers are first trained with  $AB \rightarrow X$  association and presented in a subsequent training phase with the  $A \rightarrow X$  association. As a result, both leading objects A and B are equally novel and salient during the first training phase and therefore should be learnt equally well. Therefore, we reasoned that backward blocking may allow us to study blocking without the potentially confounding factors related to novelty and salience. Crucially, this paradigm also allowed us to test for the first time whether retrospective reevaluation takes place during incidental statistical learning.

## Method

**Participants.** The experiment was performed online by using the Gorilla platform [56], and participants were recruited through the Prolific platform (<https://www.prolific.co/>). Eighty-four participants performed the experiment. Thirty-three of them were excluded before they finished the experiment based on a priori exclusion criteria (see section ‘Exclusion and inclusion criteria’ below). One participant was excluded from the final data analysis due to overall excessively fast responses (i.e., 93% of responses being less than 200 ms). As a result, 50 participants were included in the data analysis, as preregistered. This final number of included participants was based on the same sample size approach explained above.

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 2 was identical in all respects to Experiment 1, apart from the fact that the order of elemental and compound training phases was reversed (see [Table 1](#) and [Fig 2A](#)).

**Data analysis.** The data analysis of Experiment 2 was the same as for Experiment 1. Also here, we adjusted the priors of the main effect of Expectation and Exposure and the prior of their interaction based on the posteriors of the previous experiment, i.e., Experiment 1, because the stimuli and procedure regarding effects of Expectation and Exposure were exactly the same. Note that specifying the priors in this way turns the results of Experiment 2 with respect to Expectation and Exposure effects into the combined evidence from Experiments 1 and 2. Crucially, the pattern of results from Experiment 2 was exactly the same when the priors for Expectation and Exposure were also centered at zero.

## Results

**Analysis of RT data in test phase.** First, we compared the reaction times of expected and unexpected trials in the companion condition to test whether repeated exposure to the pairs of the companion leading object A and trailing object X led to learning their temporal association (see [Table 4](#)). We observed faster reaction times in expected (477 ms) than unexpected (487 ms) trials ( $b = 8.90$ ,  $CI = [3.53, 14.27]$ , Cohen’s  $d_z = 0.35$ , see [Fig 2B](#)), indicating successful

**Table 4. Posterior fixed effects of the model of companion condition on reaction times in Experiment 2. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.**

Predictors	Estimate	Est. Error	CI (95%)
Intercept	483.97	11.08	462.31–505.62
Expectation	8.90	2.75	3.53–14.27
Exposure	-14.12	4.38	-22.81 -- -5.66
Expectation × Exposure	-4.63	3.93	-12.41–3.04

<https://doi.org/10.1371/journal.pone.0306797.t004>

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 tended to decrease as the exposure increased (i.e., extinction). We did not observe any interaction effect between Expectation and Exposure ( $b = 4.63$ ,  $CI = [-12.41, 3.04]$ ), indicating that learning did not show reliable extinction over time (expectation effect for run 1: 13 ms, run 2: 4 ms; see Fig 2C).

Next, we moved to our main question and tested for the presence of backward blocking (see Table 5 and Fig 2D). There was an interaction effect between expectation and condition ( $b = 9.45$ ,  $CI = [1.34, 17.63]$ , Cohen's  $d_z = 0.32$ , see Fig 2B). 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 were faster in expected (491 ms) than in unexpected (501 ms) trials of the control condition ( $b = 8.44$ ,  $CI = [3.60, 13.29]$ , Cohen's  $d_z = 0.39$ , see S3 Table in S1 File). On the other hand, there was no evidence that reaction times in expected (496 ms) and unexpected (496) trials were different from each other in the blocked condition ( $b = 2.69$ ,  $CI = [-2.08, 7.44]$ , Cohen's  $d_z = -0.01$ , see S4 Table in S1 File). This pattern of results supports the presence of backward blocking. There was no extinction in blocked and control conditions ( $b = -6.22$ ,  $CI = [-18.63, 6.38]$ ; expectation effect in blocked condition for run 1: 0 ms, run 2: 6 ms; expectation effect in control condition for run 1: 9 ms, run 2: 15 ms; see Fig 2C).

**Analyses of accuracy data in pair recognition test.** Participants showed slightly above chance-level performance in indicating whether the trailing object was likely or unlikely given the leading stimulus in the companion (proportion correct = 56%;  $b = 0.25$ ,  $CI = [0.10, 0.40]$ ), but not in blocked (proportion correct = 50%;  $b = 0$ ,  $CI = [-0.11, 0.11]$ ) and control (proportion correct = 53%;  $b = 0.09$ ,  $CI = [-0.04, 0.22]$ ) conditions (see Fig 2E).

## 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

**Table 5. Posterior fixed effects of the model of blocked and control conditions on reaction times in Experiment 2. Estimate, estimation error, lower/upper limit of 95% profile credible intervals.**

Predictors	Estimate	Est. Error	CI (95%)
Intercept	497.47	9.40	478.97–515.83
Expectation	6.81	1.98	2.94–10.65
Condition	2.86	2.54	-2.11–7.87
Exposure	-25.51	4.63	-34.68 -- -16.49
Expectation × Condition	9.45	4.16	1.34–17.63
Expectation × Exposure	3.53	2.83	-1.95–9.11
Condition × Exposure	0.54	3.17	-5.57–6.80
Expectation × Condition × Exposure	-6.22	6.29	-18.63–6.38

<https://doi.org/10.1371/journal.pone.0306797.t005>

results, for example, in more efficient behavioral processing [29,55,63–65] and more efficient neural processing [47,48,50–52] for predictable than unpredictable events. While the benefits of statistical learning are obvious, the mechanisms of statistical learning itself are less clear. In separate experiments, we used respectively forward and backward blocking [7,11] to examine whether cue competition and retrospective revaluation, which have been observed during reinforcement learning, also apply to statistical learning. We found backward blocking, suggesting a retrospective revaluation process during the incidental extraction of statistical regularities.

In Experiment 1, participants learned the associations for the blocked (B) stimulus condition; in fact, learning was even stronger for B stimuli compared to control (D) condition, a phenomenon which is sometimes referred to as ‘augmentation’ [36,66,67]. This pattern of results is opposite to the predictions of forward blocking and suggests the absence of forward blocking in statistical learning. One might argue that overall learning in the antedating and blocked conditions may not have been strong enough to generate forward blocking, given that the reduction in response speed was less than 20 ms. Such a small reaction time difference is, however, common in statistical learning [49,51,52] and similar in magnitude to RT benefits elicited by other cognitive factors such as probabilistic attentional cues [68].

We speculate that selective attention may provide a parsimonious explanation for the observed augmented learning in the blocked condition in Experiment 1. Pearce-Hall model is one of the traditional models explaining forward blocking based on attention and prediction error. According to the model, during the second training phase, attention is divided equally to both antedating (A) and blocked (B) leading objects, and the outcome of blocked leading object is less surprising because the antedating leading object (A) already predicts the outcome (X). Thus, the association between B and X cannot be formed. On the other hand, in learning, stimuli whose consequences are initially unexpected may attract more attention [69,70], leaving open the possibility of the associability of B and X. Similarly, 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 [71–73]. In other words, their attention is drawn to stimuli that offer maximum information gain. 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 object 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 (D) 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 with the trailing object and hence their individual predictive power was reduced [9]. In Experiment 2, we aimed to eliminate this potentially attentional effect by applying a backward blocking procedure. Given that the blocked leading object (B) was presented together with a companion leading object (A) in training phase 1, both the companion leading object (A) and blocked leading object (B) were equally familiar and salient in the first phase of the study. As a result, we removed the potentially confounding factors related to novelty and salience, and crucially our results suggested that backward blocking occurs in statistical learning.

One may wonder whether the present forward and backward blocking experiments provide contradictory results regarding the presence of blocking in statistical learning. Here, it is worth noting that it is more difficult to obtain backward blocking than forward blocking, because more criteria need to be met to observe backward blocking [14]. Forward blocking only requires a strong  $A \rightarrow X$  association, which is learned in the elemental training phase, to prevent learning the relationship between cue B and outcome X during the compound training

phase. On the other hand, in backward blocking, a strong  $A \rightarrow X$  association learned in the elemental training phase is not enough to observe blocking. In addition to that, the second important condition of backward blocking is that cue A needs to be associated with cue B in order to decrease the associability of cue B in its absence, which is supported by previous studies [74–76].

By showing backward blocking in Experiment 2, our results suggest the presence of retrospective revaluation in statistical learning. Such retrospective revaluation cannot be explained by the traditional Rescorla–Wagner model, which assumes that the relevant cue must be present in order to change the associative strength [9,12,13]. However, a number of alternative models are able to explain this observation. Van Hamme and Wasserman [14] proposed a modification of the traditional Rescorla–Wagner model, by allowing an update in the weight of an absent cue if the cue that is associated with the absent cue is present in that trial. Backward blocking can also be explained by a Bayesian generalization of the Rescorla–Wagner model, the Kalman filter [12,15]. In sum, our results in Experiment 2 can be explained by both the Van Hamme–Wasserman model and the Kalman filter, both of which claim that learning is based on prediction errors [12]. At the computational level, this implies that statistical learning may be error-driven. At the implementation level, it supports the view that statistical learning may follow the principles of predictive coding [77].

Critically, retrospective revaluation may be explained also by the probabilistic contrast model, which does not rely on prediction error [78,79]. This model simply calculates how frequently events occur during learning. That is, X appears after either AB or A during training phases, and the probability of X increases after A and in the absence of B. As a result, observers associate A with X. Given that the probabilistic contrast model disregards the order of elemental (i.e.,  $A \rightarrow X$ ) and compound (i.e.,  $AB \rightarrow X$ ) training phases [80], it explains both forward and backward blocking using the same approach. Although our backward blocking results can be explained by the probabilistic contrast model, the model fails to explain our forward blocking results. This supports the importance of the order of training phases in blocking [80].

Furthermore, it is important to acknowledge that blocking may not arise due to learning deficits, as explained by the models reviewed above, but instead may depend on the failure to express cue–outcome associations at test, as explained by the so-called comparator hypothesis [13,81]. In other words, retrospective revaluation would not occur because of the increase or decrease in the associative strength between cue and outcome, but rather because of a change in its expression at test. Although we observed backward blocking in statistical learning in Experiment 2, we do not know whether it is because of a learning deficit during training or because of a performance deficit observed at test. Thus, further studies are required to better understand the cause underlying backward blocking in statistical learning.

Crucially, learning regularities is usually thought to be incidental rather than intentional in statistical learning paradigms [27,47–50]. However, this can nevertheless result in the development of explicit knowledge of the regularities [47,49,64]. Indeed, testing for explicit knowledge is often used to assess the development of explicit knowledge [49,64]. In Experiment 1 and Experiment 2, we observed that people developed some minimal amount of explicit knowledge (on average 56% correct, with chance level of 50%) in the pair recognition test (i.e., how likely the trailing object was given the leading object). The temporal association between leading and trailing object was unknown to participants at the beginning of the experiment and participants were not instructed to learn these associations. Also, participants performed the categorization task at ceiling level (overall above 97%), suggesting that their categorization judgments were not affected by knowledge of the statistical structure between stimuli. Therefore, it appears likely that learning occurred incidentally, without strong explicit knowledge of the associations that were learnt. This is a clear difference between our studies and previous

‘classic’ blocking paradigms where learning occurs intentionally and in the presence of reinforcement [16,21–23,25,82].

Further, in the context of reinforcement learning, some highlight the key role of inferential reasoning for blocking to occur. Accordingly, learning associations between events does not depend on transitional probabilities but, instead, depends on the observers’ belief about the nature of the relationship between cue and outcome [80,83,84]. Specifically, the intentional evaluation of causal associations between cues and outcomes (e.g., cue A is the cause of outcome X) appears necessary for forward blocking [4,37,38] and backward blocking [80,83,84]. As a result, there is so far evidence that conscious inferential reasoning contributes to backward blocking. To the best of our knowledge, the present study is the first to examine backward blocking in incidental statistical learning. In our experiment, participants were not instructed about any possible relationship between leading and trailing objects, and they learned the associative relationship incidentally. Therefore, our finding supports that conscious inferential reasoning is not required for backward blocking to occur; instead, retrospective reevaluation can happen during incidental statistical learning.

To sum up, while we did not find forward blocking, our results are compatible with the presence of backward blocking in statistical learning, a form of learning that develops incidentally and in the absence of rewarding outcomes or feedback. Our results are compatible with the Van Hamme–Wasserman model and Kalman filter, and thus support the idea that statistical learning may be error-driven, similar to reinforcement learning (though see the comparator hypothesis). Most importantly, our results suggest a retrospective reevaluation process in statistical learning and thus support a functional similarity between statistical learning and reinforcement learning.

## Supporting information

**S1 File. Supplementary information 1.**  
(DOCX)

**S2 File. Supplementary information 2.**  
(DOCX)

**S1 Fig. Experimental procedure and results of Experiment S1.** 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) Statistical regularities depicted as image transition matrix with stimuli pairs in training phase 1 and training phase 2. Ls represent leading stimuli, and Ts represent trailing stimuli. (c) Statistical regularities depicted as image transition matrix with stimuli pairs in test phase. Green cells represent expected pairs, and red cells represent unexpected pairs. (d) 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. (e) 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 test phase. Extinction was not different between conditions. (f) 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. (g) Across participants' proportion correct responses in pair recognition test. Participants showed slightly above chance-level performance in indicating whether the trailing object was likely or unlikely given the leading stimulus in all conditions.

(TIF)

**S2 Fig. Results of Experiment S1 as a function of Stimulus Type.** (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 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.

(TIF)

**S3 Fig. Experimental procedure and results of Experiment S2.** (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. (e) Across participants' proportion correct responses 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 all conditions.

(TIF)

## Author Contributions

**Conceptualization:** Ambra Ferrari, Floris P. de Lange.

**Data curation:** İlayda Nazlı.

**Formal analysis:** İlayda Nazlı, Christoph Huber-Huber.

**Methodology:** İlayda Nazlı, Ambra Ferrari, Floris P. de Lange.

**Project administration:** İlayda Nazlı.

**Supervision:** Ambra Ferrari, Christoph Huber-Huber, Floris P. de Lange.

**Writing – original draft:** İlayda Nazlı, Ambra Ferrari, Christoph Huber-Huber.

**Writing – review & editing:** Ambra Ferrari, Christoph Huber-Huber, Floris P. de Lange.

## References

1. Gershman SJ. Context-dependent learning and causal structure. *Psychon Bull Rev.* 2017; 24: 557–565. <https://doi.org/10.3758/s13423-016-1110-x> PMID: 27418259
2. Pavlov IP. *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex.* Oxford, England: Oxford Univ. Press; 1927. pp. xv, 430.
3. Boddez Y, Haesen K, Baeyens F, Beckers T. Selectivity in associative learning: a cognitive stage framework for blocking and cue competition phenomena. *Front Psychol.* 2014; 5: 1305. <https://doi.org/10.3389/fpsyg.2014.01305> PMID: 25429280
4. De Houwer J, Vandorpe S, Beckers T. Evidence for the role of higher order reasoning processes in cue competition and other learning phenomena. *Learn Behav.* 2005; 33: 239–249. <https://doi.org/10.3758/bf03196066> PMID: 16075842
5. Luque D, Vadillo MA, Gutiérrez-Cobo MJ, Le Pelley ME. The blocking effect in associative learning involves learned biases in rapid attentional capture. *Q J Exp Psychol.* 2018; 71: 522–544. <https://doi.org/10.1080/17470218.2016.1262435> PMID: 27874321
6. Schmidt JR, De Houwer J. Cue Competition and Incidental Learning: No Blocking or Overshadowing in the Colour-Word Contingency Learning Procedure Without Instructions to Learn. Zwaan R, Grange J, editors. *Psychol Collabra.* 2019; 5: 15. <https://doi.org/10.1525/collabra.236>
7. Kamin LJ. Predictability, surprise, attention, and conditioning. in campbell ba & church rm (eds.), *Punishment and aversive behavior* (pp. 279–296). N Y Appleton-Century-Crofts. 1969.
8. Hebb DO. *The Organization of Behaviour* (Hoboken, NJ. Wiley; 1949.
9. Rescorla R, Wagner A. A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. *Class Cond Curr Res Theory.* 1972.
10. Spicer S, Wills A, Jones P, Mitchell C, Dome L. Representing uncertainty in the Rescorla-Wagner model: Blocking, the redundancy effect, and outcome base rate. *Open J Exp Psychol Neurosci.* 2021; 14–21. <https://doi.org/10.46221/ojepn.2021.6623>
11. Shanks DR. Forward and Backward Blocking in Human Contingency Judgement. *Q J Exp Psychol Sect B.* 1985; 37: 1–21. <https://doi.org/10.1080/14640748508402082>
12. Kruschke JK. Bayesian approaches to associative learning: From passive to active learning. *Learn Behav.* 2008; 36: 210–226. <https://doi.org/10.3758/lb.36.3.210> PMID: 18683466
13. Miller RR, Witnauer JE. Retrospective Revaluation: The Phenomenon and Its Theoretical Implications. *Behav Processes.* 2016; 123: 15–25. <https://doi.org/10.1016/j.beproc.2015.09.001> PMID: 26342855
14. Van Hamme LJ, Wasserman EA. Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learn Motiv.* 1994; 25: 127–151. <https://doi.org/10.1006/lmot.1994.1008>
15. Gershman SJ. A Unifying Probabilistic View of Associative Learning. Diedrichsen J, editor. *PLOS Comput Biol.* 2015; 11: e1004567. <https://doi.org/10.1371/journal.pcbi.1004567> PMID: 26535896
16. Aggarwal M, Akamine Y, Liu AW, Wickens JR. The nucleus accumbens and inhibition in the ventral tegmental area play a causal role in the Kamin blocking effect. *Eur J Neurosci.* 2020; 52: 3087–3109. <https://doi.org/10.1111/ejn.14732> PMID: 32250479
17. Aggarwal M, Wickens JR. The Kamin Blocking Effect in Sign and Goal Trackers. *bioRxiv.* 2020.
18. Sharpe MJ, Chang CY, Liu MA, Batchelor HM, Mueller LE, Jones JL, et al. Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nat Neurosci.* 2017; 20: 735–742. <https://doi.org/10.1038/nn.4538> PMID: 28368385

19. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci*. 2013; 16: 966–973. <https://doi.org/10.1038/nn.3413> PMID: 23708143
20. Blanco F, Baeyens F, Beckers T. Blocking in human causal learning is affected by outcome assumptions manipulated through causal structure. *Learn Behav*. 2014; 42: 185–199. <https://doi.org/10.3758/s13420-014-0137-y> PMID: 24737045
21. Kruschke JK, Blair NJ. Blocking and backward blocking involve learned inattention. *Psychon Bull Rev*. 2000; 7: 636–645. <https://doi.org/10.3758/bf03213001> PMID: 11206204
22. Le Pelley ME, Oakeshott SM, McLaren IP. Blocking and unblocking in human causal learning. *J Exp Psychol Anim Behav Process*. 2005; 31: 56. <https://doi.org/10.1037/0097-7403.31.1.56> PMID: 15656727
23. Le Pelley ME, Beesley T, Suret MB. Blocking of human causal learning involves learned changes in stimulus processing. *Q J Exp Psychol*. 2007; 60: 1468–1476. <https://doi.org/10.1080/17470210701515645> PMID: 17853191
24. Mitchell CJ, Killeddar A, Lovibond PF. Inference-based retrospective reevaluation in human causal judgments requires knowledge of within-compound relationships. *J Exp Psychol Anim Behav Process*. 2005; 31: 418–424. <https://doi.org/10.1037/0097-7403.31.4.418> PMID: 16248728
25. Mitchell CJ, Lovibond PF, Minard E, Lavis Y. Forward blocking in human learning sometimes reflects the failure to encode a cue–outcome relationship. *Q J Exp Psychol*. 2006; 59: 830–844. <https://doi.org/10.1080/17470210500242847> PMID: 16608749
26. Gershman SJ, Daw ND. Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annu Rev Psychol*. 2017; 68: 101–128. <https://doi.org/10.1146/annurev-psych-122414-033625> PMID: 27618944
27. Batterink LJ, Paller KA, Reber PJ. Understanding the Neural Bases of Implicit and Statistical Learning. *Top Cogn Sci*. 2019; 11: 482–503. <https://doi.org/10.1111/tops.12420> PMID: 30942536
28. Frost R, Armstrong BC, Christiansen MH. Statistical learning research: A critical review and possible new directions. *Psychol Bull*. 2019; 145: 1128–1153. <https://doi.org/10.1037/bul0000210> PMID: 31580089
29. Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science*. 1996; 274: 1926–1928. <https://doi.org/10.1126/science.274.5294.1926> PMID: 8943209
30. Sherman BE, Graves KN, Turk-Browne NB. The prevalence and importance of statistical learning in human cognition and behavior. *Curr Opin Behav Sci*. 2020; 32: 15–20. <https://doi.org/10.1016/j.cobeha.2020.01.015> PMID: 32258249
31. Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM. Implicit perceptual anticipation triggered by statistical learning. *J Neurosci Off J Soc Neurosci*. 2010; 30: 11177–11187. <https://doi.org/10.1523/JNEUROSCI.0858-10.2010> PMID: 20720125
32. Keiflin R, Pribut HJ, Shah NB, Janak PH. Ventral Tegmental Dopamine Neurons Participate in Reward Identity Predictions. *Curr Biol CB*. 2019; 29: 93–103.e3. <https://doi.org/10.1016/j.cub.2018.11.050> PMID: 30581025
33. Klein-Flügge MC, Wittmann MK, Shpektor A, Jensen DEA, Rushworth MFS. Multiple associative structures created by reinforcement and incidental statistical learning mechanisms. *Nat Commun*. 2019; 10: 4835. <https://doi.org/10.1038/s41467-019-12557-z> PMID: 31645545
34. McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. *Neuron*. 2003; 38: 339–346. [https://doi.org/10.1016/s0896-6273\(03\)00154-5](https://doi.org/10.1016/s0896-6273(03)00154-5) PMID: 12718866
35. O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *science*. 2004; 304: 452–454. <https://doi.org/10.1126/science.1094285> PMID: 15087550
36. Beesley T, Shanks DR. Investigating cue competition in contextual cuing of visual search. *J Exp Psychol Learn Mem Cogn*. 2012; 38: 709–725. <https://doi.org/10.1037/a0024885> PMID: 21895393
37. De Houwer J, Beckers T. Secondary task difficulty modulates forward blocking in human contingency learning. *Q J Exp Psychol Sect B*. 2003; 56: 345–357. <https://doi.org/10.1080/02724990244000296> PMID: 14578079
38. Vandorpe S, De Houwer J, Beckers T. Further evidence for the role of inferential reasoning in forward blocking. *Mem Cognit*. 2005; 33: 1047–1056. <https://doi.org/10.3758/bf03193212> PMID: 16496725
39. Morís J, Cobos PL, Luque D, López FJ. Associative repetition priming as a measure of human contingency learning: evidence of forward and backward blocking. *J Exp Psychol Gen*. 2014; 143: 77. <https://doi.org/10.1037/a0030919> PMID: 23230993

40. Griffiths TL, Sobel DM, Tenenbaum JB, Gopnik A. Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cogn Sci*. 2011; 35: 1407–1455. <https://doi.org/10.1111/j.1551-6709.2011.01203.x> PMID: 21972897
41. McCormack T, Butterfill S, Hoerl C, Burns P. Cue competition effects and young children's causal and counterfactual inferences. *Dev Psychol*. 2009; 45: 1563–1575. <https://doi.org/10.1037/a0017408> PMID: 19899914
42. McCormack T, Simms V, McGourty J, Beckers T. Blocking in children's causal learning depends on working memory and reasoning abilities. *J Exp Child Psychol*. 2013; 115: 562–569. <https://doi.org/10.1016/j.jecp.2012.11.016> PMID: 23499322
43. Sobel DM, Tenenbaum JB, Gopnik A. Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cogn Sci*. 2004; 28: 303–333. [https://doi.org/10.1207/s15516709cog2803\\_1](https://doi.org/10.1207/s15516709cog2803_1)
44. Sobel DM, Kirkham NZ. Blickets and babies: the development of causal reasoning in toddlers and infants. *Dev Psychol*. 2006; 42: 1103–1115. <https://doi.org/10.1037/0012-1649.42.6.1103> PMID: 17087545
45. Sobel DM, Kirkham NZ. Bayes nets and babies: infants' developing statistical reasoning abilities and their representation of causal knowledge. *Dev Sci*. 2007; 10: 298–306. <https://doi.org/10.1111/j.1467-7687.2007.00589.x> PMID: 17444971
46. Mackintosh N. A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychol Rev*. 1975; 82: 276–298. <https://doi.org/10.1037/h0076778>
47. Batterink LJ, Paller KA. Online neural monitoring of statistical learning. *Cortex J Devoted Study Nerv Syst Behav*. 2017; 90: 31–45. <https://doi.org/10.1016/j.cortex.2017.02.004> PMID: 28324696
48. Henin S, Turk-Browne NB, Friedman D, Liu A, Dugan P, Flinker A, et al. Learning hierarchical sequence representations across human cortex and hippocampus. *Sci Adv*. 2021; 7: eabc4530. <https://doi.org/10.1126/sciadv.abc4530> PMID: 33608265
49. Turk-Browne NB, Jungé J, Scholl BJ. The automaticity of visual statistical learning. *J Exp Psychol Gen*. 2005; 134: 552–564. <https://doi.org/10.1037/0096-3445.134.4.552> PMID: 16316291
50. Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK. Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness. *J Cogn Neurosci*. 2009; 21: 1934–1945. <https://doi.org/10.1162/jocn.2009.21131> PMID: 18823241
51. Richter D, Ekman M, de Lange FP. Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *J Neurosci*. 2018; 38: 7452–7461. <https://doi.org/10.1523/JNEUROSCI.3421-17.2018> PMID: 30030402
52. Richter D, de Lange FP. Statistical learning attenuates visual activity only for attended stimuli. *eLife*. 2019; 8: e47869. <https://doi.org/10.7554/eLife.47869> PMID: 31442202
53. Kaposvari P, Kumar S, Vogels R. Statistical learning signals in macaque inferior temporal cortex. *Cereb Cortex*. 2018; 28: 250–266. <https://doi.org/10.1093/cercor/bhw374> PMID: 27909007
54. Meyer T, Olson CR. Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc Natl Acad Sci U S A*. 2011; 108: 19401–19406. <https://doi.org/10.1073/pnas.1112895108> PMID: 22084090
55. Hunt RH, Aslin RN. Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *J Exp Psychol Gen*. 2001; 130: 658–680. <https://doi.org/10.1037/0096-3445.130.4.658> PMID: 11757874
56. Anwyl-Irvine AL, Massonnié J, Flitton A, Kirkham N, Evershed JK. Gorilla in our midst: An online behavioral experiment builder. *Behav Res Methods*. 2020; 52: 388–407. <https://doi.org/10.3758/s13428-019-01237-x> PMID: 31016684
57. Ferrari A, Richter D, de Lange FP. Updating Contextual Sensory Expectations for Adaptive Behavior. *J Neurosci Off J Soc Neurosci*. 2022; 42: 8855–8869. <https://doi.org/10.1523/JNEUROSCI.1107-22.2022> PMID: 36280262
58. Al-Salom P, Miller CJ. The Problem with Online Data Collection: Predicting Invalid Responding in Undergraduate Samples. *Curr Psychol*. 2019; 38: 1258–1264. <https://doi.org/10.1007/s12144-017-9674-9>
59. Brühlmann F, Petralito S, Aeschbach LF, Opwis K. The quality of data collected online: An investigation of careless responding in a crowdsourced sample. *Methods Psychol*. 2020; 2: 100022. <https://doi.org/10.1016/j.metip.2020.100022>
60. Brady TF, Konkle T, Alvarez GA, Oliva A. Visual long-term memory has a massive storage capacity for object details. *Proc Natl Acad Sci*. 2008; 105: 14325–14329. <https://doi.org/10.1073/pnas.0803390105> PMID: 18787113
61. McLaren IPL, Mackintosh NJ. An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Anim Learn Behav*. 2000; 28: 211–246. <https://doi.org/10.3758/BF03200258>

62. Bürkner P-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *J Stat Softw.* 2017; 80: 1–28. <https://doi.org/10.18637/jss.v080.i01>
63. Fiser J, Aslin RN. Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol Sci.* 2001; 12: 499–504. <https://doi.org/10.1111/1467-9280.00392> PMID: 11760138
64. Fiser J, Aslin RN. Statistical learning of higher-order temporal structure from visual shape sequences. *J Exp Psychol Learn Mem Cogn.* 2002; 28: 458–467. <https://doi.org/10.1037//0278-7393.28.3.458> PMID: 12018498
65. Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tone sequences by human infants and adults. *Cognition.* 1999; 70: 27–52. [https://doi.org/10.1016/s0010-0277\(98\)00075-4](https://doi.org/10.1016/s0010-0277(98)00075-4) PMID: 10193055
66. Batson JD, Batsell Jr. WR. Augmentation, not blocking, in an A+/AX+ flavor-conditioning procedure. *Psychon Bull Rev.* 2000; 7: 466–471. <https://doi.org/10.3758/BF03214358> PMID: 11082852
67. Vadillo MA, Matute H. Augmentation in contingency learning under time pressure. *Br J Psychol.* 2010; 101: 579–589. <https://doi.org/10.1348/000712609X477566> PMID: 19917150
68. Posner MI. Orienting of attention. *Q J Exp Psychol.* 1980; 32: 3–25. <https://doi.org/10.1080/00335558008248231> PMID: 7367577
69. Holland PC, Schiffino FL. Mini-Review: Prediction errors, attention and associative learning. *Neurobiol Learn Mem.* 2016; 131: 207–215. <https://doi.org/10.1016/j.nlm.2016.02.014> PMID: 26948122
70. Pearce JM, Hall G. A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev.* 1980; 87: 532–552. <https://doi.org/10.1037/0033-295X.87.6.532> PMID: 7443916
71. Gottlieb J, Oudeyer P-Y, Lopes M, Baranes A. Information seeking, curiosity and attention: computational and neural mechanisms. *Trends Cogn Sci.* 2013; 17: 585–593. <https://doi.org/10.1016/j.tics.2013.09.001> PMID: 24126129
72. Kidd C, Piantadosi ST, Aslin RN. The Goldilocks effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS One.* 2012; 7: e36399. <https://doi.org/10.1371/journal.pone.0036399> PMID: 22649492
73. Poli F, Serino G, Mars RB, Hunnius S. Infants tailor their attention to maximize learning. *Sci Adv.* 2020; 6: eabb5053. <https://doi.org/10.1126/sciadv.abb5053> PMID: 32967830
74. Luque D, Flores A, Vadillo MA. Revisiting the role of within-compound associations in cue-interaction phenomena. *Learn Behav.* 2013; 41: 61–76. <https://doi.org/10.3758/s13420-012-0085-3> PMID: 22753000
75. Melchers KG, Lachnit H, Shanks DR. Within-compound associations in retrospective revaluation and in direct learning: a challenge for comparator theory. *Q J Exp Psychol B.* 2004; 57: 25–53. <https://doi.org/10.1080/02724990344000042> PMID: 14690848
76. Melchers K, Lachnit H, Shanks DR. The comparator theory fails to account for the selective role of within-compound associations in cue-selection effects. *Exp Psychol.* 2006; 53: 316–320. <https://doi.org/10.1027/1618-3169.53.4.316> PMID: 17176664
77. Hasson U. The neurobiology of uncertainty: implications for statistical learning. *Philos Trans R Soc Lond B Biol Sci.* 2017; 372: 20160048. <https://doi.org/10.1098/rstb.2016.0048> PMID: 27872367
78. Cheng P. From Covariation to Causation: A Causal Power Theory. *Psychol Rev.* 1997; 104: 367–405. <https://doi.org/10.1037//0033-295X.104.2.367>
79. Cheng PW, Novick LR. Covariation in natural causal induction. *Psychol Rev.* 1992; 99: 365–382. <https://doi.org/10.1037/0033-295x.99.2.365> PMID: 1594730
80. De Houwer J, Beckers T. A review of recent developments in research and theories on human contingency learning. *Q J Exp Psychol B.* 2002; 55: 289–310. <https://doi.org/10.1080/02724990244000034> PMID: 12350283
81. Miller RR, Matzel LD. The Comparator Hypothesis: A Response Rule for The Expression of Associations. In: Bower GH, editor. *Psychology of Learning and Motivation.* Academic Press; 1988. pp. 51–92. [https://doi.org/10.1016/S0079-7421\(08\)60038-9](https://doi.org/10.1016/S0079-7421(08)60038-9)
82. De Houwer J, Beckers T, Glautier S. Outcome and cue properties modulate blocking. *Q J Exp Psychol A.* 2002; 55: 965–985. <https://doi.org/10.1080/02724980143000578> PMID: 12188522
83. Waldmann MR. Competition among causes but not effects in predictive and diagnostic learning. *J Exp Psychol Learn Mem Cogn.* 2000; 26: 53–76. <https://doi.org/10.1037//0278-7393.26.1.53> PMID: 10682290
84. Waldmann MR, Holyoak KJ. Predictive and diagnostic learning within causal models: asymmetries in cue competition. *J Exp Psychol Gen.* 1992; 121: 222–236. <https://doi.org/10.1037//0096-3445.121.2.222> PMID: 1534834