Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Exp Psychol Gen. Author manuscript; available in PMC 2023 August 01.

Published in final edited form as:

J Exp Psychol Gen. 2022 August ; 151(8): 1772–1792. doi:10.1037/xge0001166.

Benefitting from trial spacing without the cost of prolonged training: Frequency, not duration, of trials with absent stimuli enhances perceived contingency

Santiago Castiello[1], **Ralph R. Miller**[2], **James E. Witnauer**[3], **Doriann M. Alcaide**[2], **Ethan Fung**[2], **Riddhi J. Pitliya**[1], **Dyedra K. C. Morrissey**[1], **Robin A. Murphy**[1]

[1]Department of Experimental Psychology, University of Oxford,

[2]Department of Psychology, State University of New York (SUNY) at Binghamton

[3]Department of Psychology, State University of New York (SUNY) at Brockport

Abstract

The statistical relation between two events influences the perception of how one event relates to the presence or absence of another. Interestingly, the simultaneous absence of both events, just like their mutual occurrence, is relevant for describing their contingency. In three experiments, we explored the relevance of co-absent events by varying the duration and frequency of trials without stimuli. We used a rapid trial streaming procedure, and found that the perceived association between events is enhanced with increasing frequency of co-absent events unlike the duration of co-absent events which had little effect. These findings suggest ways in which the benefits of trial spacing, during which both events are absent, could be obtained without increasing total training time. Centrally, this can be done by frequent repeating of shortened co-absent events, each marked by a trial contextual cue. We discuss four potential accounts of how co-absent experience might be processed contributing to this effect: *i*) contingency sensitivity, *ii*) testing effect, *iii*) reduced associative interference by the context, and *iv*) reduced encoding interference.

Keywords

trial spacing; distributed practice; intertrial interval; contingency learning; rapid streaming contingency

Introduction

Learning the contingency between two stimuli, or a stimulus and a response, is a fundamental form of learning (Baker, Murphy, & Vallée-Tourangeau, 1996), involving perceiving, remembering, and learning the correlation between events. Such statistical relations can be positive, in which people learn that the presence of one event predicts another (bacon is a useful predictor of eggs as they are positively correlated in some

Open access. The data and code are available online in: https://github.com/santiagocdo/cellD_paper.

Corresponding author: Santiago Castiello, M.Sc., St Anne's College, University of Oxford, 56 Woodstock Rd, OX2 6HS, Oxford, United Kingdom, santiago.castiello@psy.ox.ac.uk.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

people's breakfast experience), but predictive relations can be negative, where the presence of one event signals the absence of another (toothpaste and wine, blessedly, tend to be experiences that are negatively correlated). While the simultaneous presence of events is clearly relevant for learning a positive relation, and the occurrence of either event alone is relevant for negative correlations, the mutual co-absence of both events is less intuitively relevant for either type of correlation. However, the absence of events is not the absence of information. For example, in determining the cause of stress-related headaches, the absence of stress and headaches might be informative for inferring the relation between stress and headaches (Murphy, Byrom, & Msetfi, 2017). But, how do absences become relevant and inform particular predictive relations? Moreover, tasks that assess human reasoning find that information about absence is difficult to process and use, suggesting an underlying bias against processing these type of experiences. Here, we investigate how manipulating the frequency and duration of co-absent events impacts judgements of perceived association between pairs of stimuli, in other words, how can experience with "nothing" impact judgements about "something".

To consider how contingent relations are formed, it is worth considering the two well-established empirical factors for establishing the perception of a relationship between two stimuli (e.g., S1 and S2) which are: (1) the temporal and spatial *contiguity*, and (2) the statistical dependency or *contingency* which defines a correlative relationship. The first factor has been identified to be *necessary*, if not *sufficient*, for associative learning and memory co-activation (Miller & Barnet, 1993; for a review, see Savastano & Miller, 1998); a stimulus will become associated with other stimuli and be able to retrieve that stimulus if presented in the same moment in time and space (Molet, Jozefowiez & Miller, 2010). This idea has been formalised in the Sometimes-Competing Retrieval model (SOCR; Stout & Miller, 2007). The second factor is contingency, a descriptor of the correlation between the two events. When two stimuli are positively contingent, there is a greater likelihood of observing S1 and S2 together (i.e., conjunction or co-occurrence) and absent together (i.e., co-absence). Seeing only one stimulus without the other characterizes negative contingencies (i.e., disjunction). Contingency has been formalised by Δp (delta p; Allan, 1980), the one-way relation between two binary dichotomous variables. Δp has been widely used as an empirical model of learning and as a proxy for causal sensitivity (Baker, Murphy, Mehta, & Baetu, 2005; Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993; Dickinson, & Shanks, 1987; Dickinson, Shanks, & Evenden, 1984; Vallée-Tourangeau, Murphy, & Baker, 2005).

The Δp contingency between two stimuli, as shown in Figure 1, is defined as the difference in the conditional probability of the presence of S2 (the shoe) given the presence of S1 (the geometric shape) [p(S2|S1)], and the presence of S2 given the absence (~) of S1 [p(S2|~S1)]. When there is no relation between the two events, that is, $\Delta p = 0$, S1 provides no information about the occurrence of S2. However, if $\Delta p > 0$, the presentation of S1 signals an increase in the likelihood of S2 relative to the baseline occurrence of S2, and if $\Delta p < 0$, the presence of S1 indicates a decrease in the likelihood of S2 relative to the baseline occurrence of S2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

As illustrated in Figure 1, the number of co-occurrences of S1 and S2 (A trials), the number of occurrences of S1 alone (B trials), the number of occurrences of S2 alone (C trials), and the number of co-absences of S1 and S2 (D trials) combine to provide a metric of contingency. Perhaps not surprisingly, increasing the number of times the two stimuli are paired together (A trials), everything else being equal, increases judgments of a positive relation (e.g., Murphy et al., 2011). Perhaps less intuitively, according to Δp, increasing the number of times that both stimuli are absent together (D trials) should also increase their contingency and by extension perceived association. In contrast, presenting either event alone (B or C trials), is consistent with a negative contingency, and a decrease in the perceived association.

Previous experiments have shown that the four types of events are not equally effective in changing people's judgments (e.g., Murphy, Witnauer, Castiello, Tsvetkov, Li, Alcaide, & Miller, 2021). Particularly, a form of D trial neglect has been found, such that participants tend to give weight to the events in the following manner: A > B ≈ C > D (Wasserman, Elek, Chatlosh, & Baker, 1993). This selectivity for positive evidence has also been found in other behavioural and cognitive tasks in which participants show sensitivity to the relevance of all types of information but regularly demonstrate a bias towards confirmatory evidence and away from so called negative evidence (Wason, 1968).

In behavioural tasks, the absence of a stimulus is less effective for learning than its presence in humans (Mata, Garcia-Marques, Ferreira, & Mendonça, 2015), birds (Nieder, Wagener, & Rinnert, 2020), and rats (Sainsbury, 1971). The feature-positive effect (Jenkins & Sainsbury, 1969) demonstrates that learning about the presence of a stimulus (X) for an outcome (+) against a background (Y; i.e., XY+, Y−) is easier than learning that the absence of a stimulus is a signal for an outcome (XY−, Y+; with humans, Newman, Wolff & Hearst, 1980; with rats and pigeons, Bouton & Nelson, 1998; Young & Pearce, 1984, respectively).

In other domains, cognitions about evidence show similar biases. For instance, Wason (1968; also see Johnson-Laird & Wason, 1970) demonstrated a reliance on confirmatory evidence in testing rules (e.g., if *p* then *q*), in which participants sought evidence to confirm a hypothesis and were less likely to seek disconfirmatory evidence (if ~*p* then ~*q*; Sperber, Cara, & Girotto, 2000 although see Fiddick, Cosmides, & Tooby, 2000). Interestingly, paranormal believers are less likely to use absent information than paranormal nonbelievers (Blanco, Barberia, & Matute, 2015). Furthermore, extreme expressions of such biases against disconfirmatory evidence may reflect underlying mental health related or clinical conditions such as schizophrenia (Juarez-Ramos, Rubio, Delpero, Mioni, Stablum, & Gomez-Milan, 2014; Moritz, & Woodward, 2006).

In a previous set of experiments using a contingency learning task, in addition to finding a difference in the weighting of the four types of events of a contingency situation, we demonstrated that changes in perceived association were determined by the frequencies of events but not their durations (Murphy et al., 2021). In five experiments Murphy et al. (2021), using the rapid streaming procedure of Crump, Hannah, Allan, and Hord (2007; see also, Maia, Lefèvre, & Jozefowiez, 2018), participants received presentations of two stimuli that were objectively correlated (positively or negatively) or uncorrelated. In all experiments,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

increasing the frequency but not the duration of A trials increased ratings of a positive association, and increasing the frequency but not duration of B or C trials decreased ratings of a positive association. Increasing the number of trials while simultaneously decreasing their duration had the same effect on learning as increasing the number of trials without reducing trial duration. However, neither increasing the frequency nor the duration of D trials had any discernible effect on the rating of an association between the two events.

The present investigation, therefore, attempted to understand the role of mutual absent stimuli in contingency learning by manipulating the frequency and duration of only the trials where 'nothing', but the experimental trial context, is presented. Our previous finding was surprising because there is strong evidence that increasing the duration *between* trials in which no events are present, that is, the intertrial intervals (ITIs), increases the strength of a learned association (e.g., Balsam, Drew, & Gallistel, 2010; Escobar, Arcediano, & Miller, 2002; Harris & Bouton, 2020; Harris, Patterson, & Gharaei, 2015; Msetfi, Murphy & Simpson, 2007; Msetfi, Murphy, Simpson, & Kornbrot, 2005; Urcelay, Wheeler, & Miller, 2009). This concept of spacing of training trials is a well-researched phenomenon in psychology, in which learning improves by increasing the intervals between pairings of associated stimuli (e.g., Gibbon, Church, & Meck, 1984), leading to our expectation that increasing either the frequency or duration of D trials would enhance the perceived association.

It is also possible that the failure to detect any effect of frequency or duration of D trials (Murphy et al., 2021) may have been due to the rapid presentation procedure itself and its impact on trial salience or relevance. In the previous experiments, the temporal periods between events were of the order of hundreds of milliseconds. The effect of absent events may differ on this time scale. Presenting a participant with intervals in which no stimuli occur may provide a brief period to reset attention (attentional blink; Ophir, Sherman, & Lamy, 2018). Presenting stimulus-absent periods has been claimed to attenuate subsequent stimulus detection (Nieuwenstein, Potter, Theeuwes, 2009). In addition to the temporal differences, the participant's expectations of the relevance of an absent stimulus based on the experimental instructions may also play a role. Mata et al. (2015) found that people learn to use the co-absent information when they are instructed with respect to a specific goal, for example, when participants are asked to explain associations to someone else.

The present experiments aimed to test how absent information affects judgments of association between stimuli. The potential effect of increasing the frequency or duration of the co-absent information can be approached from at least the following four theoretical perspectives. *i*) Contingency Sensitivity: D trials are perceived along with the other events and combined with them provide a metric of association, but with the metric biased by the lower weight assigned to the low salience D trials in comparison to the other trial types (Wasserman et al., 1993). In this view, Δp could either be a process or descriptive model (Cheng, 1997; White, 2005). *ii*) Testing Effect: Performance during a test of association is enhanced if information acquired on earlier A, B, and C training trials is retrieved on later A, B, and C training trials, respectively, provided that the immediately preceding trial of that type is no longer in short-term memory (i.e., retrieval practice; Roediger & Butler, 2011); from this perspective, D trials clear STM of task-relevant trials, so, on next relevant trial,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

information is retrieved from LTM. *iii*) Reduced Associative Interference by the Context: In addition to the association presumed to form between the two target stimuli, it is assumed that the trial context becomes associated with two nominal stimuli on A, B and C trials. Theories of association assume that on an A trial for instance, both S1 and S2 become associated, but they also associate with the trial context. Similarly, on B and C trials, S1 and S2 independently associate with context (in Figure 1, the trial markers). In this manner, the S1–S2 association being acquired may be subject to interference from the associations that each element has with the context. Theories of association such as the Rescorla-Wagner model (Rescorla & Wagner, 1972) assume that cues compete for association (e.g., Vallee-Tourangeau et al., 1998). Increasing the frequency or duration of D trials would be expected to extinguish the context's associations with S1 and S2, freeing them up to associate with each other (e.g., Bouton, 1993). *iv)* Reduced Encoding Interference: Other theories assume that memory for the pairing between any two events takes time to process, and increasing the spacing between each A trial permits more time to process each new trial (Engle, 2002; Waugh & Norman, 1965). From this perspective, D trials provide a relatively empty period that acts as a temporary protector of information active in short-term memory (STM) which boosts later long-term memory (LTM) encoding or retrieval of the association (Aust, Haaf, & Stahl, 2019). Notably, these four mechanisms by which the increase of context alone presentation might facilitate learning and memory are not necessarily mutually exclusive.

Experiment 1

The objective of Experiment 1 was to assess whether frequency and duration of D trials are involved in perception of association. We manipulated duration and frequency of D trials (co-absent stimuli), and introduced a retention interval between stimulus presentation and testing. The retention manipulation was included because there is some evidence that immediate testing following training is not as sensitive to the benefit of trial spacing as testing following a retention interval (e.g., Smith & Kimball, 2010). Additionally, research has found that retention intervals, just like context changes, reduce retroactive interference (e.g., Rosas, Vila, Lugo, & López, 2001). Therefore, by introducing an interval between training and testing (the retention interval), we may reduce interference among the various trial types in the experiment and perhaps induce a D trial effect. However, it is possible that a retention interval produces an increase in subjective similarity between events; hence, an overall decrease in sensitivity to discriminable contingencies is also possible (King, Jones, Pearlman, Tishman, & Felix, 2002). Previous work using a predictive judgment task with partial reinforcement (only A and B trials) with a longer retention interval (48 hrs) found evidence supporting a primacy effect over a recency effect when there was a retrieval interval, but a recency over primacy effect when there was no retention interval (Stout, Amundson, & Miller, 2005). During the retention interval, we introduced a secondary irrelevant cognitive task to reduce the likelihood that the retention interval was not treated as relevant to the contingency task.

Methods

Participants—Forty participants, 26 females and 14 males (age between 18 and 55; M = 21.5, SD = 7.76) were recruited and offered course credit or £10.00 for their participation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

This research was approved by the Medical Science Interdivisional Research Ethics Committee of University of Oxford; reference number: R60840/RE001. Our participant exclusion criteria were participants who provided the same rating on all conditions (i.e., were not sensitive to the manipulation) or those who did not finish their session for whatever reason; based on these criteria, no participants were excluded.

Design—We used a fully within-subjects design which consisted of one practice condition followed by twelve experimental conditions in random order. Six of the experimental conditions had delayed ratings and six had immediate ratings. Each condition consisted of a sequence of trials in which each trial was one of the four trial types (A, B, C, or D; see Figure 1 and Table 1) involved in creating the overall Δp contingency between two stimuli (S1 and S2). An example of a trial sequence is shown in Figure 1.

To manipulate the retention interval, the judgment ratings were made either immediately following training or after a 3-minute retention interval during which participants solved a word search puzzle that was novel for each condition (no participant finished the puzzle within 3 minutes). We independently manipulated, by a factor of three, D trial frequency [0 (None), 36 (Baseline), 108 (More), and 324 (Many)] and D trial duration [0 ms (None), 450 ms (Baseline), 1350 ms (Long), and 4050 ms (Longer); see Table 1]. The baseline condition was 36 A, B, C, and D trials, all presented for 450 ms, with a $\Delta p = 0$. For all the experimental conditions, the A, B, and C trials had the same frequency and duration, in which they were presented 36 times each for 450 ms. Table 1 depicts each of the conditions for the 12 D trial manipulations: *i*) *none*, no D trials; *ii*) *baseline*, 36 D trials at 450 ms duration; *iii*) *more*, 108 D trials at 450 ms; *iv*) *many*, 324 D trials at 450 ms; *v*) *long*, 36 D trials at 1350 ms; and vi) *longer*, 36 D trials at 4050 ms. There was no conventional ITI in any condition. The only periods of stimulus absence during training were the programmed and marked D trials. This experiment has not been preregistered.

Procedure—We used ePrime 2.0 (Psychology Software Tools, Pittsburgh, PA) to programme the task (see Supplementary Materials, for more details). Participants provided informed consent and were asked to turn off their mobile phones. The participants completed the computer task alone in a computer cubicle and followed on-screen instructions.

The task involved observing the appearance of two figures, a geometric shape (S1) and a drawing (S2; for an example, see Figure 1) selected from two independent lists. Shapes and drawings were consistently yoked for all participants, but assignment of yoked pairs to conditions was completely randomized without replacement, independently across participants. As presented in the right-hand panel of Figure 1, every trial was displayed in a sequentially alternating position on the screen respective to the fixation cross (e.g., Right-Left-Right-Left, etc.). We did this to distinguish trials and avoid ambiguities arising when two consecutive trials of the same type are presented, which would have made it difficult to differentiate whether there were two of the same type of trial or one longer duration trial. Additionally, using the trial markers (TM; black squares around the stimuli as depicted in Figure 1) provided another feature to mark individual trials.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

For the delay judgment tests with a retention interval, participants were instructed at the end of each condition stream to complete a word search on a sheet of paper until they saw a flashing signal on the monitor which was timed to occur three minutes from the start. Participants stopped the flashing screen by clicking on “continue”, then the testing phase commenced wherein they were instructed to rate on the computer the association of the two stimuli.

At the beginning of the experiment, the participant was instructed to provide their age and gender, after which the instructions read:

In this experiment, you will be watching numerous series of rapidly presented shapes and drawings. After each series, a question screen will appear and you will be asked to rate the degree of relatedness between the shape and drawing on a scale from −10 to +10. Please keep your eyes on the cross in the center of the screen.

> A STRONG POSITIVE RATING should be given when the shape and picture are always presented together and when one is absent the other is also absent.
>
> A STRONG NEGATIVE RATING should be given when the shape is always presented without the picture and the picture is always presented without the shape.

At the end of each training condition, participants were tested by being asked the question below. Responses were made pressing a labelled key in the keyboard, using a Likert scale from −10 to 10 with increases of 1 (i.e., 21 possible responses):

Please indicate the degree of relatedness between the [shape (S1) image inserted here] *and the* [drawing (S2) image inserted here] *in the series that you just saw. Use the rating scale below to enter the degree of relatedness:*

> −10 = strongly negatively related
>
> 0 = not related
>
> +10 = strongly positively related

Each condition was programmed into four training blocks, with the Δp for that condition being met within each block; participants were not informed of the transitions between blocks. Within each of the four blocks, the trials were randomized without replacement. Presenting trials by blocks avoided random selection biasing the overall relationship, either at the beginning or end of the condition. Note that manipulating the frequency of D trials influences the Δp as given by the formula in Figure 1, as does manipulation of the duration of D trials if Δp is calculated based on duration. The nominal Δp for every condition is given in Table 1. Before starting the 12 experimental conditions, participants experienced a practice condition, which was not used for the analysis. In the practice condition, each trial type was presented 36 times at 450 ms, and the Δp = 0; thus, it was identical to the baseline condition and *post hoc* analysis found that the practice condition received similar ratings to the randomly presented baseline condition.

Analysis plan—We used linear mixed models (LMM) to analyse the ratings. The Full model included trial frequency, trial duration, and retention interval as regressors; we

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

added condition order as a covariate (as a discrete variable). Given that the experiment is not a full factorial design, we only included the pertinent interactions. Hence, the Full model was [in *lmer()* R notation]: *ratings ~ 1 + (frequency + duration) * retention + cond_order + (1 + retention + frequency + duration | participant)*, where the random intercept effects were participant, and random slope effects were retention, frequency, and duration, resulting in a maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). The variable retention was binary, coded as 1 for 3-mins retention and 0 for no retention, the frequency and duration variables were coded as the contrast values (see below). We followed two approaches for the analysis, a *data-driven* and hypothesis-based *model comparison* approach.

In order to find out which regressors were the most relevant from the *data-driven* approach we used a backward stepwise algorithm (Kuznetsova, Brockhoff, & Christensen 2015). The Full model was fitted for frequency (0, 36, 108, 324) and for duration (0 ms, 450 ms, 1350 ms and 4050 ms) using the following contrast −13, −9, −1, and 23 (Schad, Vasishth, Hohenstein, & Kliegl, 2020). After filtering the regressors using the stepwise algorithm, we obtained a Final model. The LMM analysis was performed in R using the function *lmer()* from the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). The stepwise algorithm was implemented by using the function *step()* from the package *lmeTest* (Kuznetsova et al., 2015). For the *model comparison* approach, we first compared the reduced models (no retention, no duration, and no frequency) against the Full model by using a likelihood ratio test [in R: *anova(*[Reduced model], [Full model], *test = "LRT")*]. Then, we compared all possible alternative models with the same random effect structure against each other by using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Finally, in the supplementary materials, we report the Bayes Factor approximation (BF; Wagenmakers, 2007) for all model comparisons. For all null hypothesis significance tests used on the estimates from the full models, we used an α threshold of 0.05.

The R scripts code used for statistical analysis, plots, and data files (for the 3 experiments) are available online: https://github.com/santiagocdo/cellD_paper; file called: "coabsent_stats&plots_v6.R" (instructions provided in a *readme* file; Castiello, Miller, Witnauer, Alcaide, Fung, Pitliya, Morrissey, & Murphy, 2021). All analysis and visualizations were made using R (R Core Team, 2019). Additional packages used for data processing were *reshape2* and *dplyr* (Wickham, 2007; Wickham, 2019); models effect sizes and exporting to csv used *report* (Makowski, & Lüdecke, 2019), calculation of Bayes factors used *bayestestR* (Makowski, Ben-Shachar, & Lüdecke, 2019), and data visualization used *ggplot2* (Wickham, 2016), *viridis* (Garnier, 2018), and *ggpubr* (Kassambara, 2018).

Results & discussion

The ratings in this experiment, in which there was only a single zero contingency (i.e., uncorrelated) baseline condition, showed a clear effect of D trial frequency, unlike our previous work (Murphy et al., 2021). The mean ratings from each of the 12 conditions are shown in Figure 2 and the Full model results are in Table 2. Increasing the frequency of co-absent events resulted in higher ratings of association between the stimuli even though the explicit pairings (A trials) as well as the B and C trials were held constant across all

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

conditions in both frequency and duration. The right-hand panel of Figure 2 does not show a corresponding effect of D trial duration: increasing D trial duration while holding all other trial types constant did not result in a corresponding increase in judgments. It is also clear that the retention interval, included to potentially reduce interference from training on the test, had no effect.

Data-driven approach:

After the backward stepwise eliminations of regressors, the Final model included only the frequency factor, frequency random slope, and participant random intercept, suggesting that frequency predicted ratings more accurately than retention, duration, and other interactions. The Full model is presented in Table 2; participant was a random intercept and retention, frequency, and duration were random slopes.

Model comparison approach:

First, we compared the Full model (deviance = 2567.6) against three Reduced models which omitted independent variables. The first Reduced model omitted frequency (deviance = 2573.1), the second duration (deviance = 2571.3), and the last omitted the retention interval (deviance = 2573.7). We found that the Full model was no better than any of the reduced models: $\chi^2(3) = 5.526$, $p = 0.06$; $\chi^2(2) = 3.7$, $p > 0.05$; and $\chi^2(3) = 6.08$, $p > 0.05$; respectively. In addition, we fit all the possible alternative models (see Table S1) and compared them with AIC and BIC. The model with the lowest BIC and AIC values was the one which included only frequency as an independent variable (Table S1). This suggests that frequency provides a better model. See also comparison of all models using Bayes Factors in Figure S2.

In summary, we found that when we increased the frequency of the co-absent events (D trials), participants' judgments of the association between the two stimuli increased. However, this was not the case for the conditions in which we increased the duration of the D trials, nor were judgments impacted when a retention interval was added between the exposures (training) and the test phases.

A previous attempt (Murphy et al., 2021, Experiment 1, 3 and 5) failed to find an effect of a D-trial manipulation. In those experiments, participants received a range of randomly presented positive, zero, and negative baseline contingencies at different trial frequency (E1: 9, 36, 144; E3: 4, 12, 36; E5: 6, 18, 54) and trial duration values (E1: 200, 800 and 3200 ms; E3: 600, 1800, 5400 ms; E5:150, 900 or 5400 ms). It is possible that by including a condition with zero frequency D-events (equivalent to zero duration D-events) here, participants may have been more able to perceive the relatively smaller shifts in contingency achieved by manipulating the frequency of D trials. This would be consistent with the finding that discrimination performance is influenced by the range of comparison stimuli (e.g., Harris, 1952; Matthews & Stewart, 2008). Experiment 2 was motivated by the question of whether it was possible to replicate the effect found in Experiment 1 with a wider range of baseline contingencies.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Experiment 2

Experiment 2 sought to *i*) extend the D trial frequency effect result, found in Experiment 1, to positive and negative contingencies at baseline, and *ii*) replicate the D trial frequency effect with different frequencies at zero contingency at baseline. Given that D trial duration and the retention interval had no appreciable effect in Experiment 1, we only manipulated trial frequency in Experiment 2. We used a different set of frequencies with a lower baseline value in Experiment 2 (i.e., 24) than in Experiment 1 (baseline = 36), doing so permitted us to reduce the overall training time. Using fewer trials seemed justifiable because data from other studies involving the rapid streaming procedure suggested sensitivity to contingency with fewer trials than the number of trials used in Experiment 1 (e.g., Maia et al., 2018). Thus, we hypothesised that by reducing the total number of trials but maintaining the three-fold increase in trials between conditions, we would find a similar effect.

In Experiment 2, we presented the different experimental conditions in separate blocks for the three baseline contingencies (negative, zero, and positive) rather than randomizing them in an attempt to reduce variability in ratings. All conditions with the same number of A, B, and C trials (manipulated to create the baseline negative, zero, and positives contingencies) were presented together in blocks, but we randomised the order of the three contingency blocks across participants.

We also made two more changes from the procedure of Experiment 1. First, the trial markers (TM) borders were made unique for each condition (see Figure S1). We did this to minimise TM-contextual carryover effect that might have increased similarity between conditions, given that trial context influences judgments and behaviour in humans (e.g., Msetfi, Wade, & Murphy, 2013; Soares, Polack, & Miller, 2016), as well as expression of learning by nonhumans (Jozefowiez, Witnauer, & Miller, 2012; Murphy, Baker, & Fouquet, 2001; Urcelay, Witnauer, & Miller, 2012; for a review see also Urcelay & Miller, 2014). Second, we added a secondary discrete context stimulus unique for each condition (Murphy & Baker, 2004; Vallée-Tourangeau, Murphy, Drew & Baker, 1998; Waugh & Norman, 1965).

Methods

Participants—Forty-five participants, 24 females and 21 males (age between 19 and 38, M = 20.4, SD = 2.82), volunteered for this experiment. The experiments were conducted in an open laboratory space, where multiple participants were tested at the same time. We used the same exclusion criterion as Experiment 1 and excluded no participants. This research was approved by the Medical Science Interdivisional Research Ethics Committee of University of Oxford; reference number: R60840/RE001.

Design—The experimental design consisted of twelve conditions (plus one practice condition; total 13 conditions), four negative, four zero, and four positive contingencies (see Table 3). We generated blocks of conditions by contingencies based on the assumption that one potential factor that promoted the D trial effect in Experiment 1 was that participants may have been more attentive to events of low salience (i.e., D trials) when other types of trials did not vary across adjacent conditions. The order of contingency stages (negative,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

zero, and positive) was randomly assigned, and within a contingency stage the four frequency conditions were randomized.

In Experiment 1, we had used the same visual border (trial marker; TM) around the stimuli or absence of stimuli for each condition. Here we varied the TM in each condition (different border textures) and included a further discrete stimulus (a Greek letter that was unique for each condition) on the opposite side of the screen from the stimuli on each trial (see Figure S1).

For each of the three baseline contingencies (negative, zero, and positive), we performed the same manipulation of the four D frequencies: 0 (None), 24 (Baseline), 72 (More), and 216 (Many). The baseline D trial frequency for every contingency was 24. Negative contingencies included 4 A trials and 44 B trials; zero contingencies included 24 A and 24 B trials, and positive contingencies included 44 A trials and 4 B trials, with all three contingency types including 24 C trials. These numbers resulted in the negative baseline contingency having a $\Delta p = [4/(4+44)]-[24/(24+24)] \approx -0.42$; zero baseline contingency having a $\Delta p = [24/(24+24)]-[24/(24+24)] = 0$, and positive baseline contingency having a $\Delta p = [44/(44+4)]-[24/(24+24)] \approx +0.42$ (see Δp values for each condition in Table 3). For example, the Negative More condition ($\Delta p = -0.17$; negative baseline contingency with 72 D trials) was presented in four blocks, each one with 1 A trial, 11 B trials, 6 C trials, and 18 D trials; so the calculated Δp for this block was the same as the overall contingency (i.e., $1/(1+11) - 6/(6+18) \approx -0.17$). To minimize potential trial order effects within conditions, the trials within all conditions were presented in three blocks, with the Δp for that condition being met in each block. The four trial types were always 450 ms in duration. As in Experiment 1, there was no ITI other than the D trials. The practice condition was a replication of the baseline condition with a zero contingency, with 24 A, 24 B, 24 C, and 24 D trials. This experiment has not been preregistered.

Procedure—Participants provided informed consent before receiving instructions for the task. An adapted program (ePrime 2.0; Psychology Software Tools, Pittsburgh, PA) with the same instructions as those provided in Experiment 1 was used (see Experiment 1 and the Supplementary Materials for more details concerning the task).

Analysis plan—For the statistical analysis we used the same program, packages, functions, and α threshold as Experiment 1. The Full model was [in *lmer()* R notation]: *ratings ~ 1 + frequency * contingency + contingency order + (1 + frequency + contingency | participant)*, where participant was a random intercept and frequency and contingency were random slopes. We used the same contrasts as in Experiment 1. The *data-driven* approach was the same as in Experiment 1, that is, we applied a stepwise to the Full model. For the *model comparison* approach, we compared the Full model against a Reduced model without frequency as an independent variable. Using AIC and BIC, we then compare all possible alternative models, including the Null model [*rating ~ 1 + stage_order + (1 + frequency + contingency | subject)*]. To explore the interaction *frequency * contingency,* we assessed three more models, one per contingency stage, that is, negative, zero, or positive data sets. All models included contingency stage order as a covariate.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Results & discussion

Consistent with the findings of Experiment 1 we found a significant D trial frequency effect for each of the three contingencies. Figure 3 illustrates how judgments increased for each contingency as D trials frequency was increased. In addition, we found that the frequency effect interacted with contingency, with the negative contingency having a larger D trial frequency effect than the zero or positive contingencies.

Data-driven approach:

After the backward stepwise algorithm, we found that frequency, contingency, and its interaction, all remained in the Final model, as well as the random effect structure. The Full model results is shown in Table 4.

Model comparison approach:

Next, we compared the Full model (deviance = 2748.6) against a Reduced model without frequency (deviance = 2809.5), and found evidence supporting the Full model in the likelihood ratio test, $\chi^2(2) = 60.97$, $p < 0.001$. In addition, when comparing the Full model against all other alternative models, the former model performed better with lowest BIC and AIC values (Table S2). See also all models' comparisons using Bayes Factors in Figure S3.

Post-hoc analysis.

We analysed each contingency individually, and found that the Frequency effect was significant for positive, zero, and negative contingencies [$\beta_{std} = 0.35$ (small effect size), Estimate = 0.07, 95% CI 0.04 to 0.09; $\beta_{std} = 0.38$ (small effect size), Estimate = 0.11, 95% CI 0.07 to 0.16; $\beta_{std} = 0.52$ (medium effect size), Estimate = 0.18, 95% CI 0.13 to 0.25; respectively]. The complete analysis is presented in Supplementary Materials (Table S3).

Experiment 3

Consistent with Experiment 1, in Experiment 2 increasing the frequency of the co-absent events (i.e., D trials) enhanced the perceived association between the two stimuli. In addition, this effect of D-trial frequency was extended from zero-baseline contingencies (Experiment 1) to positive- and negative-baseline contingencies (Experiment 2). If frequency is more effective than duration, then it is possible to ask whether the observed effect of frequency might be due in part to a masked effect of duration since increasing the trial frequency also increases the overall duration of D trial exposure. That is, our results of frequency may be partially a product of duration. One way of answering this question is to see whether the effect of increasing frequency is still present if we reduce trial duration proportionately.

In Experiment 3, we inversely adjusted D trial duration as D trial frequency was altered. By so doing, we maintained the overall D trial exposure time constant across conditions. Therefore, in Experiment 3 we included conditions with more but shorter D trials and conditions with fewer but longer D trials. On the basis of the assumptions of Δp or our

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

associative mechanism, we predicted that more D trials would contribute to enhanced perception of association even if the total exposure duration was shorter.

Methods

Participants—Forty participants, 27 females and 13 males (age between 18 and 55, M = 27.37, SD = 8.84), served in the experiment. The participants were paid £10.00 at the completion of the experiment. No participants were excluded using the exclusion criterion of Experiment 1. We used the same exclusion criterion as Experiment 1 and excluded no participants. This research was approved by the Medical Science Interdivisional Research Ethics Committee of University of Oxford; reference number: R60840/RE001.

Design—The experimental design consisted of 24 conditions, plus a practice condition which was not used for the analysis. We tested the D frequency effect with respect to three baseline contingencies, Negative ($\Delta p \approx -0.42$), Zero ($\Delta p \approx 0$), and Positive ($\Delta p \approx 0.42$). Table 5 depicts all 24 conditions. The manipulations were: *i*) four D frequencies (0, 8, 24, and 72) with the constant duration (450 ms), *ii*) four D durations (0, 150, 450, and 1350 ms) with the same frequency (24), and *iii*) two D frequency × duration inverse varied conditions: a frequency of 8 trials of 1350 ms, and a frequency of 72 trials of 150 ms; for each of the three baseline contingencies (see Table 5).

We used 24 as the baseline frequency. This was done to avoid a potential ceiling effect that we found with positive contingencies in Experiment 2 as shown in Figure 3. That this change did not undermine sensitivity to D trial frequency was corroborated by fitting a model omitting the 216-frequency condition in Experiment 2; here we found a reliable effect of frequency, stage, and their interaction, Fs > 10, ps < 0.01 (see Table S4).

Negative contingencies included 4 A trials and 44 B trials; zero contingencies, 24 A and 24 B trials, and finally, positive contingencies, 44 A trials and 4 B trials with all three contingency types including 24 C trials. These numbers resulted in the negative baseline contingency having a $\Delta p = [4/(4+44)]-[24/(24+24)] \approx -0.42$; zero baseline contingency having a $\Delta p = [24/(24+24)]-[24/(24+24)] = 0$, and positive baseline contingency having a $\Delta p = [44/(44+4)]-[24/(24+24)] \approx +0.42$ (see Δp values for each condition in Table 5). All conditions were presented in four blocks, with the Δp for that condition being met in each block. A, B, and C trials were always 450 ms in duration, with only D trials being manipulated. The practice condition was a replication of the baseline condition with a zero contingency, with 24 A, 24 B, 24 C, and 24 D trials. This experiment has not been preregistered.

Procedure—We used the same procedure as in Experiment 2. See Supplementary Materials for more details.

Analysis plan—For this statistical analysis, we used the same program, packages, functions, and α threshold as in Experiments 1 and 2. The Full model involved: *rating ~ frequency + contingency + adjusted_duration + all the interactions + contingency_order + (1 + frequency + contingency | participant)*. Participants had a random intercept effect and contingencies and frequency were the random slope effects; and as a covariate we

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

included the contingency stage order. The same contrasts were used, and the same *data-driven* approach was used as in Experiments 1 and 2. Given that our main hypothesis was related to the effect of adjusting the total D trial time (i.e., inversely varying duration) and its effect on the frequency effect, the *model comparison* approach was oriented to test a Reduced model without the 'adjusted by duration' factor against the Full model. In order to have the 'frequency adjusted by duration' as a complete design factor, we duplicated the 0 frequency condition data, so adjusted and non-adjusted at 0 frequency were the same. Because repeating the same condition may bias the analysis by undermining the assumptions of independent observations and variability, we performed a sensitivity analysis with the zero frequency conditions removed. The pattern of results was similar with both analysis strategies (see Table S5).

Results & Discussion

Consistent with the findings of Experiments 1 and 2, in Experiment 3 we found a D trial frequency effect as illustrated in Figure 4 (upper panel). Learning about each of three baseline contingencies suggests an effect of increased ratings of association as a function of increasing the number of co-absent D trials. Adjusting frequency by inversely modifying duration did not attenuate the D trial frequency effect as illustrated in Figure 4. These results suggest that the duration of absent events played very little role in learning the relation between the two stimuli.

Data-driven approach:

After the backward stepwise algorithm, the Final model included only the regressors frequency, contingency, and their interaction; it did not include adjusted by duration nor the stage order covariate. All of the random effects structure survived the stepwise analysis. The Full model results are shown in Table 6.

Model comparison approach:

To explore the effect in more detail and determine whether matching the total D duration was a relevant explanatory factor, we compared the Full model (deviance = 5641.6) against the Reduced model without the adjusted-by-duration factor (deviance = 5641.9), we found a nonsignificant difference in likelihood ratio test, $\chi^2(4) = 0.29$, $p > 0.05$. This suggests that adjusting by duration did not increase the goodness of fit. In addition, using the BIC and AIC, we compared all possible alternative models, including the Null model (Table S6), with the same random effect structure. We found that the most parsimonious model was the one that includes frequency, contingency, and their interaction, with a BIC = 5725.7 and an AIC = 5665.9; where the Full model (which included adjusted by duration) had larger BIC = 5753.3 and AIC = 5673.6. Finally, a Bayes factor analysis pointed in the same direction (Figure S3). This suggests that frequency, rather than duration, explains the differences in ratings across conditions. We reported a similar effect for frequency over duration for A, B, and C trials in our previous work (Murphy et al., 2021).

In summary, participant's contingency judgments increased with more co-absent trials. Even when we inversely varied overall D trial duration with respect to D trial frequency,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

participants were similarly influenced by the frequency and not the duration. With this third experiment, we showed that it is possible to enhance the perceived association between two stimuli without the explicit pairing of the stimuli (A trials), just by adding trials with the mutual absence of the target events in the training context (trial markers). In addition, this result complements previous findings (Murphy et al., 2021) in which we found that learning the objective contingency can be altered by increasing the frequencies of A, B, or C events, that is, trial events with actual stimuli contained within them, even while proportionately decreasing the durations of A, B, or C trials. In those studies, we found there was a benefit for perceiving a contingency from increasing frequency even with a proportionally shorter overall duration of A, B, or C trials. Here, we found a similar effect for the D trials, in Experiment 3 although we did not investigate whether we could actually shorten (as opposed to holding constant) the overall training session length and still observe a benefit. This result is surprising given a general experimental assumption in psychology that the effectiveness of a learning trial is influenced both by experience with that trial and duration of exposure (e.g., Lattal, 1999; Pashler, Rohrer, Cepeda, & Carpender, 2007).

Meta-analysis

A summary comparison of the data pooled from Experiments 1 and 3 (80 participants and 1440 ratings) was conducted. It supports the conclusion that frequency and a frequency-based metric of contingency provides a better fit than one involving duration. The first metric, Δp was calculated using the cell frequency (Δp; Figure 5A); for the second, we calculated Δp by only using cell duration (i.e., $[A_{dur}/(A_{dur}+B_{dur})] - [C_{dur}/(C_{dur}+D_{dur})] = \Delta p_{duration}$; Figure 5B); in the third metric, the Δp was calculated by multiplying frequency and duration, i.e., the total condition duration ($[A_{freq*dur}/(A_{freq*dur}+B_{freq*dur})] - [C_{freq*dur}/(C_{freq*dur}+D_{freq*dur})] = \Delta p_{freq*dur}$; Figure 5C); and finally, the fourth metric was calculated by considering the relative rates at which events occurred, here we took the natural logarithm of the rates ratio [i.e., $\ln(\lambda_{S1}/\lambda_{TM})$] between the rate of occurrence of S2 during S1($\lambda_{S1} = A_{freq} / [(A_{freq} * A_{dur}) + (B_{freq} * B_{dur})]$) and the rate of occurrence of S2 during the context alone or ~S2 (TM; $\lambda_{TM} = C_{freq} / ((C_{freq} * C_{dur}) + (D_{freq} * D_{dur})]$), which is a metric of frequency divided by duration (Figure 5D; see Gallistel & Papachristos, 2020).

Furthermore, we fit four LMMs to predict ratings with each one of the contingency metrics as regressors; participants were random intercepts and nested within experiments. By comparing the BIC and AIC (lower indicates a better fit), we found that the Frequency-based contingency metric (i.e., Δp; Figure 5A) was the best fitting model of ratings compared with Duration-based ($\Delta p_{duration}$), Frequency * Duration-based ($\Delta p_{freq*dur}$), and rates-based $[\ln(\lambda_{S1}/\lambda_{TM})]$ contingency metrics.

This section summarized and corroborated the main effect found in Experiment 1, 2 and 3, in which D trial frequency, not D trial duration, was the critical factor to increase ratings. In addition, in Figure 5 we present evidence that Frequency-based contingency metric (Δp), provides a better linear fit metric than one which includes duration.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

General Discussion

These experiments explored the effect of presenting trials without stimuli on learning an association between two absent target stimuli. We demonstrated the surprising result that frequency of repeated absent event trials is more predictive of performance than overall duration of those trials. Experiment 1 found an effect of D trial (co-absent stimuli) frequency, but not D trial duration, on judgments of a single type of contingency (zero contingency: no correlation between the two stimuli). Increasing D-trial frequency enhanced the perceived relation between the two stimuli. This result is consistent with an account for contingency learning in which both the presentation of the two target stimuli and their absence increases their perceived association (Allan, 1980). The result is also consistent with an increase in response criterion as defined in signal detection theory (SDT), however SDT discriminability does not describe correctly our results (Siegel et al., 2009; see bellow and Figure A1). There was no evidence that the increase of D trial duration nor the introduction of a 3-min delay (retention interval) between learning and judgments resulted in any change in the expression of the learned relationship. Experiment 2 extended the findings to positive and negative baseline contingencies, where the trial frequency effect was replicated. Experiment 3 provided a control for overall D trial duration that corrected for increases in D trial frequency potentially having a corresponding effect on overall duration. Here, the D trial frequency effect was found to be independent of changes in overall D trial duration. The consistent result of the three experiments is that increasing the frequency of co-absent stimuli (D trials) enhances the perceived association between the two stimuli.

The enhanced perceived association produced by D trials may be understood as an increase in contingency discriminability (d'). Applying SDT to contingency judgments results (Allan, Hannah, Crump, & Siegel, 2008; Perales et al., 2005; Siegel et al., 2009), participants are assumed to use experience with the trial types to detect positive or negative contingencies. In the case of negative contingencies, increasing the D-trial frequency should have little effect on the proportion of hits [hit / (hit + miss)], but the proportion of false alarms [false alarm / (false alarm + correct rejection)] is predicted to increase due to a decrease in correct rejections (see Table A2). The effect of this would be to decrease d' as the number of co-absent trials increases. Furthermore, in positive contingencies, the proportion of hits should increase and the proportion of false alarms should remain stable, which would lead to an increase in d' as the number of co-absent trials increase. In summary, and contrary to our results, SDT predicts a decrease in d' as a function of the number of D trials for negative contingencies (see also Mercier & Parr, 1996). The SDT analysis is presented in Figure A1 of the Appendix. SDT does predict our results at least as conceived as the result of a change to the response criterion (see bottom panel Figure A1) owing to an increase in the proportion of false alarms.

Memory theory also explains aspects of the present results. Testing memory for single items, Hintzman (1970) demonstrated that frequency of an item was more relevant for memory performance than the duration of exposure to that item. Those experiments varied frequency and duration of target item exposure and found that total-time affected neither recall nor recognition memory performance, but the frequency of presentation did (see also Hintzman, 2010). In our case, the frequency effect applied to negative, zero and positive

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

baseline contingencies, but was most robust with negative baseline contingencies. Finally, we found that increasing the frequency and reducing the duration of co-absent stimulus trials (maintaining the same overall duration) did not attenuate the effect of D trial frequency. Next, we describe four different accounts for how more or longer absent events, so-called D trials, have been hypothesized to influence contingency ratings: *i*) contingency sensitivity, *ii*) testing effect, *iii*) reduced associative interference by the context, and *iv*) reduced encoding interference.

Contingency Sensitivity.

The first account of the trial spacing effect is that the mutual absence of the two stimuli (D trials) is information supportive of the two stimuli being related, as indicated by Δp. This account may be referred as the process model of Δp (Cheng, 1997; White, 2004). In this framework, learning involves four types of trials and it is the difference in the proportions of the trials in which one event occurs conditional on the other (A/[A+B]) and the proportion of the trials in which one event occurs conditional on the absence of the other (C/[C+D]). Increasing the D trials has the effect of reducing the latter proportion, thereby increasing the objective contingency (e.g., Allan, 1980; Msefti et al., 2005).

Testing Effect.

The second account is the testing effect which is a boost in test performance as a result of retrieval from LTM on a prior test (Karpicke, & Roediger, 2008). This explanation assumes that ‘irrelevant’ D trial experience (at least partially) clears STM of task-relevant information (i.e., A, B, or C trial) so that on the next relevant trial previously acquired relevant information is retrieved from LTM as opposed to its already being in STM as a result of a recent prior trial. From this perspective, each D trial provide a mask for STM and forces retrieval from LTM on the subsequent trial, which in turn boosts later retrieval from LTM of A-, B-, and C-trial information on later A-, B-, and C-trials, respectively (Aust, Haaf, & Stahl, 2019). This retrieval practice mechanism is the presumed basis of the Testing Effect (e.g., Roediger & Butler, 2011) and may explain our enhanced effect, although the effect does not differentiate between frequency and duration and so does not anticipate our frequency-superiority effect.

Associative Interference by the Context.

A third account of the trial spacing effect suggests that during each A trial people learn associations between the two presented stimuli, and on B and C trials the stimuli are associated with the trial context, where the trial context is the set of stimuli unique to the experimental situation (e.g., the rectangular frames in the present research). This context becomes associated with both B and C events, and ultimately presentation of the context alone may retrieve memories of the individual stimuli. One consequence of these context associations is that representations of the two stimuli (S1 and S2) in each trial may be retrieved by the context on A trials and interfere with further learning of the target association (Bouton, 1993; Miller & Escobar, 2002). However, to the extent that the context is extinguished during D trials, these associations might weaken and consequently interfere less with A-trial learning. Therefore, increasing the frequency or duration of D trials would be expected to reduce the effect of interference arising from context learning.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Reduced Encoding Interference.

A fourth account is that including 'empty' D events in each training condition increases the opportunity to further process the information from the preceding A, B or C trial before the next one of these trials begins, thereby providing more available cognitive resources and allowing better processing of the content of each trial (Waugh & Norman, 1965). Our rapidly presented stimuli are presumably processed in STM which has limited capacity (Cowan, 2008). Thus, the faster a continuing stream of information enters STM, the less time is available to process each new stimulus event. This is apt to decrease attentional control and hence increase proactive interference (Engle, 2002). The view here is that slowing down the input of information allows participants more time to consolidate the content of each successive trial into LTM (McGaugh, 1966).

The trial spacing effect

The testing effect and the reduced encoding interference accounts of the trial spacing effect both anticipate that the impacts of A, B, and C trials will be boosted by more D trials, here treated as low cognitive load periods. This means that with more D trials, baseline positive contingencies should have been rated as more positive as was observed, but baseline negative contingencies should have been rated as more negative (see Mercier & Parr, 1996) which is contrary to what we observed. The contingency sensitivity and interference by the training context accounts both fare better in that, regardless of the baseline contingency, increasing the frequency of D trials increases the objective contingency as is evident from the equation $\Delta p = (A/[A+B]) - (C/[C+D])$ and provides more time for extinction of the training context, respectively. However, with the number of C trials fixed across conditions as it was in each of the present experiments, according to the contingency account, variation in the frequency of D trials would be expected to have a similar positive effect on both positive and negative baseline contingencies. Our results revealed a somewhat larger effect with the negative contingencies, supported by the interaction between frequency and contingency in Experiment 2 and 3 (i.e., increasing the number of D trials increased ratings marginally more when the baseline contingency was negative than when it was positive). Hence, the superiority of the contingency sensitivity account over the other accounts with respect to the present data is qualified. Turning to interference by the context, context-S1 and context-S2 associations might be expected to interfere more with memories of A trials (i.e., S1–S2) than B or C trials. Thus, more D trials (i.e., more extinction of the context) would be expected in increase contingency ratings with both positive and negative contingencies, as was observed. One might think that extinction would also be enhanced by longer D trials, but considerable prior research has found extinction to depend more on number of extinction trials than the duration of the extinction trials (Harris & Andrews, 2017). Consequently, based on the present data alone, the benefit of more D trials seemingly arises from greater opportunity for extinction of relevant associations to the context.

Probabilities versus rates

It was surprising to us that we did not find an effect of duration on judgments of association given that duration is well known to play a fundamental role in learning (Balsam & Gallistel, 2009; Miller & Barnet, 1993). We briefly summarize three families of learning models

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

which treat time in different ways (for a more detailed review of the first two, see Bonardi, Cheung, Mondragón, & Tam, 2016). First, there are *time-accumulation models* (e.g., Gallistel & Gibbon, 2000; Gibbon & Balsam, 1981), in which events are accumulated in a period of times (i.e., rates). Second, there are *trial-based models* (e.g., Rescorla & Wagner, 1972) which do not explicitly code the duration of events; this appears problematic given the often-reported relevance of temporal information as a learning factor in both humans and animals (e.g., Buriticá & Alcalá, 2019; Nasser & Delamater, 2016). However, a number of attempts to integrate timing into the classic *trial-based models* have been suggested (e.g., Delamater, Desouza, Rivkin, & Derman, 2014; Donahoe, Burgos, & Palmer, 1993; Gershman, 2015; Ghirlanda, Lind, & Enquist, 2020; Luzardo, Alonso, & Mondragon, 2017). Third, there are *behavioural timing models* (e.g., Killeen & Fetterman, 1988; Machado, 1997), in which an agent's internal states (behavioural states) serve as events that are associated with moments in the environment at a particular time. Our results support a trial-based approach of learning where trial duration has no clear effect and the critical factor is how many times an event is presented. The difference between multiple trials and a long duration is the repeated change of the state of the cues, from off to on or on to off which may itself support learning. There is evidence that transitions are more informative than steady states (Hintzman, 1970), and that transitions are themselves cues for learning (Murphy, Mondragon & Murphy, 2009).

Animal Literature

Comparisons between human contingency learning and animal conditioning have been previously suggested to reflect a similar underlying learning process (Baker & Mackintosh, 1977; Baker & Mackintosh, 1979; Miller, 1982; Shanks, 1995). While animal behaviour does map onto contingency (Rescorla, 1967; Murphy, & Baker, 2004), there is evidence that supports duration (time) instead of frequency during the acquisition process (e.g., Balsam et al., 2010; Escobar et al., 2002; Gallistel & Papachristos, 2020; Harris & Bouton, 2020; Harris et al., 2015; Urcelay et al., 2009). However, during extinction, the presentation of the CS without the US, CS frequency seems to be more relevant for predicting decreases in conditioned responding (Harris & Andrew, 2017; Chan & Harris, 2019). Some of this research has involved absent events being defined by extended exposure to the constellation of stimuli provided by the context or chamber in which the animals were housed. With the right/left alternation of stimuli and use of frames, we endeavoured to make our trials more discrete and enhance the salience of the D trials to encourage context relevance for learning.

Limitations

An important point concerning D trials, consistent with Murphy et al.'s (2021) findings concerning A, B, and C trials, is that we found enhanced association ratings even when the increased frequency of D trials was accompanied by a proportionately shorter duration of the D trials. It remains unclear whether this effect would also occur in paradigms that are not built on an explicit frequency basis, as was the case in the present streaming task (Allan et al., 2008). For instance, as previously mentioned, the Δp formulation has been criticised because of the view that counting of non-occurring events may be problematic given that it is not obvious how participants count the number of times something has not happened (see Gallistel, Craig, & Shahan, 2014), other than by laboratory strategies such

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

as we used here. As an adapted example taken from Gallistel et al. (2014), it is difficult to imagine how many absences of earthquakes in Mexico City have occurred since 2017, but we can perceive the duration of that absence. Other contingency analyses based on information theory exist in which, instead of counting events, participants use the entropy of the distribution of the intervals between cues and outcomes (Ward, Gallistel, & Balsam, 2013). In the present work, trials were explicitly discrete with different sequential topology and every trial was marked, which may have favoured a counting strategy over a strategy of accumulating event durations (rates). In this regard, the present experiments are consistent with much of experimental psychology which often uses a trial-based strategy to assess human performance. Our results contradict a general hypothesis that treats longer stimulus exposure as an increased opportunity to process a stimulus and a decreased opportunity for interference from competing trials. At least with contingency learning, increasing trial duration was less effective than providing more frequent, shorter trial bursts.

One might ask whether similar results would have been observed with trials of more conventional duration as opposed to the rapid streaming procedure. Obviously, the present data does not answer that question, although the streaming-trial procedure did diminish the likelihood that our effects are based on counting. However, in one experiment in the Murphy et al. (2021) paper, increasing all trial durations by a factor of three did not alter the greater effect of frequency than duration on A, B, or C trials.

Implications

One of the most robust phenomena in the learning literature is the trial spacing effect: greater spacing of training trials results in superior performance on test trials (Vlach, & Sandhofer, 2021). In applied situations such as educational and psychotherapeutic settings, one challenge to the spacing of training trials is that spacing training ordinarily increases the total time needed for training. The present findings suggest that frequency of D trials rather than D trial duration is critical for an enhanced perception of association. If we believe that there is a fundamental similarity between trial spacing and co-absent events, this would suggest that the benefit of trial spacing is not due to time per se between training trials, but the number of co-absent events that agent's perceive between training trials. Hence, more intertrial events (i.e., D trials) can be added without an extra cost in training time by making these events shorter. However, this generalization should be entertained with caution. Given that trial spacing in many other situations is seemingly time dependent and our findings suggest that duration does not plays a critical role, trial spacing and number of co-absent events may act through fundamentally different underlying processes. If this were the case, generalizing our findings to other preparations would be premature.

Conclusions

Our results support the view that humans use the mutual absence of stimuli (D trials) to guide their judgments of association (i.e., contingency). Furthermore, it is possible to enhance perceived contingencies between two stimuli by increasing the frequency (but not the duration) of D trials, even when these trials have a short duration, a frequency-superiority effect. The results support accounts of the benefit of co-absence events based on (a) participants tracking the Δp contingency, and (b) reduced associative interference with

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

information concerning A trials by the context. These findings have potential applications for boosting learning without additional explicit pairing of stimuli or lengthening of training session.

Context

One of the most ubiquitous phenomena in Psychology is the trial spacing effect; spaced training trials result in better learning and retention. However, the cost of spacing training trials is that training sessions must be longer or there have to be more training sessions. In contingency learning sessions, trials with no cues or outcomes (D trials), are similar to standard intertrial intervals (ITIs). Different theories have been proposed to explain why human and non-human animals learn better with longer intervals between trials. Although the very term 'intertrial interval' focuses on the duration of the interval, typically the duration of the interval is confounded with seemingly irrelevant events that occur within the ITIs. In the present research, we dissociated ITI duration from inter trial frequency (the ITF), and found that frequency of D trials rather than duration of D trials is what enhances perceived association, at least in contingency learning. Thus, stronger associative learning can be obtained by increasing D trials while making the D trials proportionally shorter.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

This research was supported in part by research funds from Corpus Christi College to RAM, NIH grant MH033881 to RRM, and D.Phil. (Ph.D.) scholarship by the University of Guadalajara, Mexico (V/2018/1476) to SC.

The authors thank Sarah Chew, Dennis Elengickal, Jovin Huang, Sammi Lin, Tessa L. Livingston, Cameron M. Mccrea, Jared Silverstein, and Meghan Wiseman for their commenting on a preliminary version of the manuscript. These studies have not been preregistered. This work has been presented previously in: Virtual Associative Learning Symposium 2020, Conference on Comparative Cognition, Melbourne, FL, 2020 (accepted, scheduled, and prepared, but not delivered due to covid-19 pandemic); and Experimental Psychology Society online meeting January 2021 (doi: 10.13140/RG.2.2.30851.73766). We also than to three anonymous reviewers and Professor Mathew Crump for excellent comments on a previous version of the manuscript.

Appendix

We described the data from Experiment 3 in the framework of signal detection theory (SDT). We classified contingencies as positive and negative based on Δp calculated with the trial frequency for the baseline condition. Additionally, we divided participants' ratings into positive and negative detections, where ratings >= mean (ratings) were classified as positive detection, and ratings < mean(ratings) were classified as negative detection (Table A1).

Table A1.

Signal Detection Theory, each cell in the 2 × 2 table represent frequencies

	Positive contingency	**Negative contingency**
Positive detection	Hit	False Alarm

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

	Positive contingency	Negative contingency
Negative detection	Miss	Correct rejection

We obtained the detectability (d') and the response criterion (c) per conditions by using the following equations:

$$p(h) = Hit/(\ Miss + Hit\ ), \quad \text{(eq. 1)}$$

$$p(f) = False\ Alarm/(\ False\ Alarm + Correct\ Rejection\ ), \quad \text{(eq. 2)}$$

$$d' = z[p(h)] - z[p(f)], \quad \text{(eq. 3)}$$

$$c = -1 * (z[p(h)] + z[p(f)])/2 \quad \text{(eq. 4)}$$

where z[] scores represent trials that correspond to these right-tail, cumulative normal probabilities. The complete code is available online in the file called "coabsent_stats&plots_v6.R" in https://github.com/santiagocdo/cellD_paper.

When we analyse each condition in Experiment 3 with SDT and plot d' and c as a function of D frequency, we generate Figure A1, panel C. This analysis shows that discriminability increases as a function of D-trial frequency for both zero and positive contingencies; however, it decreases for negative contingencies. Thus, given zero and positive contingencies, higher contingency ratings with increasing frequency of D trials may be explained as an increase in detecting positive contingencies. However, in the case of negative contingencies, a different process may be at work, one in which increases in frequency of D trials decrease detectability (d'). Finally, SDT suggests that the response criterion should decrease as a function of D-trial frequency (Figure A1, panel C).

Table A2

Signal Detection Theory (SDT) analysis of Experiment 3

	Name	Frequency	Duration	Δp	p(h)	p(f)	d'	c	Hit	FA	Miss	CR
Negative	None	0	0	−0.92	0.025	0.125	−0.81	1.56	0	5	0	35
	Few	8	450	−0.67	0.025	0.125	−0.81	1.56	0	5	0	35
	Few Adj.	8	1350	−0.67	0.025	0.050	−0.32	1.80	0	2	0	38
	Baseline	24	450	−0.42	0.025	0.150	−0.92	1.50	0	6	0	34
	Short	24	150	−0.42	0.025	0.100	−0.68	1.62	0	4	0	36
	Long	24	1350	−0.42	0.025	0.175	−1.03	1.45	0	7	0	33
	More Adj.	72	150	−0.17	0.026	0.333	−1.52	1.19	0	13	0	26
	More	72	450	−0.17	0.026	0.462	−1.85	1.02	0	18	0	21
Zero	None	0	0	−0.50	0.025	0.500	−1.96	0.98	0	20	0	20

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

	Name	Frequency	Duration	Δp	p(h)	p(f)	d’	c	Hit	FA	Miss	CR
Positive	Few	8	1350	−0.25	0.025	0.550	−2.09	0.92	0	22	0	18
	Few Adj.	8	450	−0.25	0.026	0.487	−1.92	0.99	0	19	0	20
	Baseline	24	150	0.00	0.525	0.025	2.02	0.95	21	0	19	0
	Short	24	450	0.00	0.550	0.025	2.09	0.92	22	0	18	0
	Long	24	1350	0.00	0.550	0.025	2.09	0.92	22	0	18	0
	More Adj.	72	450	0.25	0.700	0.025	2.48	0.72	28	0	12	0
	More	72	150	0.25	0.775	0.025	2.72	0.60	31	0	9	0
	None	0	0	−0.08	0.026	0.795	−2.77	0.56	0	31	0	8
	Few	8	450	0.17	0.850	0.025	3.00	0.46	34	0	6	0
	Few Adj.	8	1350	0.17	0.875	0.025	3.11	0.40	35	0	5	0
	Baseline	24	450	0.42	0.925	0.025	3.40	0.26	37	0	3	0
	Short	24	150	0.42	0.850	0.025	3.00	0.46	34	0	6	0
	Long	24	1350	0.42	0.950	0.025	3.60	0.16	38	0	2	0
	More Adj.	72	450	0.67	0.900	0.025	3.24	0.34	36	0	4	0
	More	72	150	0.67	0.925	0.025	3.40	0.26	37	0	3	0

Note: **Frequency** and **Duration** relative to the D trials. **Δp** as calculated in Figure 1. **p(h)** = **Hit** / (**Miss** + **Hit**) and **p(f)** = **FA** / (**FA** + **CR**), where **FA** is False Alarm and **CR** is Correct rejection. **d’** = z[**p(h)**] - z[**p(f)**] and **c** = (z[**p(h)**] - z[**p(f)**]) / 2, where z[] is the right-tail, cumulative normal probabilities.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

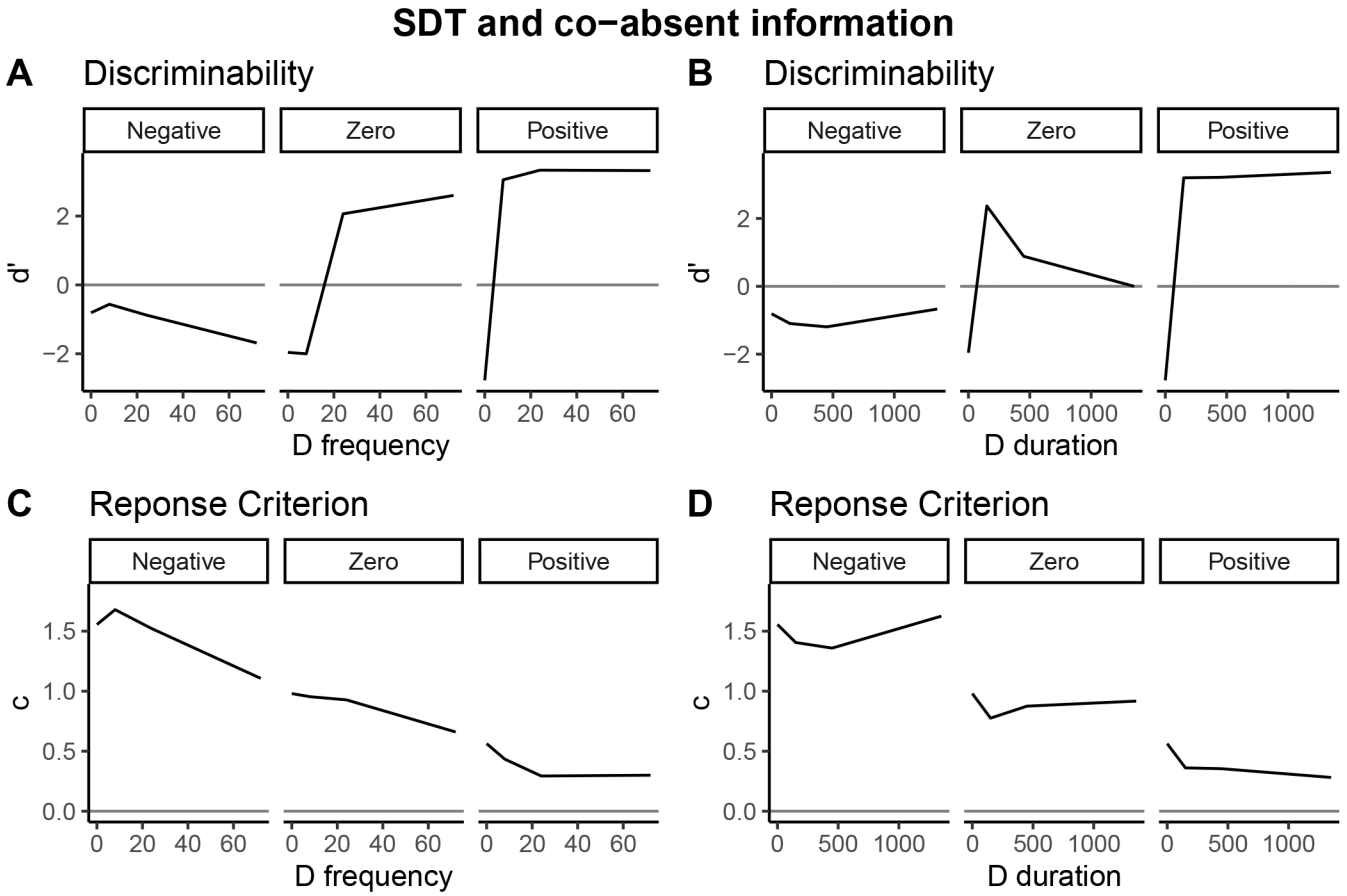

Figure A1.

Signal detection theory analysis of Experiment 3. **A** is discriminability (d') as a function of D-trial frequency; **B** is discriminability (d') as a function of D-trial duration; **C** is response criterion (c) as a function of D-trial frequency; **D** is response criterion (c) as a function of D-trial duration.

References

Allan LG (1980). A note on measurement of contingency between two binary variables in judgment tasks. Bulletin of the Psychonomic Society, 15(3), 147–149. 10.3758/BF03334492

Allan LG, Hannah SD, Crump MJ, & Siegel S (2008). The psychophysics of contingency assessment. Journal of Experimental Psychology: General, 137(2), 226. 10.1037/0096-3445.137.2.226 [PubMed: 18473655]

Aust F, Haaf JM, & Stahl C (2019). A memory-based judgment account of expectancy-liking dissociations in evaluative conditioning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 45(3), 417. 10.1037/xlm0000600 [PubMed: 30024247]

Baker AG, & Mackintosh NJ (1977). Excitatory and inhibitory conditioning following uncorrelated presentations of CS and UCS. Animal Learning & Behavior, 5(3), 315–319. 10.3758/BF03209246

Baker AG, & Mackintosh NJ (1979). Preexposure to the CS alone, US alone, or CS and US uncorrelated: Latent inhibition, blocking by context or learned irrelevance?. Learning and Motivation, 10(3), 278–294. 10.1016/0023-9690(79)90034-1

Baker AG, Mercier P, Vallée-Tourangeau F, Frank R, & Pan M (1993). Selective associations and causality judgments: Presence of a strong causal factor may reduce judgments of a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

weaker one. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(2), 414. 10.1037/0278-7393.19.2.414

Baker AG, Murphy RA & Vallée-Tourangeau F (1996). Associative and normative models of causal induction: Reacting to versus understanding cause. In Shanks DR, Holyoak KJ & Medin DL (Eds.), The Psychology of Learning and Motivation, Vol. 34, (pp. 1–45) San Diego, CA: Academic Press.

Baker AG, Murphy R, Mehta R, & Baetu I (2005). Mental models of causation: A comparative view. In Wills AJ (Ed.), New directions in human associative learning (p. 11–40). Lawrence Erlbaum Associates Publishers, New Jersey, USA.

Balsam PD, Drew MR, & Gallistel CR (2010). Time and associative learning. Comparative Cognition & Behavior Reviews, 5, 1–22. 10.3819/ccbr.2010.50001 [PubMed: 21359131]

Balsam PD, & Gallistel CR (2009). Temporal maps and informativeness in associative learning. Trends in neurosciences, 32(2), 73–78. 10.1016/j.tins.2008.10.004 [PubMed: 19136158]

Barr DJ, Levy R, Scheepers C, & Tily HJ (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68(3), 255–278. 10.1016/j.jml.2012.11.001

Bates D, Mächler M, Bolker B, Walker S (2015). Fitting linear mixed-effects models Using lme4. Journal of Statistical Software, 67(1), 1–48. 10.18637/jss.v067.i01

Blanco F, Barberia I, & Matute H (2015). Individuals who believe in the paranormal expose themselves to biased information and develop more causal illusions than nonbelievers in the laboratory. PloS one, 10(7), e0131378. [PubMed: 26177025]

Bonardi C, Cheung TH, Mondragón E, & Tam SK (2016). Timing and conditioning. In Murphy RA, & Honey RC (Eds) The Wiley handbook on the cognitive neuroscience of learning, 348–379.

Bouton ME (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. Psychological Bulletin, 114(1), 80. 10.1037/0033-2909.114.1.80 [PubMed: 8346330]

Bouton ME, & Nelson JB (1998). The role of context in classical conditioning: Some implications for cognitive behavior therapy. In O'Donohue WT (Ed.), Learning and behavior therapy (p. 59–84). Allyn & Bacon, USA.

Buriticá J, & Alcalá E (2019). Increased generalization in a peak procedure after delayed reinforcement. Behavioural Processes, 169, 103978. 10.1016/j.beproc.2019.103978 [PubMed: 31580905]

Castiello S, Miller RR, Witnauer JE, Alcaide DM, Fung E, Pitliya RJ, Morrissey D, & Murphy RA (2021, March 3). cellD_paper. GitHub repository. https://github.com/santiagocdo/cellD_paper

Chan CK, & Harris JA (2019). The partial reinforcement extinction effect: The proportion of trials reinforced during conditioning predicts the number of trials to extinction. Journal of Experimental Psychology: Animal Learning and Cognition, 45(1), 43. 10.1037/xan0000190 [PubMed: 30604994]

Cheng PW (1997). From covariation to causation: A causal power theory. Psychological Review, 104(2), 367–405. 10.1037/0033-295X.104.2.367

Cowan N (2008). What are the differences between long-term, short-term, and working memory?. Progress in brain research, 169, 323–338. 10.1016/S0079-6123(07)00020-9 [PubMed: 18394484]

Crump MJ, Hannah SD, Allan LG, & Hord LK (2007). Contingency judgements on the fly. Quarterly Journal of Experimental Psychology, 60(6), 753–761. 10.1080/17470210701257685

Delamater AR, Desouza A, Rivkin Y, & Derman R (2014). Associative and temporal processes: A dual process approach. Behavioural Processes, 101, 38–48. 10.1016/j.beproc.2013.09.004 [PubMed: 24076309]

Dickinson A, Shanks D, & Evenden J (1984). Judgement of act-outcome contingency: The role of selective attribution. Quarterly Journal of Experimental Psychology, 36(1), 29–50. 10.1080/14640748408401502

Dickinson A, & Shanks D (1987). Associative accounts of causality judgment. The Psychology of Learning and Motivation, 21, 229–261.

Donahoe JW, Burgos JE, & Palmer DC (1993). A selectionist approach to reinforcement. Journal of the Experimental Analysis of Behavior, 60(1), 17–40. 10.1901/jeab.1993.60-17 [PubMed: 8354965]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Engle RW (2002). Working memory capacity as executive attention. Current Directions in Psychological Science, 11(1), 19–23. 10.1111/1467-8721.00160

Escobar M, Arcediano F, & Miller RR (2002). Latent inhibition and contextual associations. Journal of Experimental Psychology: Animal Behavior Processes, 28(2), 123. 10.1037/0097-7403.28.2.123 [PubMed: 11987870]

Fiddick L, Cosmides L, & Tooby J (2000). No interpretation without representation: The role of domain-specific representations and inferences in the Wason selection task. Cognition, 77(1), 1–79. 10.1016/S0010-0277(00)00085-8 [PubMed: 10980253]

Gallistel CR, & Papachristos EB (2020). Number and time in acquisition, extinction and recovery. Journal of the Experimental Analysis of Behavior, 113(1), 15–36. 10.1002/jeab.571 [PubMed: 31856323]

Gallistel CR, Craig AR, & Shahan TA (2014). Temporal contingency. Behavioural Processes, 101, 89–96. 10.1016/j.beproc.2013.08.012 [PubMed: 23994260]

Gallistel CR, & Gibbon J (2000). Time, rate, and conditioning. Psychological Review, 107(2), 289–344. 10.1037/0033-295X.107.2.289 [PubMed: 10789198]

Garnier S, (2018). viridis: Default Color Maps from ‘matplotlib’. R package version 0.5.1. https://CRAN.R-project.org/package=viridis

Gershman SJ (2015). A unifying probabilistic view of associative learning. PLoS Computational Biology, 11(11). 10.1371/journal.pcbi.1005829

Ghirlanda S, Lind J, & Enquist M (2020). A-learning: A new formulation of associative learning theory. Psychonomic Bulletin & Review, 1–29. 10.3758/s13423-020-01749-0

Gibbon J, & Balsam P. (1981). Spreading association in time. In: Locurto CM, Terrace HS, Gibbon J, editors. Autoshaping and conditioning theory (pp. 219–253). New York: Academic Press.

Gibbon J, Church RM, & Meck WH (1984). Scalar timing in memory. Annals of the New York Academy of sciences, 423(1), 52–77. 10.1111/j.1749-6632.1984.tb23417.x [PubMed: 6588812]

Harrell FE Jr, with contributions from Charles Dupont and many others. (2019). Hmisc: Harrell Miscellaneous. R package version 4.2–0. https://CRAN.R-project.org/package=Hmisc

Harris JD (1952). The decline of pitch discrimination with time. Journal of Experimental Psychology, 43(2), 96–99. 10.1037/h0057373 [PubMed: 14927806]

Harris JA, & Bouton ME (2020). Pavlovian conditioning under partial reinforcement: The effects of nonreinforced trials versus cumulative conditioned stimulus duration. Journal of Experimental Psychology. Animal Learning and Cognition, 46(3), 256–272. 10.1037/xan0000242 [PubMed: 32162948]

Harris JA, Patterson AE, & Gharaei S (2015). Pavlovian conditioning and cumulative reinforcement rate. Journal of Experimental Psychology: Animal Learning and Cognition, 41(2), 137. 10.1037/xan0000054 [PubMed: 25734756]

Harris JA, & Andrew BJ (2017). Time, trials, and extinction. Journal of Experimental Psychology: Animal Learning and Cognition, 43(1), 15. 10.1037/xan0000125 [PubMed: 28045292]

Hintzman DL (1970). Effects of repetition and exposure duration on memory. Journal of Experimental Psychology, 83(3, Pt.1), 435–444. 10.1037/h0028865

Hintzman DL (2010). How does repetition affect memory? Evidence from judgments of recency. Memory & Cognition, 38(1), https://doi.org/102-115.10.3758/MC.38.1.102

Jenkins HM, & Sainsbury RS (1969). The development of stimulus control through differential reinforcement. In Mackintosh NJ, & Honig WK (Eds.), Foundation mental issues in associative learning, pp. 123–167. Dalhousie Univ. Press, Halifax.

Johnson-Laird PN, & Wason PC (1970). A theoretical analysis of insight into a reasoning task. Cognitive Psychology, 1(2), 134–148. 10.1016/0010-0285(70)90009-5

Jozefowiez J, Witnauer JE, & Miller RR (2012). The temporal pattern of responding in conditioned bar-press suppression: The role of the context switch and training mode. Behavioural Processes, 89(3), 239–243. 10.1016/j.beproc.2011.12.002 [PubMed: 22178451]

Juarez-Ramos V, Rubio JL, Delpero C, Mioni G, Stablum F, & Gomez-Milan E (2014). Jumping to conclusions bias, BADE and feedback sensitivity in schizophrenia and schizotypy. Consciousness and Cognition, 26, 133–144. 10.1016/j.concog.2014.03.006 [PubMed: 24747992]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Kassambara A, (2018). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.2. https://CRAN.R-project.org/package=ggpubr

King DL, Jones FL, Pearlman RC, Tishman A, & Felix CA (2002). The length of the retention interval, forgetting, and subjective similarity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(4), 660. 10.1037/0278-7393.28.4.660 [PubMed: 12109760]

Karpicke JD, & Roediger HL (2008). The critical importance of retrieval for learning. Science, 319(5865), 966–968. 10.1126/science.1152408 [PubMed: 18276894]

Kuznetsova A, Brockhoff PB, & Christensen RHB (2017). lmerTest package: tests in linear mixed effects models. Journal of Statistical Software, 82(13). 10.18637/jss.v082.i13

Maia S, Lefèvre F, & Jozefowiez J (2018). Psychophysics of associative learning: Quantitative properties of subjective contingency. Journal of Experimental Psychology: Animal Learning and Cognition, 44(1), 67. 10.1037/xan0000153 [PubMed: 29154562]

Makowski D & Lüdecke D (2019). The report package for R: Ensuring the use of best practices for results reporting. CRAN. Available from https://github.com/easystats/report

Makowski D, Ben-Shachar M, & Lüdecke D (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. Journal of Open Source Software, 4(40), 1541. 10.21105/joss.01541

Mata A, Garcia-Marques L, Ferreira MB, & Mendonça C (2015). Goal-driven reasoning overcomes cell D neglect in contingency judgements. Journal of Cognitive Psychology, 27(2), 238–249. 10.1080/20445911.2014.982129

Matthews WJ, & Stewart N (2008). The effect of stimulus range on two-interval frequency discrimination. The Journal of the Acoustical Society of America, 123(4), EL45–EL51. 10.1121/1.2884084 [PubMed: 18396920]

McGaugh JL (1966). Time-dependent processes in memory storage. Science, 153(3742), 1351–1358. [PubMed: 5917768]

Mercier P, & Parr W (1996). Inter-trial interval, stimulus duration and number of trials in contingency judgments. British Journal of Psychology, 87(4), 549–566. 10.1111/j.2044-8295.1996.tb02608.x

Miller RR (1982). Effects of intertrial reinstatement of training stimuli on complex maze learning in rats: Evidence that "acquisition" curves reflect more than acquisition. Journal of Experimental Psychology: Animal Behavior Processes, 8(1), 86–109. 10.1037/0097-7403.8.1.86

Miller RR, & Barnet RC (1993). The role of time in elementary associations. Current Directions in Psychological Science, 2(4), 106–111. 10.1111/1467-8721.ep10772577

Miller RR, & Escobar M (2002). Associative interference between cues and between outcomes presented together and presented apart: An integration. Behavioural Processes, 57(2–3), 163–185. 10.1037/0033-2909.117.3.363 [PubMed: 11947996]

Molet M, Jozefowiez J, & Miller RR (2010). Integration of spatial relationships and temporal relationships in humans. Learning & behavior, 38(1), 27–34. 10.3758/LB.38.1.27 [PubMed: 20065346]

Moritz S, & Woodward TS (2006). A generalized bias against disconfirmatory evidence in schizophrenia. Psychiatry Research, 142(2–3), 157–165. 10.1016/j.psychres.2005.08.016 [PubMed: 16631258]

Msetfi RM, Murphy RA, Simpson J, & Kornbrot DE (2005). Depressive Realism and Outcome Density Bias in Contingency Judgments: The Effect of the Context and Intertrial Interval. Journal of Experimental Psychology: General, 134(1), 10–22. 10.1037/0096-3445.134.1.10 [PubMed: 15702960]

Msetfi RM, Murphy RA, & Simpson J (2007). Depressive realism and the effect of intertrial interval on judgements of zero, positive, and negative contingencies. Quarterly Journal of Experimental Psychology (2006), 60(3), 461–481. 10.1080/17470210601002595 [PubMed: 17366312]

Msetfi RM, Wade C, & Murphy RA (2013). Context and time in causal learning: Contingency and mood dependent effects. PLoS One, 8(5). e64063. 10.1371/journal.pone.0064063 [PubMed: 23691147]

Murphy RA, & Baker AG (2004). A role for CS-US contingency in Pavlovian conditioning. Journal of Experimental Psychology: Animal Behavior Processes, 30(3), 229. 10.1037/0097-7403.30.3.229 [PubMed: 15279513]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Murphy RA, Baker AG, & Fouquet N (2001). Relative validity of contextual and discrete cues. Journal of Experimental Psychology: Animal Behavior Processes, 27(2), 137–152. 10.1037/0097-7403.27.2.137 [PubMed: 11296489]

Murphy RA, Schmeer S, Vallee-Tourangeau F, Mondragon E, & Hilton D (2011). Making the illusory correlation effect appear and then disappear: The effects of increased learning. Quarterly Journal of Experimental Psychology, 64(1), 24–40.

Murphy R, Byrom N, & Msetfi RM (2017). The problem with explaining symptoms: The origin of biases in causal processing. European Journal for Person Centered Healthcare, 5(3). 10.5750/ejpch.v5i3.1318

Murphy RA, Mondragon E & Murphy VA (2009). Contingency, temporal order and generalization: Building blocks of causal cognition. International Journal of Comparative Psychology, 22, 61–74. 10.5070/P4221020518

Murphy RA, Witnauer JE, Castiello S, Tsvetkov A, Li A, Alcaide DM, & Miller RR (in press). More frequent, shorter trials enhance acquisition in a training session: There is a free lunch! Journal of Experimental Psychology: General.

Nasser HM, & Delamater AR (2016). The determining conditions for Pavlovian learning. In Murphy RA, & Honey RC (Eds) The Wiley handbook on the cognitive neuroscience of learning, 5–46. Wiley Blackwell, USA.

Nieder A, Wagener L, & Rinnert P (2020). A neural correlate of sensory consciousness in a corvid bird. Science, 369(6511), 1626–1629. 10.1126/science.abb1447 [PubMed: 32973028]

Nieuwenstein MR, Potter MC, & Theeuwes J (2009). Unmasking the attentional blink. Journal of Experimental Psychology: Human perception and performance, 35(1), 159–169. 10.1037/0096-1523.35.1.159 [PubMed: 19170478]

Newman JP, Wolff WT, & Hearst E (1980). The feature-positive effect in adult human subjects. Journal of Experimental Psychology: Human Learning and Memory, 6(5), 630–650. 10.1037/0278-7393.6.5.630 [PubMed: 7437120]

Lattal KM (1999). Trial and intertrial durations in Pavlovian conditioning: Issues of learning and performance. Journal of Experimental Psychology: Animal Behavior Processes, 25(4), 433–450. 10.1037/0097-7403.25.4.433 [PubMed: 17763570]

Luzardo A, Alonso E, & Mondragón E (2017). A Rescorla-Wagner drift-diffusion model of conditioning and timing. PLOS Computational Biology, 13(11), e1005796. 10.1371/journal.pcbi.1005796 [PubMed: 29095819]

Ophir EA, Sherman E, & Lamy D (2020). An attentional blink in the absence of spatial attention: a cost of awareness?. Psychological Research, 84(4), 1039–1055. 10.1007/s00426-018-1100-x [PubMed: 30264360]

Pashler H, Rohrer D, Cepeda NJ, & Carpenter SK (2007). Enhancing learning and retarding forgetting: Choices and consequences. Psychonomic bulletin & review, 14(2), 187–193. 10.3758/BF03194050 [PubMed: 17694899]

Perales JC, Catena A, Shanks DR, & González JA (2005). Dissociation Between Judgments and Outcome-Expectancy Measures in Covariation Learning: A Signal Detection Theory Approach. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(5), 1105–1120. 10.1037/0278-7393.31.5.1105 [PubMed: 16248753]

Psychology Software Tools, Inc. [E-Prime 2.0]. (2016). Retrieved from https://support.pstnet.com/.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rescorla RA (1967). Pavlovian conditioning and its proper control procedures. Psychological Review, 74(1), 71–80. 10.1037/h0024109 [PubMed: 5341445]

Rescorla RA, & Wagner AR (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black AH, & Prokasy WF (Eds.), Classical Conditioning II: Current Research and Theory (pp. 64–99). New York: Appleton Century Crofts.

Roediger III HL, & Butler AC (2011). The critical role of retrieval practice in long-term retention. Trends in Cognitive Sciences, 15(1), 20–27. 10.1016/j.tics.2010.09.003 [PubMed: 20951630]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Rosas JM, Vila NJ, Lugo M, & López L (2001). Combined effect of context change and retention interval on interference in causality judgments. Journal of Experimental Psychology: Animal Behavior Processes, 27(2), 153–164. 10.1037/0097-7403.27.2.153 [PubMed: 11296490]

Sainsbury R (1971). The “feature positive effect” and simultaneous discrimination learning. Journal of Experimental Child Psychology, 11(3), 347–356. 10.1016/0022-0965(71)90039-7

Savastano HI, & Miller RR (1998). Time as content in Pavlovian conditioning. Behavioural Processes, 44(2), 147–162. 10.1016/s0376-6357(98)00046-1 [PubMed: 24896972]

Schad DJ, Vasishth S, Hohenstein S, & Kliegl R (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. Journal of Memory and Language, 110, 104038. 10.1016/j.jml.2019.104038

Shanks DR (1995). The psychology of associative learning (Vol. 13). Cambridge University Press, UK. 10.1017/CBO9780511623288

Siegel S, Allan LG, Hannah SD, & Crump MJC (2009). Applying signal detection theory to contingency assessment. Comparative Cognition & Behavior Reviews, 4, 116–134. 10.3819/ccbr.2009.40012

Smith TA, & Kimball DR (2010). Learning from feedback: Spacing and the delay–retention effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(1), 80–95. 10.1037/a0017407 [PubMed: 20053046]

Soares JS, Polack CW, & Miller RR (2016). Retrieval-induced versus context-induced forgetting: Does retrieval-induced forgetting depend on context shifts? Journal of Experimental Psychology: Learning, Memory, and Cognition, 42(3), 366–378. 10.1037/xlm0000171 [PubMed: 26389628]

Sperber D, Cara F, & Girotto V (1995). Relevance theory explains the selection task. Cognition, 57(1), 31–95. 10.1016/0010-0277(95)00666-M [PubMed: 7587018]

Stout SC, Amundson JC, & Miller RR (2005). Trial order and retention interval in human predictive judgment. Memory & Cognition, 33(8), 1368–1376. 10.3758/BF03193369 [PubMed: 16615384]

Stout SC, & Miller RR (2007). Sometimes competing retrieval (SOCR): A formalization of the comparator hypothesis. Psychological Review, 114(3), 759–783. 10.1037/0033-295X.114.3.759 [PubMed: 17638505]

Urcelay GP, Wheeler DS, & Miller RR (2009). Spacing extinction trials alleviates renewal and spontaneous recovery. Learning & Behavior, 37(1), 60–73. 10.3758/LB.37.1.60 [PubMed: 19122053]

Urcelay GP, & Miller RR (2014). The functions of contexts in associative learning. Behavioural processes, 104, 2–12. 10.1016/j.beproc.2014.02.008 [PubMed: 24614400]

Urcelay GP, Witnauer JE, & Miller RR (2012). The dual role of the context in postpeak performance decrements resulting from extended training. Learning & Behavior, 40(4), 476–493. 10.3758/s13420-012-0068-4 [PubMed: 22359182]

Vallée-Tourangeau F, Murphy RA & Baker AG (2005). Contiguity and the outcome density bias in action-outcome contingency judgments, Quarterly Journal of Experimental Psychology: Comparative and Physiological, 58B, 177–192. 10.1080/02724990444000104

Vallée-Tourangeau F, Murphy RA, Drew S & Baker AG (1998). Judging the importance of constant and variable candidate causes: A test of the Power PC theory. Quarterly Journal of Experimental Psychology, 51A, 65–84 10.1080/027249898391765

Vlach HA, & Sandhofer CM (2012). Distributing learning over time: The spacing effect in children’s acquisition and generalization of science concepts. Child Development, 83(4), 1137–1144. 10.1111/j.1467-8624.2012.01781.x [PubMed: 22616822]

Wagenmakers EJ (2007). A practical solution to the pervasive problems of p values. Psychonomic Bulletin & Review, 14(5), 779–804. 10.3758/BF03194105 [PubMed: 18087943]

Ward RD, Gallistel CR, & Balsam PD (2013). It’s the information!. Behavioural Processes, 95, 3–7. 10.1016/j.beproc.2013.01.005 [PubMed: 23384660]

Wason PC (1968). Reasoning about a rule. Quarterly Journal of Experimental Psychology, 20(3), 273–281. 10.1080/14640746808400161 [PubMed: 5683766]

Wasserman EA, Elek SM, Chatlosh DL, & Baker AG (1993). Rating causal relations: Role of probability in judgments of response-outcome contingency. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(1), 174–188. 10.1037/0278-7393.19.1.174

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Waugh NC, & Norman DA (1965). Primary memory. Psychological Review, 72(2), 89–104. 10.1037/h0021797 [PubMed: 14282677]

White PA (2004). Causal judgment from contingency information: A systematic test of the pCI rule. Memory & Cognition, 32(3), 353–368. 10.3758/BF03195830 [PubMed: 15285120]

Wickham H (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1–20. URL http://www.jstatsoft.org/v21/i12/.

Wickham H, François R, Henry L, and Müller K, (2019). dplyr: A Grammar of Data Manipulation. R package version 0.8.0.1. https://CRAN.R-project.org/package=dplyr

Wickham H, (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Young DB, & Pearce JM (1984). The influence of generalization decrement on the outcome of a feature-positive discrimination. Quarterly Journal of Experimental Psychology, 36(4), 331–352. 10.1080/14640748408402212

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

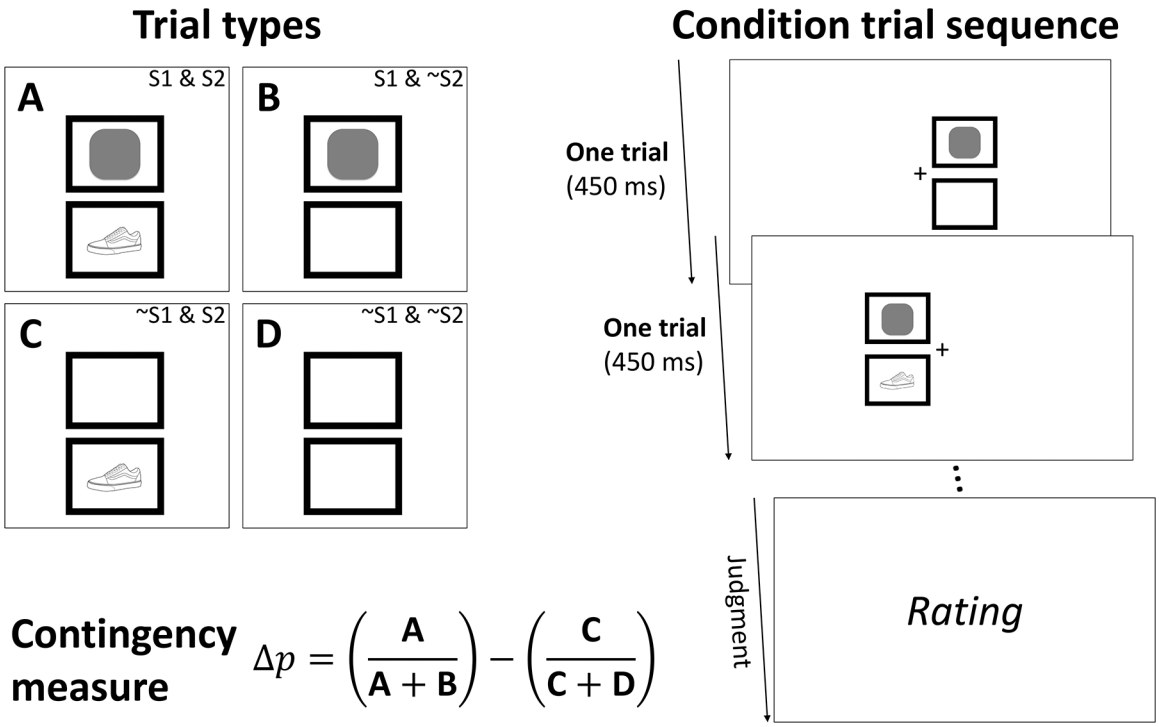

Figure 1.

Trial types. There are four possible combinations between shapes (S1) and drawings (S2) that were used as stimuli in the three experiments; ~ indicates *not* (or absence of). For details see Supplementary Materials. Trial sequence. This is an example of the first two trials in one condition. The stimuli in each trial were presented on the left and right of the fixation cross, alternating in consecutive trials to explicitly mark different trials. Black squares or trial markers (TM) were used to mark trials and their context. Contingency measure. Delta p (Δp) is a measure of continency (Allan, 1980) defined by the difference in conditional probabilities representing the four event types.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

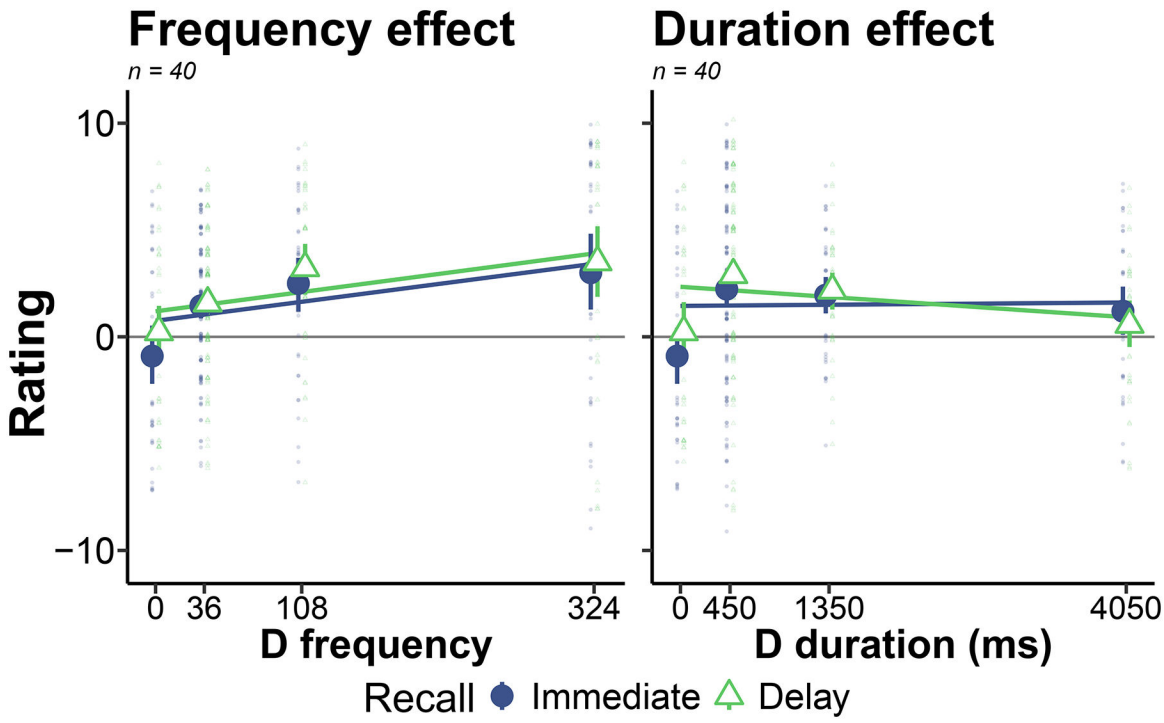

Figure 2.

Results from Experiment 1. Ratings at the end of conditions as a function of co-absent (D trials) trial frequency (left panel) and duration (right panel). Error bars are 95% confidence interval (CI) with a bootstrap method (used by ggplot based on Hmisc package; Harrell et al., 2019).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

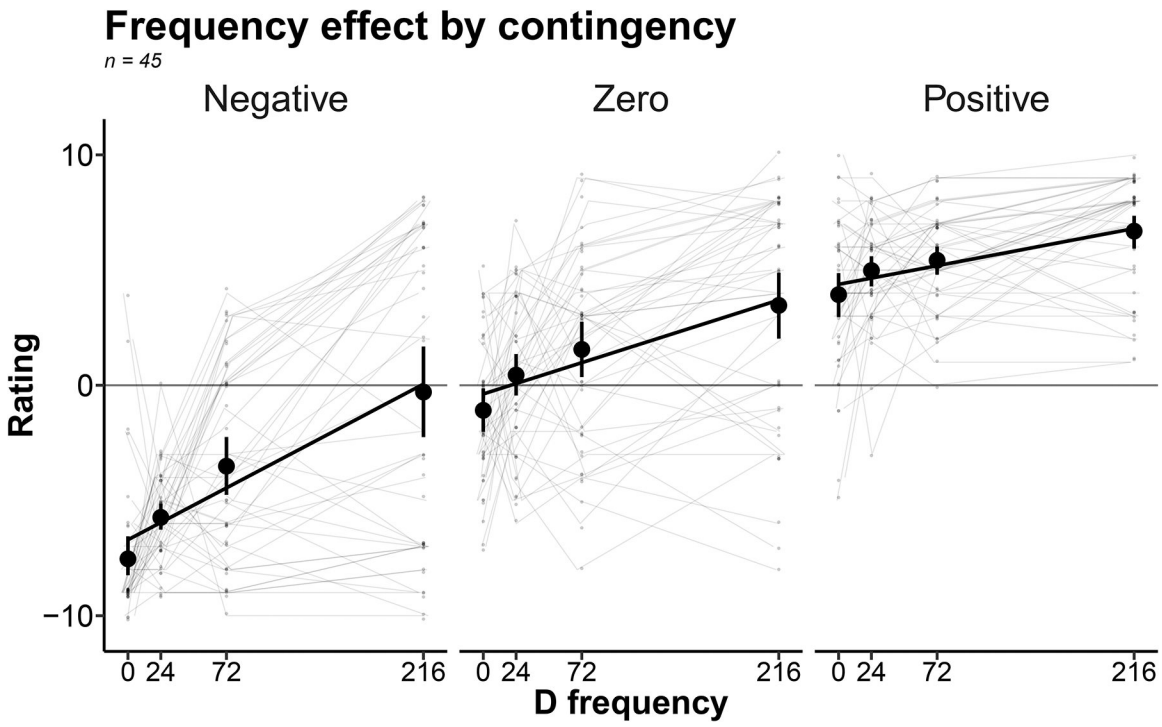

Figure 3.

Results from Experiment 2. Ratings at the end of conditions as a function of co-absent (D trials) trial frequency and contingency type (Negative, Zero, and Positive). Error bars are 95% confidence interval (CI) with a bootstrap method (same as Figure 2). Individual light grey lines represent participants.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

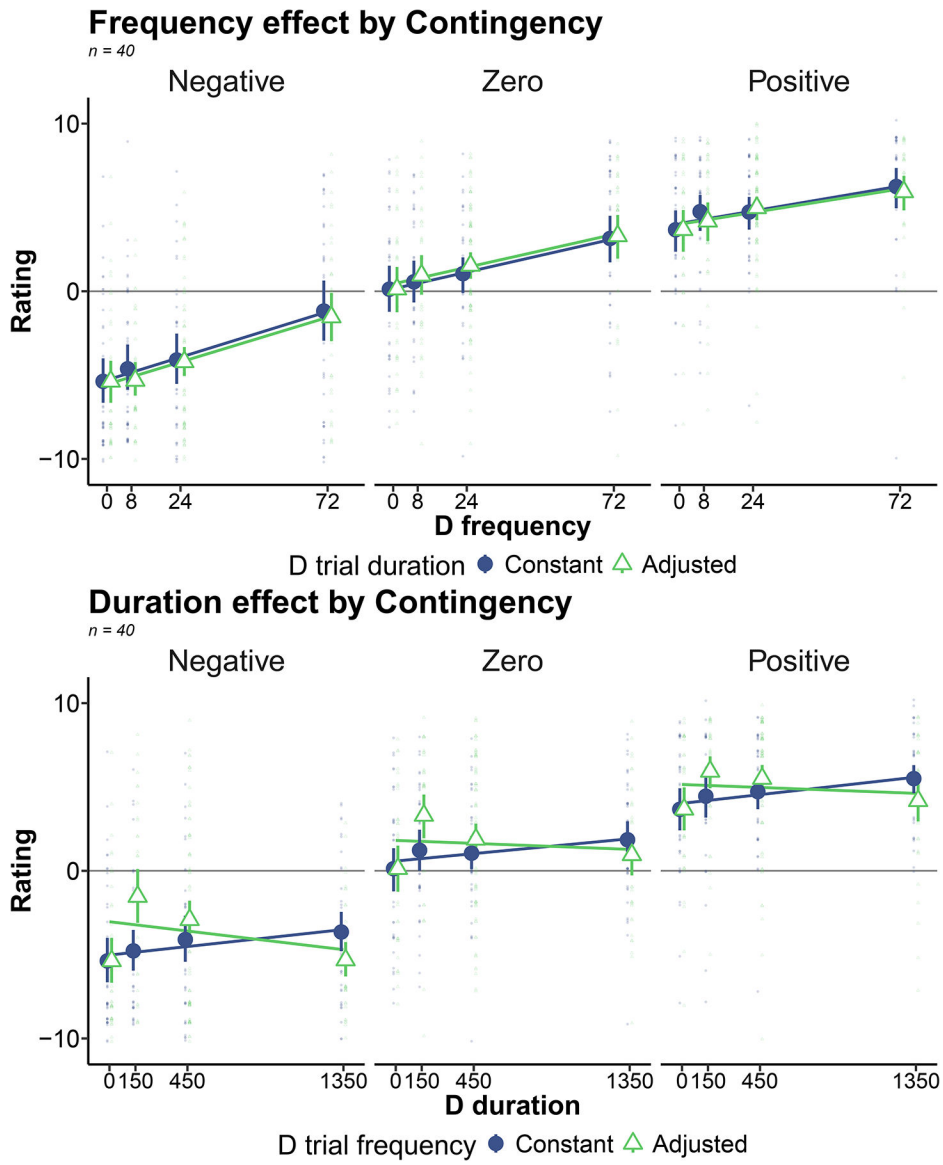

Figure 4.

Results from Experiment 3. For the upper panel, the data are split by D trial duration being inversely adjusted with respect to D trial frequency and D trial duration being held constant at 450 ms. For the lower panel, the data are split by D trial frequency being inversely adjusted with respect to D trial duration and a constant 24 D trial frequency. Error bars are 95% confidence interval (CI) with a bootstrap method (same as Figure 2 and 3). The adjusted 0 frequency and 0 duration conditions where duplicated for visualization purpose. For the analysis we perform both analyses, with duplicated 0, and removing 0 (see Supplementary Materials: Sensitivity analysis).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

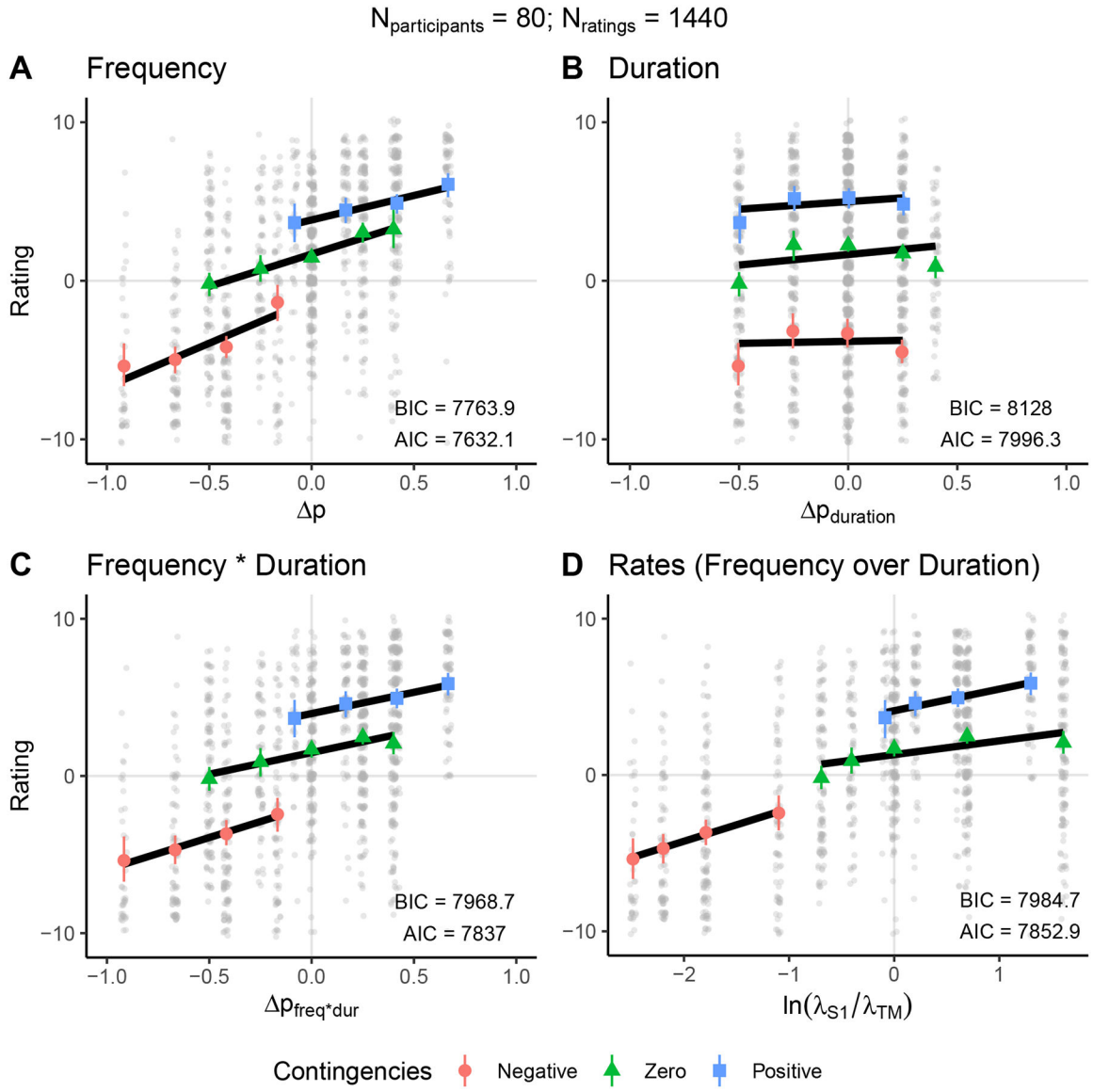

Figure 5.

Ratings for Experiments 1 and 3 pooled. **A** ratings as a function of Δp; **B** ratings as a function of $[A_{dur}/(A_{dur}+B_{dur})] - [C_{dur}/(C_{dur}+D_{dur})] = \Delta p_{duration}$; **C** ratings as a function of $[A_{freq*dur}/(A_{freq*dur}+B_{freq*dur})] - [C_{freq*dur}/(C_{freq*dur}+D_{freq*dur})] = \Delta p_{freq*dur}$; and **D** ratings as a function of natural logarithm of a rate ratio $[\ln(\lambda_{S1}/\lambda_{TM})]$. λ_{S1} is the rate of occurrence of S2 during S1 $[A_{freq} / ((A_{freq} * A_{dur}) + (B_{freq} * B_{dur}))]$; λ_{TM} is the rate of occurrence of S2 during only TM trials $[C_{freq} / [(C_{freq} * C_{dur}) + (D_{freq} * D_{dur}))]$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Experiment 1, Conditions and Δp values based on frequency (columns) and duration (rows) [Retention interval (2), Frequency (4), and Duration (4)]

Test	D trial Duration (ms)	D trial Frequency			
		0	36	108	324
Immediate	0	None −0.5			
	450		Baseline 0	More 0.25	Many 0.4
	1350		Long 0.25*		
	4050		Longer 0.4*		
Delay	0	None −0.5			
	450		Baseline 0	More 0.25	Many 0.4
	1350		Long 0.25*		
	4050		Longer 0.4*		

Note: Twelve different conditions with the number below representing Δp. Δp is conventionally based on trial frequency which is the basis of the calculation when frequency of D trials was varied (0, 36, 108, and 324).

*When D trial duration was varied (0 [same as no D trials: None], 450, 1350, and 4050 ms), Δp was calculated based on total duration of each trial type.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Experiment 1: Full model results

Variable	Estimate	CI_lower	CI_higher	*t*	*df*	*p*	*Std. Est.*
(Intercept)	1.84	0.95	2.73	4.06	466	<0.001	−0.05
delay	0.29	−0.31	0.88	0.94	466	ns	0.11
frequency	0.08	0.01	0.14	2.38	466	<0.05	0.23
duration	0.02	−0.01	0.06	1.26	466	ns	0.06
condition order	−0.01	−0.09	0.08	−0.16	466	ns	−0.01
delay × frequency	−0.01	−0.06	0.04	−0.39	466	ns	−0.03
delay × duration	−0.05	−0.09	0	−1.92	466	ns	−0.13

Note: Std. Est. = standardized estimate, Estimate = linear mixed model coefficient, CI_lower = lower confidence interval, CI_higher = higher confidence interval, *t* = t value, df = degree of freedom, *p* = p value.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Experiment 2: Conditions and Δp [Frequency (4), and baseline Contingency (3)]

Contingency and A, B, and C trials		**D trial Frequency**			
		None	Baseline	More	Many
		0	24	72	216
Positive	(44 A, 4 B, 24 C)	−0.1	0.42	0.67	0.82
Zero	(24 A, 24 B, 24 C)	−0.5	0	0.25	0.4
Negative	(4 A, 44 B, 24 C)	−0.9	−0.42	−0.17	−0.02

Note: Each square represents a condition, the number represents Δp.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4.

Experiment 2: Full model results

Variable	Estimate	CI_lower	CI_higher	*t*	*df*	*p*	*Std. Est.*
(Intercept)	1.25	0.45	2.05	3.06	527	<0.01	0
frequency	0.12	0.08	0.16	6.34	527	<0.001	0.3
contingency	4.78	4.34	5.21	21.65	527	<0.001	0.69
condition stage order	−0.28	−0.58	0.03	−1.77	527	ns	−0.04
frequency × contingency	−0.06	−0.08	−0.04	−5.82	527	<0.001	−0.12

Note: Std. Est. = standardized estimate, Estimate = linear mixed model coefficient, CI_lower = lower confidence interval, CI_higher = higher confidence interval, t = t value, df = degree of freedom, p = p value.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Experiment 3: Experimental conditions and Δp values [Contingency (3), Frequency (4), and Duration (4)]

Contingent	D trial Duration	D trial Frequency			
		0	8	24	72
Positive (44 A, 4 B, 24 C)	0	None −0.08			
	150			Short 0.17 *	More adjusted 0.67; 0.42 *
	450		Few 0.17	Baseline 0.42	More 0.67
	1350		Few adjusted 0.17; 0.42 *	Long 0.67 *	
Zero (24 A, 24 B, 24 C)	0	None −0.5			
	150			Short −0.25 *	More adjusted 0.25; 0 *
	450		Few −0.25	Baseline 0	More 0.25
	1350		Few adjusted −0.25; 0 *	Long 0.25 *	
Negative (4 A, 44 B, 24 C)	0	None −0.92			
	150			Short −0.67 *	More adjusted −0.17; −0.42 *
	450		Few −0.67	Baseline −0.42	More −0.17
	1350		Few adjusted −0.67; −0.42 *	Long −0.17 *	

Note: Each rectangle represents a condition, with the name of the condition and the number is the corresponding Δp.

* When D trial duration was varied (0 [same as no D trials: None], 150, 450, and 1350 ms), Δp was calculated based on total duration in each trial type.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 6.

Experiment 3: Full model results

Variable	Estimate	CI_lower	CI_higher	*t*	*df*	*p*	*Std. Est.*
(Intercept)	0.81	0	1.61	1.97	1059	<0.05	0
frequency	0.09	0.05	0.12	4.29	1059	<0.001	0.2
contingency	4.35	3.62	5.09	11.59	1059	<0.001	0.64
adjusted by duration	−0.03	−0.39	0.32	−0.19	1059	ns	−0.01
stage order	−0.03	−0.32	0.26	−0.2	1059	ns	0
frequency × contingency	−0.03	−0.05	0	−2.15	1059	<0.05	−0.05
frequency × adjusted by duration	0	−0.03	0.02	−0.22	1059	ns	−0.01
contingency × adjusted by duration	0.1	−0.33	0.54	0.46	1059	ns	0.02
triple interaction	0	−0.03	0.03	0.02	1059	ns	0

Note: Std. Est. = standardized estimate, Estimate = linear mixed model coefficient, CI_lower = lower confidence interval, CI_higher = higher confidence interval, *t* = t value, df = degree of freedom, *p* = p value.