Running head: CONTINGENCY ASSESSMENT AND SDT

Signal detection analysis of contingency assessment: Associative interference and nonreinforcement impact cue-outcome contingency sensitivity, whereas cue density affects bias

Jérémie Jozefowiez,

Université de Lille

Gonzalo P. Urcelay

University of Nottingham

&

Ralph. R. Miller

SUNY - Binghamton

Correspond with:
Dr. Jérémie Jozefowiez
Univ. Lille, CNRS UMR 9193 – SCALab – Sciences Cognitives et Affectives
F-59000 Lille
France
Email: jeremie.jozefowiez@univ-lille.fr
Phone: +33-320-041-6866
Fax: +33-320-041-6036

Submitted: May 2022

Abstract

In a signal detection theory (SDT) approach to associative learning, the perceived (i.e., subjective) contingency between a cue and an outcome is a random variable drawn from a Gaussian distribution. At the end of the sequence, participants report a positive cue-outcome contingency provided the subjective contingency is above some threshold. Some researchers have suggested that the mean of the subjective contingency distributions and the threshold are controlled by different variables. The present data provide empirical support for this claim. In three experiments, participants were exposed to rapid streams of trials at the end of which they had to indicate whether a target outcome O1 was more likely following a target cue X. Two interfering treatments were incorporated in some streams to impend participants' ability to identify the objective X-O1 contingency: interference trials (X was paired with an irrelevant outcome O2), nonreinforced trials (X was presented alone), plus control trials (an irrelevant cue W was paired with O2). Overall, both interference and nonreinforced trials impaired participants' sensitivity to the contingencies as measured by SDT's *d'*, but they also enhanced detection of positive contingencies through a cue density effect, with nonreinforced trials being more susceptible to this effect than interference trials. These results are explicable if one assumes interference and nonreinforced trials impact the mean of the associative strength distribution, while the cue density influences the threshold.

Keywords: contingency assessment, signal detection theory, streaming procedure, associative interference, cue density effect.

In Pavlovian conditioning, the pairing of an initially-neutral conditioned stimulus (CS) with a biologically-relevant unconditioned stimulus (US) causes the CS to subsequently trigger a conditioned response (CR). In a human contingency assessment task, participants are exposed to sequences of stimuli: pairings of a target cue (X) and a target outcome (O) within such sequences allow participants to predict the outcome on the basis of the cue. Both phenomena are thought to be the consequence of associative learning, that is, the formation of an association between the internal representations of the CS/cue and the US/outcome, allowing the former to activate the latter. Most of the variables that determine whether a CS will elicit a CR in Pavlovian conditioning also determine whether a cue will predict an outcome in contingency assessment. Among these variables is the objective predictive value of the CS/cue with respect to the US/outcome as measured by the ΔP metric of contingency. ΔP is the conditional probability of the US/outcome occurrence given the CS/cue minus its conditional probability in the absence of the CS/cue. When ΔP is positive (i.e., a positive contingency), the CS/cue is a genuine predictor of the US/outcome; when ΔP is negative (i.e., a negative contingency), the CS/cue is a genuine predictor of the absence of the US/outcome where it might otherwise be expected; finally, when ΔP is null (i.e., a null contingency), the CS/cue does not provide information regarding the occurrence or the absence of the US/outcome. As first demonstrated by Rescorla (1968), the likelihood that a CS will trigger a CR is directly related to the ΔP value between the CS and the US. Likewise, contingency judgments in human participants track the ΔP value between the cue and the outcome (Allan, 1980; see Shanks, 1995, 2007, for reviews).

Allan and collaborators have analyzed contingency assessment using a signal detection theory (SDT) framework (Allan, Hannah, Crump, & Siegel, 2008; Allan, Siegel, & Tangen, 2005; Siegel, Allan, Hannah, & Crump, 2009; see also Jozefowiez, 2021; Laux, Goedert, & Markman, 2010; Maia, Lefèvre, & Jozefowiez, 2018; Perales, Catena, Shanks, &

Gonzalez, 2005; see Wickens, 2002, for an introduction to SDT). Exposing participants to a sequence of trials with and without the cue and with and without the outcome presumably results in the development of a subjective contingency that reflects the strength of the cue-outcome association. If the subjective contingency is above a critical value, participants report is a positive contingency between the cue and the outcome; otherwise, they do not. The subjective contingency is ‘noisy;’ if a participant is exposed to two sequences of stimuli implementing the same cue-outcome contingency, the subjective contingency value at the end of those two sequences is not apt to be identical because, even though it would be drawn from the same Gaussian distribution, it is drawn anew for each stimulus sequence.

In such an SDT framework, participants belief states at the end of a stimulus sequence are the result of several potentially independent processes: (a) the sensitivity to the objective cue-outcome contingencies which determines how the mean of the subjective contingency distribution varies with the cue-outcome contingency; (b) the noise in the perception of cue-outcome contingencies which determines the standard deviation of the subjective contingency distributions; (c) the threshold value determining which of two dichotomous belief states participants are in (i.e., outcome expected or unexpected) as a function of the subjective contingency. Allan and her collaborators have argued that the sensitivity to the contingencies and the threshold value are affected by different variables (because they worked only within the framework of an SDT model assuming constant variance, they did not consider variables potentially affecting the noise in the subjective contingency). Notably, Hannah & Allan (2011), Laux et al. (2010), and Siegel, Allan, Hannah, & Crump (2009) have argued that the two conditional probabilities of ΔP do not act on the same process. They concluded that the probability of the outcome given the cue impacts the sensitivity to the contingencies, while the probability of the outcome in the absence of the cue impacts the threshold. Likewise, Allan, Siegel, and Tangen (2005) have argued that the outcome density effect (for a given ΔP value, participants are more likely to perceive a positive contingency between the cue and the outcome if the overall probability of the outcome is higher) is the result of changes in the threshold, rather than in the sensitivity to the contingencies. Likewise, Perales et al. (2005) have made a similar argument concerning the cue density effect (for a given ΔP, participants are more likely to perceive a positive contingency between the cue and the outcome if the overall probability of the cue is higher, Allan & Jenkins, 1993; Matute, Yarritu, & Vadillo, 2011; Vadillo, Musca, Blanco, & Matute, 2011; Wasserman, Kao, Van Hamme, Kategiri, & Young, 1996; White, 2003).

Unfortunately, the conclusion that different variables impact the sensitivity to the contingencies and the threshold is not entirely warranted by the data collected by these authors. As discussed in Maia et al. (2018), their reasoning only holds if one knows whether an experimental manipulation has altered the subjective contingency distributions or the threshold. But the parameters of an SDT model extracted from a data set are not absolute measurements; their values are expressed only relative to each other. For instance, suppose SDT is used to model an experiment in which participants have to discriminate between two cue-outcome contingencies C1 and C2, both between cue A and outcome O. The parameters of the model are the mean and standard deviation of the subjective contingency distribution when C1 (or C2) is presented, and the threshold value. To extract the SDT parameters, the mean and standard deviation of the subjective distribution for either C1 or C2 are clamped to arbitrary values (usually 0 for the mean and 1 for the standard deviation), and the values of the other parameters are then expressed relative to them. Hence, depending on whether C1 or C2 is used as the reference condition, the absolute values of the SDT parameters will change, but their values relative to each other will not.

As a consequence, it is impossible in many situations to know whether it is the threshold value which has shifted while the mean of the distributions remained constant, or the reverse (e.g., Witt, Taylor, Sugovic, & Wixted, 2015; Wixted & Stretch, 2000). For instance, Laux et al. (2010) manipulated the contingency between a target cue and the outcome, which could be either null or positive, and between a competing cue and the outcome, which could also be either null or positive. They found that their data could be accounted for by an SDT model assuming that the mean of the subjective contingency distribution was determined by the contingency between the target cue and the outcome, while the threshold value was controlled by the contingency between the alternate irrelevant cue and the outcome. But Maia et al. (2018) showed that this SDT model was mathematically equivalent to an SDT model assuming that both probabilities jointly affected the mean of the subjective contingency distributions while the threshold remained constant. Hence, it was not possible to distinguish between these two models from the available data. The same problem affects similar arguments made by Hannah and Allan (2011) and Siegel et al. (2009). For the same reason, the conclusions of Allan et al. (2005) and Perales et al. (2005) regarding the cue and outcome density effects are not unambiguously supported by their data.

Hence, while the idea that the sensitivity to the contingencies and the threshold are impacted by different variables is intriguing, it is not well supported by the published data. We believe that the evidence presented in this paper provide such empirical support. Yet, this was an unexpected outcome for us. The first experiment in this series was designed with a different goal in mind. It led to a surprising result that was difficult to anticipate and consequently we decided to investigate in its own right. At the conclusion of the series, we realized that, if analyzed through the lens of SDT, our results could be explained by concluding that the sensitivity to the contingencies and the threshold are affected by different variables. Instead of inventing a *post hoc* rational for Experiment 1 in the light of this conclusion, we decided it was preferable to acknowledge the serendipitous nature of our discovery and to lead the reader through the same reasoning we followed during the experiments.

Experiment 1

Associative interference (outcome interference, as opposed to cue interference) occurs whenever the representation of a cue X becomes associated with two different outcome representations, O1 and O2. In retroactive interference, the interfering X-O2 association is learned second (i.e., after the X-O1 association), whereas, in proactive interference, it is learned first. Following this sort of categorization, there is a third type of interference which is curiously rarely mentioned: interspersed interference in which the X-O1 and the X-O2 pairings occur interspersed and subsequently interfere with expression of the opposing association (Pineño, Ortega, & Matute, 2000; Polack, Jozefowiez, & Miller, 2017).

As the first step in a larger project, Experiment 1 was aimed at simply identifying parameters that would produce robust interspersed interference in a contingency assessment task. We used the streaming procedure initially developed by Allan and collaborators (a.k.a. the streamed-trial procedure: Crump, Hannah, Allan, & Hord, 2007; Hannah, Crump, Allan, & Siegel, 2009; Siegel, Allan, Hannah, & Crump, 2009; see also Jozefowiez, 2021; Laux et al., 2010; Maia et al., 2018). In this procedure, participants are exposed to very rapid streams of stimuli into which a target cue and a target outcome are repeatedly embedded. At the end of each stream, participants have to indicate whether the target outcome is more likely following the target cue, using either a Yes/No question or a Likert rating. One advantage of this procedure is that, as streams are presented multiple times to a single participant, it provides enough data to extract the parameters of a SDT model (see, for instance, Jozefowiez, 2021, or Maia et al., 2018).

Method

Participants and apparatus

We hypothesized a Cohen's *d* of 0.6. In a worst-case-scenario where performance in the various experimental conditions is not correlated, this would suggest that 45 participants are necessary to allow a *t*-test to detect an effect corresponding to a Cohen's *d* of 0.6 80% of the time. Thus, we planned on a sample size of 45 participants (10 males and 35 females, with mean age 20.28 +/- 1.48, ranging from 19 to 25 years old) recruited from the subject pool at the University of Lille. The experiment took place in individual experimental cubicles using a MacBook Pros with 15.4-inch screens with a 2880 x 1800 resolution. The program for the study was written using the Psychopy2 library (Peirce, 2007).

Procedure

After reading a screen describing the experiment, participants signed a consent form indicating that their data would be anonymized and that they could withdraw their consent to participate at any time without prejudice. The procedure generally followed that of Jozefowiez (2021) and Maia et al. (2018). The experiment was divided into two parts. The first part was a warm-up aimed at familiarizing participants with the task and at screening out participants who had difficulties discriminating contingencies. This warm-up phase was itself divided into three parts: positive warm-up, negative warm-up, and mixed warm-up. The second part was the experimental phase per se.

Positive warm-up: At the beginning of the positive warm-up procedure, participants were first presented with instructions explaining them the task. They were also reminded to turn off their cellphone and asked to not count the stimuli. They were then presented with a stimulus stream. The stream started with a 1-s grey screen with a white fixation cross at the center. The fixation cross remained visible throughout the stream. A series of 20 trials was then presented to participants. A trial was composed of (a) an 83-ms cue epoch during which a white triangle (trial marker A, h x w was 105 x 105 pixels) was shown in the upper left corner of the screen (coordinates in pixels were [-225, 170], with [0, 0] being the location of the fixation cross) while another stimulus (a green oval, radii was 90 x 78 pixels, target stimulus X) was shown during only some trials in the upper right corner (coordinates in pixels were [225, 196]; (b) then an 83-ms outcome epoch during which a red rectangle (181 x 158 pixels, target outcome O1) would or would not be presented centered in the lower part of the screen, below the fixation cross (coordinates in pixels [0, -196]); (c) finally, there was an 83-ms inter-trial interval (ITI) during which only the fixation cross was presented.

At the end of the stream, a dialog box appeared at the center of the screen with the question: "Does the red square appear more often when the green circle has just been presented than if the green circle has not presented?". Instead of the words “green circle” and “red square”, participants saw small iconic representations of the stimuli as they appeared in the experiment. Participants used the mouse to answer by clicking on either a "Yes" or a "No" button displayed below the question. The mouse pointer initially appeared equidistant from each button. Once they clicked on one of the buttons, the dialog box disappeared and another one appeared with the question “How sure are you of your decision?” with three buttons reading “Not sure”, “Sure”, “Very sure” appearing below the question. Once participants answered this question, the dialog box disappeared, and the stream ended. This confidence rating was introduced because it would have allowed us, if needed, to use a more complex SDT model assuming that the standard deviation of the subjective contingency distribution varied across conditions (see, for instance, Jozefowiez, 2021, & Maia et al., 2018). This was introduced as an option that we could have used in case we needed it, that is to say, in case the simpler equal variance SDT model did not provide a clear account of the data. As that did not turn out to be the case, we decided not to exercise this option. Jozefowiez (2021) and Maia et al. (2018) have, in any case, showed that, for the range of ΔP values we explored, the assumption that the standard deviation of the subjective contingency distribution is reasonable.

Overall, streams were composed of the following types of trials: (a) AX-O1 trials: A and X followed by O1; (b) AX- trials: A and X followed by no outcome; (c) A-O1 trials: Only A is shown followed by O1; (d) A- trials: Only A is shown followed by no outcome. The ΔP contingency between X and O1 was manipulated by varying the proportion of these four types of trials within a stream.

The trial composition for the stream shown during the positive warm-up is shown in Table 1. It implements a strong positive contingency (ΔP = 1) between the cue and the outcome. The order of presentation of the trials within the stream was determined randomly. Once the stream was over, participants were presented with a debriefing screen explaining that, as O1 was always presented following X while it was never presented when X was not presented, they should have answered YES to the contingency question. They were then prompted to press a button to move to the next part of the warm-up phase.

Negative warm-up. During the negative warm-up, participants were presented with a stream identical to the one used during the positive warm-up except for the composition of trials which is shown in Table 1. It implements a strong negative contingency (ΔP = -1) between the cue and the outcome. At end of the stream, participants were presented with a debriefing screen explaining that, as O1 was never presented following X while it was always presented when X was not presented, they should have answer NO to the contingency question. They were then prompted to press a button to move to the next part of the warm-up phase.

Mixed warm-up: Participants were given instructions explaining that they would now be presented with positive contingency and negative contingency streams and that they should try to identify them as accurately as possible. They were then presented with a block of two streams. The block was composed of a ΔP = 1 stream similar to the one shown during the positive warm-up and a ΔP = -1 stream similar to the one shown during the negative warm-up. The order of presentation of the streams within this block was determined randomly. In order to move to the experimental phase, participant had to answer correctly to two blocks in a row. A correct answer was defined as answering YES to the contingency question after a $\Delta P = 1$ stream and NO after a $\Delta P = -1$ stream. If participants did not meet the learning criterion after 10 block presentations, they were rejected from the study; a debriefing screen appeared thanking them for their participation and explained the overall goal of the experiment. In practice, this did not occur for any participants.

Experimental phase. At the beginning of the experimental phase, participants were shown instructions telling them that they were about to enter the experimental phase and that they would be presented with streams for which it would more difficult than before to decide whether they should answer YES or NO to the contingency question. They were also reminded that they should try not to count the stimuli.

During the experimental phase, participants were presented with streams identical to the one shown during the warm-up phase except that (a) they were composed of 40 trials instead of 20 and (b) two new stimuli were presented during the streams: a yellow trapezoid (95 x 105 pixels, distractor cue W) appearing at the same location as the target cue X and a blue rhombus (95 x 105 pixel, interfering outcome O2) appearing at the same location as the target outcome O1. This led to the introduction of two new kinds of trials: AW-O2 trials (A and W were shown during the cue epoch while O2 was shown during the outcome epoch phase) and AX-O2 trials (A an X were shown during the cue epoch while O2 was shown during the outcome epoch). As during the warm-up, at the end of each stream participants completed a contingency question followed by a confidence rating.

The compositions of the four types of streams participants were exposed to during the experimental phase is shown in Table 2. Each stream was composed of a set of ’core’ trials which established a baseline contingency between X and O1 (either negative or positive) and the ’interfering treatment’ aimed at impeding the learning or expression of the X-O1 association established by the core trials. Although in Table 2 the core trials are listed separately for clarity, the trials composing the core trials and the interfering trials were randomly mixed during the presentation of a stream. The interfering trials consisted of 20 AW-O2 trials in the Control condition, and of 20 AX-O2 trials in the Interference condition. Each stream was presented 40 times to each participant. The order of presentation of the streams was randomly determined.

Data analysis

In addition to the percentage of each type of trial for which participants identified a positive contingency, we computed $d'(i)$ for the streams in the Control and Interference condition which represented each participant’s sensitivity to contingency in condition i ($i \in [Control, Interference]$), that is, the participant’s ability to discriminate between a positive and a negative contingencies in condition i. This relies on the following assumptions. When presented with a negative stream in condition i, the subjective contingency is drawn from a Gaussian distribution with mean $\mu_N(i)$ and standard deviation σ, whereas, for a positive contingency, it is drawn from a Gaussian distribution with mean $\mu_P(i)$ and standard deviation σ (see General Discussion for elaboration) When the subjective contingency is above a threshold value $C(i)$, the participant believes that there is a positive contingency between the cue and the outcome.

Under these assumptions, $d'(i)$ is the difference $\mu_P(i) - \mu_N(i)$. Let $P_i(+|Neg)$ be the probability of detecting a positive contingency after a negative stream in condition i and $P_i(+|Pos)$ be the probability of detecting a positive contingency after a positive stream in condition i. Then,

$$d'(i) = Z[P_i(+|Pos)] - Z[P_i(+|Neg)] \qquad (1)$$

where $Z(x)$ is the inverse of the cumulative standard distribution (Wickens, 2002). Note that *d'(i)* cannot be computed if either $P_i(+|Neg)$ or $P_i(+|Pos)$ is equal to 0 or 1. This happened for just one participant in the Control condition: whenever a comparison involved the Control condition, the data from this participant were excluded from the analysis.

$P_i(+|Neg)$ is what is usually called the probability of false alarms for condition *i* whereas $P_i(+|Pos)$ is the hit rate for condition i. This entails that the subjective contingency distribution following a negative stream is the noise distribution for condition *i*, whereas the subjective contingency distribution following a positive stream is the signal distribution for condition *i*. Those terms imply that all the SDT parameters are expressed relative to $\mu_N(i)$.

Inferential analysis was carried out based on within-subject 95% confidence intervals (CI) computed using Student's *t*-distribution. Cohen's *d* was used as a measure of effect size between two conditions. Following Cummings' (2012) recommendation for the computation of Cohen's *d* in a within-subject design, we used the following formula (also called Hedge's g)

$$d = \left[1 - \frac{3}{8(n-1)-1}\right] \frac{\sum_{j=1}^{n}(x_{k1} - x_{k2})}{n\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$

where x_{kj} is the score of participant *k* in condition *j*, and *n* is the number of participants. 95% CI for Cohen's *d* were computed using ESCI (https://thenewstatistics.com/itns/esci/) which implements the method described by Algina and Kesselman (2003). It provides an approximation of the 95% CI for Cohen's *d* in a paired design if (a) the number of participants is larger than 10, (b) Cohen's *d* in the population is between -1.8 and +1.8, and (c) the correlation in the population between the two conditions is between 0 and 0.8. The linear correlation between x_{k1} and x_{k2} is also provided because it can be used to inform future power analysis.

Results and discussion

As shown in the top panel of Figure 1, participants discriminated between the positive and the negative streams in both the Control and the Interference condition (Control: Cohen's *d* = 1.89, no CI could be computed because the convergence properties of the Algina-Kasselman method were not met, correlation = 0.61; Interference: Cohen's *d* = 0.80, 95% CI [0.54, 1.09], correlation = 0.73). This is confirmed by the bottom panel of Figure 1 which shows participants' sensitivity to the contingencies as measured by *d'*. It is clearly positive in both conditions. Moreover, as one would expect as a result of associative interference, the sensitivity to the contingency is lower in the Interference condition than in the Control one (Cohen's *d* = 0.93, 95% CI [0.58, 1.30], correlation = 0.49).

But, contrary to expectations, participants were also more likely to detect a positive contingency in the Interference condition than in the Control condition (Cohen's *d* = 0.90, 95% CI [0.44, 1.38], correlation = -0.10). This is true for both the negative streams (Cohen's *d* = 1.20, 95% CI [0.74, 1.70], correlation = 0.04) and the positive ones (Cohen's *d* = 0.44, 95% CI [0.01, 0.88], correlation = -0.09). At first glance, there are two potential explanations for this puzzling result. As X was presented 20 more times in the Interference condition than the Control condition, this could be the result of a cue density effect. Another possibility is that participants did not discriminate between O1 and O2, perhaps because the stimuli were so short. Indeed, if we assume that O1 and O2 are processed as the same outcome and we then compute ΔP for each stream, the various conditions are ranked as follows from the higher ΔP to the lowest: Interference-Positive, Control-Positive, Interference-Negative, Control-Positive. If we rank the conditions according to the empirical probability of detecting a positive contingency, we obtain exactly the same order.

Experiment 2

Experiment 2 was designed to differentiate between these two possibilities (i.e., cue density effect vs. outcome discrimination failure). It basically replicates Experiment 1 but adds a Nonreinforcement control in which the AX-O2 interference trials are replaced by AX- nonreinforced trials. If the cue density explanation is correct, the same pattern observed in the Interference condition should also be observed in the Nonreinforcement condition because the cue density is the same in both conditions. The sensitivity to the contingencies would likely be lower because of the presentation of AX- nonreinforced trials, but participants should be biased toward detecting a positive contingency. Conversely, if the outcome discrimination failure hypothesis is correct, the overall probability of detecting a positive contingency should be lower, or at most identical, in the Nonreinforcement condition than in the Control condition.

Method

Participants and apparatus

The target sample size was 64. As the effect size observed in Experiment 1 was large, a lower number of participants in theory should have sufficed. But, as the number of interference AX-O2 trials was decreased from 20 in Experiment 1 to 10 in Experiment 2 for methodological reasons that are explained in the next section, and because we did not know how this would affect the effect size, we aimed at 80% power for a *t*-test under the assumption that interference would correspond to a Cohen's *d* of 0.5 and that performance in the various conditions would not be correlated within subjects.

We ended up recruiting overall 69 participants from the University of Lille subject pool, but 5 of them failed the warm-up phase. The remaining 64 participants were composed of 55 females and 9 males (age: 21.06 +/- 2.73, ranging from 18 years old to 33 years old). The apparatus was identical to the one used in Experiment 1.

Procedure

The procedure was the same as that used in Experiment 1 except for the composition of the streams during the experimental phase which is depicted in Table 3. Note that the 20 interfering trials in the Control condition in Experiment 2 consisted of 10 AW-O2 trials and 10 AW- trials (in AW- trials, A and W were shown during the cue epoch whereas no stimulus was shown during the outcome epoch). The 10 AW-O2 trials were replaced by 10 interference AX-O2 trials in the Interference condition and by 10 AX- trials in the Nonreinforcement condition. This design has the advantage to keep the O2 density constant across conditions. This would not have been the case if, as in Experiment 1, we had 20 AW-O2 trials in the Control condition and 20 AX-O2 trials in the Interference conditions. That is, replacing the 20 AX-O2 and AW-O2 trials by 20 AX- trials in the Nonreinforcement condition would have created a potential confound because O2 would never have been shown in the Nonreinforcement streams.

Note that, in Table 3, the 10 AX- nonreinforced trials are shown separately in the Nonreinforcement condition for convenience, but they were not different from the AX- trials which were part of the ‘core’ trials that determined the baseline contingency between X and O1. Data analysis proceeded as in Experiment 1. $d'(i)$ $(i \in [Control, Interference, Nonreinforcement])$ was computed for each participant across condition.

Results and discussion

As in Experiment 1, participants had no difficulty discriminating between the positive and negative streams in all conditions, as evidenced both by the probability of detecting a positive contingency (Figure 2, upper panel. Control: Cohen’s *d* = 2.05, no CI could be computed, correlation = 0.03; Interference: Cohen’s *d* = 1.50, 95% CI [1.15, 1.87], correlation = 0.51; Nonreinforcement: Cohen’s *d* = 1.10, 95% CI [0.82, 1.40], correlation = 0.62), and the sensitivity to the contingencies (Figure 2, lower panel). Moreover, again as in Experiment 1, the sensitivity to the contingencies was lower in the Interference condition than in the Control condition (Cohen’s *d* = 0.35, 95% CI [0.07, 0.62], correlation = 0.41). This was also the case for the Nonreinforcement condition (Cohen's *d* = 0.42, 95% CI [0.15, 0.70], correlation = 0.41). The effect size for the effect of the interference trials on the sensitivity to the contingencies was weaker than in Experiment 1, which probably reflects the number of interfering trials having been reduced from 20 to 10. The experiment overall shows that performance is remarkably resistant to interference. Notably, one might have thought that 10 additional AX- trials would have had a larger impact on the sensitivity to the contingencies. This could indicate that, as already established by previous studies (i.e., Murphy et al., in press; Perales et al., 2005), AX-O1 trials weight more in contingency judgements than do AX- trials. We have evidence from unpublished studies that performance in this version of the streaming procedure is under the control of ΔP and not merely of $P(O1|AX)$.

There was no appreciable difference between Nonreinforcement and Interference in sensitivity to the contingency as measured by *d'* (Cohen's *d* = 0.09, 95% CI [-0.17, 0.35], correlation = 0.43). This could indicate that either the AX-O2 interference and the AX- nonreinforced trials were processed the same way or that the experiment lacked the statistical power necessary to detect a difference between the two interfering treatments. If the latter is correct, the differences between the two treatments were small, which is line with the conclusions reached by Jozefowiez et al. (2020) in which they compared extinction and counterconditioning with a neutral nontarget outcome.

The critical question is whether participants were more likely to detect a positive contingency in the Nonreinforcement condition than the Control condition regardless of the objective cue-outcome contingency because this is the distinction between the cue density effect hypothesis and the outcome discrimination failure hypothesis. Compared to participants in the Control condition, those in the Nonreinforcement condition were more likely to report that the outcome was more likely to appear if the cue had been presented than if it had not (Cohen's *d* = 1.11, 95% CI [0.80, 1.45], correlation = 0.46, see top panel of

Figure 2), a conclusion which holds not only for the positive streams (Cohen's *d* = 0.66, 95% CI [0.39, 0.93], correlation = 0.52) but the negative ones (Cohen's d = 1.18, 95% CI [0.84, 1.53], correlation = 0.37). This refutes the outcome discrimination failure hypothesis and points toward a cue density effect.

The problem with a cue density account is that a cue density effect was not observed in the Interference condition despite the fact that the Interference and Nonreinforcement conditions had the same cue density. As far as the overall probability of detecting a positive contingency is concerned, there was no appreciable difference between the Control and the Interference conditions (Cohen's *d* = 0.12, 95% CI [-0.15, 0.39], correlation = 0.41). This is also the case for the probability to detect a positive contingency after a positive stream (Cohen's *d* = -0.12, 95% CI [-0.39, 0.15], correlation = 0.40). It is only after a negative stream that participants were more likely to detect a positive contingency in the Interference condition than in the Control condition. However, this effect is much weaker than the similar effect observed in Experiment 1 or in the Nonreinforcement condition of the present experiment (Cohen's *d* = 0.33, 95% CI [0.05, 0.61], correlation = 0.39).

A possible explanation for the lack of a cue density effect in the Interference condition is that the cue density effect interacts with other experimental parameters. For example, the cue density effect might occur only when cue density is above a certain threshold value and that value might be lower in the Nonreinforcement condition (where 10 AX- trials seemed to be enough to trigger a cue density effect) than in the Interference condition (where 10 AX-O2 trial were not enough, but 20 were). [Note that threshold value should not be confused with the threshold *C* in the SDT model: The former is not part of the SDT framework and determines whether a cue density is triggered; the latter is part of the SDT framework and determines whether the participant perceives a positive cue-outcome contingency.] If this explanation is correct, then a cue density effect should be observed in both the Interference and the Nonreinforcement conditions if Experiment 2 were replicated using the interference parameters from Experiment 1 (20 AX-O2 trials in the Interference condition and 20 AX-trials in the Nonreinforcement condition). The goal of Experiment 3 was to test this prediction.

Experiment 3

Method

Participants and apparatus

Due to the COVID19 pandemics, Experiment 3 was conducted online using the Gorilla experiment builder (www.gorilla.sc) to create and host the experiment (Anwyl-Irvine, Massonié, Flitton, Kirkham, & Evershed, 2020). Because we hypothesized that moving data collection online might negatively impact effect sizes, we aimed for a sample size of 100. We ended up recruiting 122 participants from the SUNY-Binghamton subject pool. Fifteen participants failed the warm-up, and 8 participants indicated in debriefing that they were distracted during the task (see the procedure below). This left us with 99 participants (57 females, 41 males, and 1 participant declined to provide gender information). Their mean age was 18.98 +/- 1.02 with age ranging from 18 years old to 24 years old).

Procedure

The following changes from Experiment 2 were made to the procedure due to data collection taking place online: (a) as a health precaution, participants with a prior history of epilepsy could not participate in the study; (b) besides being reminded to turn off their cell phone, participants were asked to make sure that they would not be disturbed in the next hour; (c) the duration of the stimuli and the ITI were both changed from 83 ms to 100 ms; (d) we eliminated the confidence ratings because we were not using them; (e) during a contingency question, a timer counting down from 20 s appeared in the lower part of the screen. If participants failed to provide a response within 20 s, no data was recorded on that trial (this only occurred very occasionally and rarely twice for the same participant); (f) when participants answered the contingency question or if the counter reached 0, a new screen was presented to participants with a button reading “Click here to continue” at its center: clicking on it moved the procedure forward (ordinarily starting the next stream). This allowed us to control the position of the mouse at the beginning of the stream because Gorilla does not allow experiments to directly control the location of a user’s mouse.

The positive and negative warm-up phases proceeded as in Experiments 1 and 2. The mixed warm-up differed slightly. Participants were presented with a block of 4 randomly presented streams: two of those streams implemented of positive $\Delta P = 1$ contingency whereas the other two implemented a negative $\Delta P = -1$ contingency. Participants moved to the experimental phase if they answered correctly on all 4 streams composing a block. Otherwise, the block was presented again; if it was presented 10 times without the participant advancing to the experimental phase, the participant failed the warm-up phase and was eliminated.

During the experimental phase, participants were presented with 30 blocks. A block was composed of 6 streams, one for each of the conditions shown in Table 4, presented in a random order. Hence, overall, participants were exposed 30 times to each of the conditions described in Table 4 (in contrast to the 40 blocks of Experiments 1 and 2). Every time participants completed 10 blocks, they were invited to take a break lasting no more than 10 minutes.

Once the experimental phase was complete, participants were presented with a screen at the bottom of which were 5 stimuli (a blob, a circle, a crescent, a square, and a hexagon). They were filled black and appeared over a white background. Participants were asked to click on the square. They were then asked whether they devoted their full attention to the task. Participants were eliminated from the study if they failed to identify the square or said that they had been distracted during the task. As previously mentioned, while no participant failed to identify the square, 8 participants indicated that they did not devote their full attention to the task. All participants were then shown the same debriefing screen shown to participants who failed the warm-up. Data analysis proceeded as in Experiments 1 and 2. It was not possible to compute a *d'* for 1 participant in the Control condition, 3 participants in the Interference condition, and 4 participants in the Nonreinforcement condition.

Results and discussion

First, participants once again were able to discriminate between the positive and negative contingencies in all the conditions as indicated both by the probabilities of detecting a positive contingency (Figure 3, top panel. Control: Cohen's *d* = 0.94, 95% CI [0.73, 1.15], correlation = 0.64; Interference: Cohen's *d* = 0.41, 95% CI [0.26, 0.56], correlation = 0.75; Nonreinforcement: Cohen's *d* = 0.51, 95% CI [0.35, 0.65], correlation = 0.78) and the sensitivity to the contingencies as measured by *d'* (Figure 3, bottom panel). Note that the effect sizes are much lower than in Experiments 1 and 2, which is to be expected from the fact that online data collection does not allow the same experimental control as in situ data collection: this increases the variability in the data, which negatively impacts effect sizes. This precludes meaningful quantitative comparisons between Experiment 3 and Experiments 1 and 2. Otherwise, the data from Experiment 3 are entirely consistent with those of Experiments 1 and 2. First, as in Experiments 1 and 2, compared to the Control condition, the sensitivity to the contingencies was lower in both the Interference condition (Cohen's *d* = 0.55, 95% CI [0.33, 0.78], correlation = 0.47) and the Nonreinforcement condition (Cohen's *d* = 0.39, 95% CI [0.15, 0.64], correlation = 0.32). Moreover, as in Experiment 2, there was no appreciable difference between the Interference and Nonreinforcement conditions (Cohen's *d* = 0.16, 95% CI [-0.76, 0.40], correlation = 0.33).

Second and more importantly, as we predicted, a cue density effect was observed for both the Interference and the Nonreinforcement conditions. Compared to the Control condition, participants were more likely to detect a positive contingency in both the Interference condition (Cohen's *d* = 1.41, 95% CI [1.06, 1.78], correlation = -0.15) and the Nonreinforcement condition (Cohen's *d* = 1.38, 95% CI [1.01, 1.77], correlation = -0.38). This effect was observed for both the positive streams (Control vs. Interference: Cohen's *d* = 1.07, 95% CI [0.76, 1,39], correlation = 0.01; Control vs. Nonreinforcement: Cohen's *d* = 1.13, 95% CI [0.78, 1.47], correlation = -0.23) and for the negative streams (Control vs. Interference: Cohen's *d* = 1.55, 95% CI [1.18, 1.92], correlation = -0.14; Control vs. Nonreinforcement: Cohen's *d* = 1.41, 95% CI [1.04, 1.79], correlation = -0.29).

Hence, it appears that the cue density effect interacts with the interfering treatment. As the critical cue density necessary to trigger a cue density effect appears to be lower in the Nonreinforcement condition than in the Interference condition, we might have expected that, in Experiment 3, the cue density effect would have been larger in the Nonreinforcement condition than in the Interference one. This was not the case: there was no appreciable difference between Interference and Nonreinforcement (Overall probability of detecting a positive contingency: Cohen's *d* = 0.04, 95% CI [-0.11, 0.19], correlation = 0.71; Positive streams: Cohen's *d* = 0.08, 95% CI [-0.07, 0.22], correlation = 0.72; Negative streams: Cohen's *d* = -0.03, 95% CI [-0.20, 0.14], correlation = 0.65).

General Discussion

Framing the conclusions through the lens of SDT

Considered as a whole, Experiments 1-3 provide a convincing case that interfering trials (either X-O2 interference trials or X- nonreinforced trials) affect performance in at least two different ways: First, probably through associative interference, they decreased the ability of participants to discriminate between contingencies, as indicated by a lower *d'* in the Interference and Nonreinforcement conditions in all three experiments; Second, by increasing the cue density, interfering trials made participants more likely to report a positive contingency between the cue and the outcome. One way to make sense of these conclusions is to analyze them through the lens of SDT.

Let's focus on the Control and Nonreinforcement conditions in Experiments 2 and 3. The simplest SDT model compatible with the data is depicted in the left panel of Figure 4. In the Control condition, the subjective contingency is drawn after a positive, relative to a Negative, stream from a Gaussian distribution with mean $\mu_P(Ctr)$, with respective to $\mu_N(Ctr)$, and standard deviation σ. The participant perceives a positive X-O1 contingency when the subjective contingency falls above a threshold $C(Ctr)$. Likewise, in the Nonreinforcement condition, the subjective contingency is drawn after a positive, relative to a negative stream, from a Gaussian distribution with mean $\mu_P(Nonreinf)$, relative to $\mu_N(Nonreinf)$, and standard deviation σ. The participant perceives a positive X-O1 contingency when the subjective contingency falls above a threshold $C(Nonreinf)$. Finally, and crucially, $C(Ctr) = C(Nonreinf)$.

According to this model, $\mu_N(Nonreinf) > \mu_N(Ctr)$ and $\mu_P(Nonreinf) > \mu_P(Ctr)$. As we assumed that the subjective contingency reflects the strength of the cue-outcome association, this means that, following a negative stream, the cue-outcome association was stronger in the Nonreinforcement condition than in the Control condition, and likewise following a positive stream. This assumption is simply not plausible in light of what we know about nonreinforcement and extinction. The only way to save this model would be to assume that the additional X-Null trials in the Nonreinforcement condition (where the Null outcome is the absence of an outcome) resulted in additional backward O1-X pairings that would have been incorrectly processed by the participant as forward X-O1 pairings because of the fast stimulus presentation rate, hence boosting the strength of the X-O1 association[1]. This could

[1] We thank Andrew Delamater for this suggestion.

explain the pattern of results described in the present series, but we do not believe this account is correct based on the results of another series which used the streaming procedure to compare proactive, interspersed, and retroactive interference (Jozefowiez et al., 2022). The interspersed condition in that series used the same parameters (ΔP values, number of interspersed interference X-O2 trials, stimulus and ITI duration); yet, this did not lead to a cue density effect. We will comment below on the reasons that we think explain the discrepancy between that result and the one observed in the present series. For now, let us just say that the lack of a cue density effect in Jozefowiez et al. (2022) appears to refute the backward pairing hypothesis: the present series and the one by Jozefowiez et al. (2022) are identical with respect to the opportunity for interspersed interference to induce backward O1-X pairings that would have been mistakenly processed by the participants as forward X-O1 pairings. Hence, if the backward pairing hypothesis was true, a cue density effect should have also observed in Jozefowiez et al.'s (2022) interspersed interference condition.

Hence, as intriguing as the backward pairing hypothesis is, we do not think it provides a satisfactory account of the cue density effect observed in Experiments 1 to 3. It seems more plausible to assume that the model in the left panel of Figure 4 is incorrect and to assume instead that $\mu_N(Nonreinf) < \mu_N(Ctr)$ and $\mu_P(Nonreinf) < \mu_P(Ctr)$. This implies that $C(Nonreinf) < C(Ctr)$. To understand why, consider the right panel of Figure 4. This is almost the same SDT model as the one we considered previously, except that we no longer assume that $C(Ctr) = C(Nonreinf)$. Imagine the subjective distributions for the Nonreinforcement conditions sliding left and right on the x-axis. The value of $d'(Ctr)$ and the location of $C(Ctr)$ relative to $\mu_N(Ctr)$ cannot change because they are determined by the data[2]. The same is true for $d'(Nonreinf)$ and the location of $C(Nonreinf)$ relative to $\mu_N(Nonreinf)$. Finally, remember we assumed that $\mu_N(Nonreinf)$ must always be below $\mu_N(Ctr)$. Given these constraints, no matter the value of $\mu_N(Nonreinf)$ is, $C(Nonreinf)$ will always be below $C(Ctr)$. Hence, we can conclude that the nonreinforced trials have affected both the subjective contingency distribution [because $d'(Nonreinf)$ is lower than $d'(Ctr)$] and the threshold value. The same reasoning, reaching the same conclusion, could be carried out substituting the Interference condition in Experiments 1 and 3 for the Nonreinforcement condition.

Overall, the current study provides support for the proposal that the various parameters of an SDT model are not equally sensitive to various experimental manipulations. At least if one assumes that the strength of the cue-outcome association constitutes the main component of the subjective contingency, the data indicate that associative interference and nonreinforcement affect the subjective contingency distributions whereas cue density affects the threshold value. Stated another way, the SDT analysis revealed that associative interference and nonreinforcement impacted the strength of the cue-outcome association whereas cue density did not.

Why does cue density affect the threshold value?

Let $a(O1)$ be the activation of the representation of outcome O1 when a participant is asked to assess the X-O1 contingency. Participant reports that the X-O1 contingency is positive when

$$a(O1) > \lambda \quad (2)$$

otherwise, they do not. In classic associative fashion,

---

[2] The value of $d'(Ctr)$ is given by Equation (1). -Z[$P_{Ctr}(+|Neg)$] $= C(Ctr) - \mu_N(Ctr)$ (Wickens, 2002).

$$a(O1) = V_{O1}(X)a(X) \qquad (3)$$

where $V_{O1}(X)$ is the strength of the X-O1 association and $a(X)$ is the activation level of the representation of X. From Equations (2) and (3), we can deduce a participant reports the X-O1 contingency to be positive whenever

$$V_{O1}(X) > \frac{\lambda}{a(X)} \qquad (4)$$

As $a(X)$ is bound to be affected by cue density, this explains why the cue density effect impacts the threshold.

Equation (4) suggests that the rapid and repetitive presentation of X during a stream drove $a(X)$ to a high value, thereby explaining the unusually large cue density effect in the present series. For comparison purposes, the cue density effect observed by Vadillo et al. (2010) corresponds to a Cohen's d of 0.34, which is quite typical. In contrast, the cue density effect observed in the present series ranged from 0.90 to 1.41). If the cues had been longer so that $a(X)$ reached a lower asymptotic value, or if a delay had occurred between the end of a stream and the contingency question, thereby allowing $a(X)$ to decay, the cue density effect might not have been as large or might have even disappeared. Indeed, this would explain why Jozefowiez et al. (2022) failed to find a cue density effect in their interspersed interference condition. Their streams were composed of three phases, each one identical to the one used in the present series, except for the stimuli which were presented in each phase. In the interspersed interference condition, X, O1, and the X-O2 interference trials occurred only in the Phase 2 stream (Phase 1 and Phase 3 streams used different stimuli, see Jozefowiez et al., 2022, for details). As the X-O1 contingency rating occurred at the end of the Phase 3 stream, this effectively created a delay between the X-O1 and X-O2 pairings in Phase 2 and the request for a contingency rating. As X was not shown during Phase 3, this would have left enough time for $a(X)$ to decrease, thereby explaining the lack of a cue density effect in Jozefowiez et al. (2022).

Limitation and further comments

Equation (2) highlights an important limitation of our conclusions. Equations (2) and (4) are mathematically equivalent, but they correspond to two different SDT models: In Equation (2), the subjective contingency is the outcome activation while, in Equation (4), it is the associative strength. Our conclusion that the cue density impacts the threshold while associative interference and nonreinforcement impacts the subjective contingency holds if one looks at the data from the point of view of Equation (3). It does not if one looks at them from the point of view of Equation (2): in this case, both associative interference/nonreinforcement and the cue density affect the subjective contingency. As Equations (2) and (4) are mathematically equivalent, the choice between each decision variable (associative strength vs. outcome activation) cannot be decided on an empirical basis and is, in some way, a matter of convenience. Choosing the associative strength as the decision variable identifies everything that does not affect the associative strength as a source of bias (that is to say, variables affecting the threshold, not the decision variable). Indeed, no matter the choice of the decision variable, the conclusion stands that associative interference This might be a more appropriate choice if one is trying to determine whether a given variable impacts the associative strength or not. and nonreinforcement impact the strength of the cue-outcome association whereas cue density does not. Using the associative strength as the decision variable makes this conclusion more compelling. This perspective also helps put the emphasis on the contribution of nonassociative processes to contingency learning.

A final novel result of this series is that the cue density effect interacted with interspersed interference and nonreinforcement treatments. It was greater with interspersed nonreinforcement than with interspersed interference. There is a large literature (reviewed in Jozefowiez et al., 2020), evaluating whether extinction is more or less efficient than other treatments intended to decrease either US expectancy in presence of the CS or acquired emotional reaction to the CS. The focus in those studies was usually to determine which treatment had the larger impact on one or the other of these dependent variables and, eventually, which was less susceptible to context effects (i.e., renewal). The present series suggests that there may be subtle differences of clinical relevance between extinction and nonreinforcement on one hand, and other interfering treatments on the other beyond the question of the relative efficiency of various interfering treatments and their context dependence. For example, if cue density effects similar to the one observed in the present study were observed in clinically relevant situations, knowing that interference is less susceptible than nonreinforcement to them would be useful information for a clinician when deciding which type of exposure therapy to implement.

References

Algina, J., & Kesselman, H. J. (2003). Approximate confidence intervals for effect size. *Educational and Psychological Measurements, 63,* 537-553.

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15,* 147-149.

Allan, L. G., Hannah, S. D., Crump, M. J. C., & Siegel, S. (2008). The psychophysics of contingency assessment. *Journal of Experimental Psychology: General, 137,* 226-243.

Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning & Motivation, 14,* 381-405.

Allan, L. G., Siegel, S., & Tangen, J. M. (2005). A signal detection analysis of contingency data. *Learning & Behavior, 33,* 250-263.

Anwyl-Irvine, A. L., Massonié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: an online behavioral experiment builder. *Behavior Research Methods, 52,* 388-407.

Crump, M. J. C., Hannah, S. D., Allan, L. G., & Hord, L. K. (2007). Contingency judgments on the fly. *Quarterly Journal of Experimental Psychology, 60,* 753-761.

Cummings, G. (2012). *Understanding the new statistics: effect size, confidence intervals, and meta-analysis.* New York, NY: Routledge.

Hannah, S. D., & Allan, L. G. (2011). The criterion-calibration model of cue interacton in contingency judgments. *Learning & Behavior, 39,* 171-190.

Hannah, S. D., Crump, M. J. C., Allan, L. G., & Siegel, S. (2009). Cue-interaction effects in contingency judgments using the streamed-trial procedure. *Canadian Journal of Experimental Psychology, 63,* 103-112.

Jozefowiez, J. (2021). Individual differences in the perception of cue-outcome contingencies: A signal detection analysis. *Behavioural Processes, 188,* 104398.

Jozefowiez, J., Berruti, A. S., Moshchenko, Y., Peña, T., Polack, C. W., & Miller, R. R. (2020). Retroactive interference: Counterconditioning and extinction with and without biologically significant outcomes. *Journal of Experimental Psychology: Animal Learning & Cognition, 46,* 443-459.

Jozefowiez, J., Witnauer, J. E., Hang, J., Silverstein, J. W., Woltag, S., Chew, S., & Miller, R. R. (2022). Comparing associative interspersed interference with proactive and retroactive interference. *Manuscript submitted for publication.*

Laux, J. P., Goedert, K. M., & Markman, A. B. (2010). Causal discounting in the presence of a stronger cue is due to bias. *Psychonomic Bulletin & Review, 17,* 213-218.

Maia, S., Lefèvre, F., & Jozefowiez, J. (2018). Psychophysics of associative learning: quantitative properties of subjective contingency. *Journal of Experimental Psychology: Animal Learning & Cognition, 44,* 67-81.

Matute, H., Yarritu, I., & Vadillo, M. A. (2011). Illusions of causality at the heart of pseudoscience. *British Journal of Psychology, 102,* 392-405.

Murphy, R. A., Witnauer, J. E., Castiello, S., Tsvetkov, A., Li, A., Alcaide, D. M., & Miller, R. R. (2022). More frequent, shorter trials enhance acquisition in a training session: there is no free lunch! *Journal of Experimental Psychology: General, 151,* 41-64.

Polack, C. W., Jozefowiez, J., & Miller, R. R. (2017). Stepping back from “persistence and relapse” to see the forest: Associative interference. *Behavioural Processes, 141,* 128-136.

Peirce, J. W. (2007). Psychopy: Psychophysics software in Python. *Journal of Neuroscience Methods, 162,* 8-13.

Perales, J. C., Catena, A., Shanks, D. R., & González, J. A. (2005). Dissociation between judgments and outcome-expectancy measures in covariation learning: A signal detection approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1105-1120.

Pineño, O., Ortega, N., & Matute, H. (2000). The relative activation of the associations modulates interference between elementally-trained cues. *Learning and Motivation,* 31, 128–152.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology, 66,* 1-5.

Siegel, S, Allan, L. G., Hannah, S. D., & Crump, M. J. C. (2009). Applying signal detection theory to contingency assessment. *Comparative Cognition & Behavior Reviews, 4*, 116-134.

Shanks, D. R. (1995). *The psychology of associative learning.* Cambridge: Cambridge University Press.

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology, 60,* 291-309.

Vadillo, M. A., Miller, R. R., & Matute, H. (2005). Causal and predictive-value judgments, but predictions, are based on cue-outcome contingency. *Learning & Behavior, 33,* 172-183.

Vadillo, M. A., Musca. S. C., Blanco, F., & Matute, H. (2010). Contrasting cue-densitiy effects in causal and prediction judgments. *Psychonomic Bulletin & Review, 18,* 100-115.

Wasserman, E. A., Kao, S. F., Van Hamme, L., Kategiri, M., & Young, M. E. (1996). Causation and association. In D. R. Shanks, K. Holyoak, & D. Medin (Eds)*, The Psychology of Learning and Motivation: Causal Learning* (pp. 207-264). San Diego, CA: Academic Press.

Wickens, T. D. (2002). *Elementary signal detection theory.* Oxford: Oxford University Press.

Witt, J. K., Taylor, E. T., Sugovic, M. & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception, 44*, 289-300.

Wixted, J. T., & Stretch, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review, 107*, 368-376.

White, P. A. (2003). Making causal judgments from the proportion of confirming instances: The pCI rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 710-727.

Author Notes

Jérémie Jozefowiez, Univ. Lille, CNRS, UMR 9193 – SCALab – Sciences Cognitives et Sciences Affectives, F-59000 Lille, France; Gonzalo P. Urcelay, School of Psychology, University of Nottingham, UK; Ralph R. Miller, Department of Psychology, SUNY-Binghamton, USA. We would like to thank Sara Ampen, Tara Bousebha, Camille Chanthavong, Thomas Delmotte, Flavien Desmaret, Ophélie Goulas, Mariette Marrant, Mathilde Owczarek, and Sam Teneul for their help in collecting the data for Experiments 1 and 2. Correspondence concerning this article should be addressed to Jérémie Jozefowiez, Laboratory of Cognitive and Affective Sciences (SCALab UMR CNRS 9193), Université de Lille, Domaine Universitaire du Pont de Bois, BP 60149, 58653 Villeneuve d'Ascq Cedex, France. E-mail: jeremie.jozefowiez@univ-lille.fr. This research was supported by a Projet de soutien à l'internationalsation from the University of Lille, by a grant from Agence Nationale de la Recherche (ANR-ACES, J. Jozefowiez), and by NIH Award MH033881. G U was supported by a UK ESRC Grant (ES/R011494/2). All raw data and computer code for the experiments are available at https://orb.binghamton.edu/archive-joz-urc-mil-2021/1 or upon request from J. Jozefowiez. The authors report no conflicts of interest.

Table 1

Number and type of trials for the streams used during the warm-up in Experiments 1, 2, and 3.

Stream name	Composition	ΔP
Positive warm-up	10 AX-O1/10 A-	1.00
Negative warm-up	10 AX-/10 A-O1	-1.00

Table 2

Number and type of trials for the streams used during the experimental phase of Experiment 1. To compute Δ*P, the AX-O2 trials were treated as AX- trials while the AW-O2 trials have been treated as A- trials.*

Name	Composition	ΔP
Ctr-Positive	7 AX-O1/3 AX-/3 A-O1/7 A- (Core) + 20 AW-O2 (interference)	0.60
Ctr-Negative	3 AX-O1/7 AX-/7 A-O1/3 A- (Core) + 20 AW-O2 (interference)	0.07
Int-Positive	7 AX-O1/3 AX-/3 A-O1/7 A- (Core) + 20 AX-O2 (interference)	-0.07
Int-Negative	3 AX-O1/7 AX-/7 A-O1/3 A- (Core) + 20 AX-O2 (interference)	-0.60

Ctr = Control; Int = Interference

Table 3

Number and type of trials for the streams used during the experimental phase of Experiment 2. To compute ΔP, the AX-O2 trials were treated as AX- trials while the AW-O2 trials have been treated as A- trials. Nonreinf = Nonreinforcement.

Name	Composition	ΔP
Control-Positive	7 AX-O1/3 AX-/3 A-O1/7 A- + 10 AW-/10 AW-O2	0.60
Control-Negative	3 AX-O1/7 AX-/7 A-O1/3 A- + 10 AW-/10 AW-O2	0.07
Interference-Positive	7 AX-O1/3 AX-/3 A-O1/7 A- + 10 AX-O2/10 AW-	0.20
Interference-Negative	3 AX-O1/7 AX-/7 A-O1/3 A- + 10 AX-O2/10 AW-	-0.20
Nonreinf-Positive	7 AX-O1/3 AX-/3 A-O1/7 A- + 10 AX-/10 AW-O2	0.20
Nonreinf-Negative	3 AX-O1/7 AX-/7 A-O1/3 A- + 10 AX-/10 AW-O2	-0.20

Table 4

Number and type of trials for the streams used during the experimental phase of Experiment 3. To compute ∆P, the AX-O2 trials were treated as AX- trials while the AW-O2 trials have been treated as A- trials. Nonreinf = Nonreinforcement.

Name	Composition	∆P
Control-Positive	7 AX-O1/3 AX-/3 A-O1/7 A- + 20 AW-O2	0.60
Control-Negative	3 AX-O1/7 AX-/7 A-O1/3 A- + 20 AW-O2	0.07
Interference-Positive	7 AX-O1/3 AX-/3 A-O1/7 A- + 20 AX-O2	-0.07
Interference-Negative	3 AX-O1/7 AX-/7 A-O1/3 A- + 20 AX-O2	-0.60
Nonreinf-Positive	7 AX-O1/3 AX-/3 A-O1/7 A- + 20 AX-	-0.07
Nonreinf-Negative	3 AX-O1/7 AX-/7 A-O1/3 A- + 20 AX-	-0.60

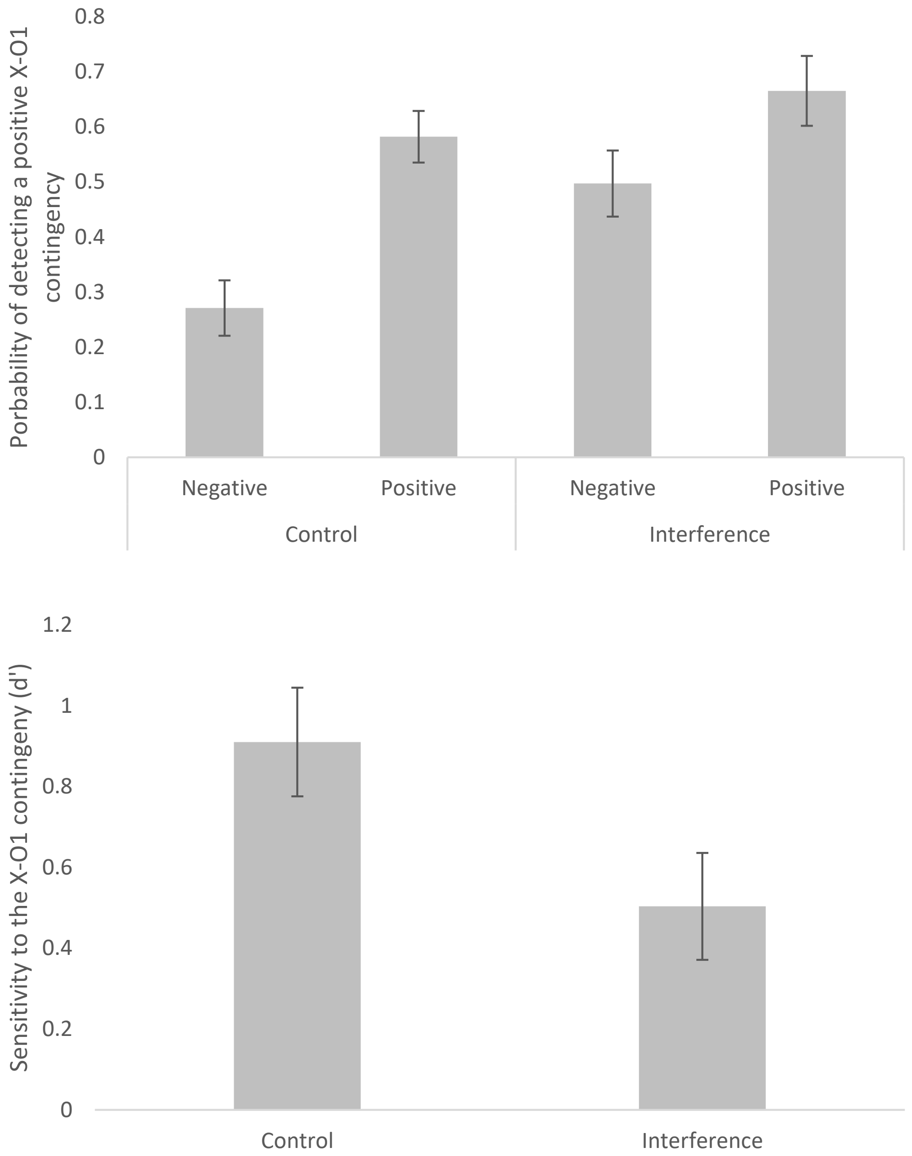

Figure 1. Results from Experiment 1. Top: Probability of detecting a positive contingency as a function of condition. Bottom: Sensitivity to the X-O1 contingency as measured through *d'* as a function of condition. Within-subject error bars are 95% CIs.

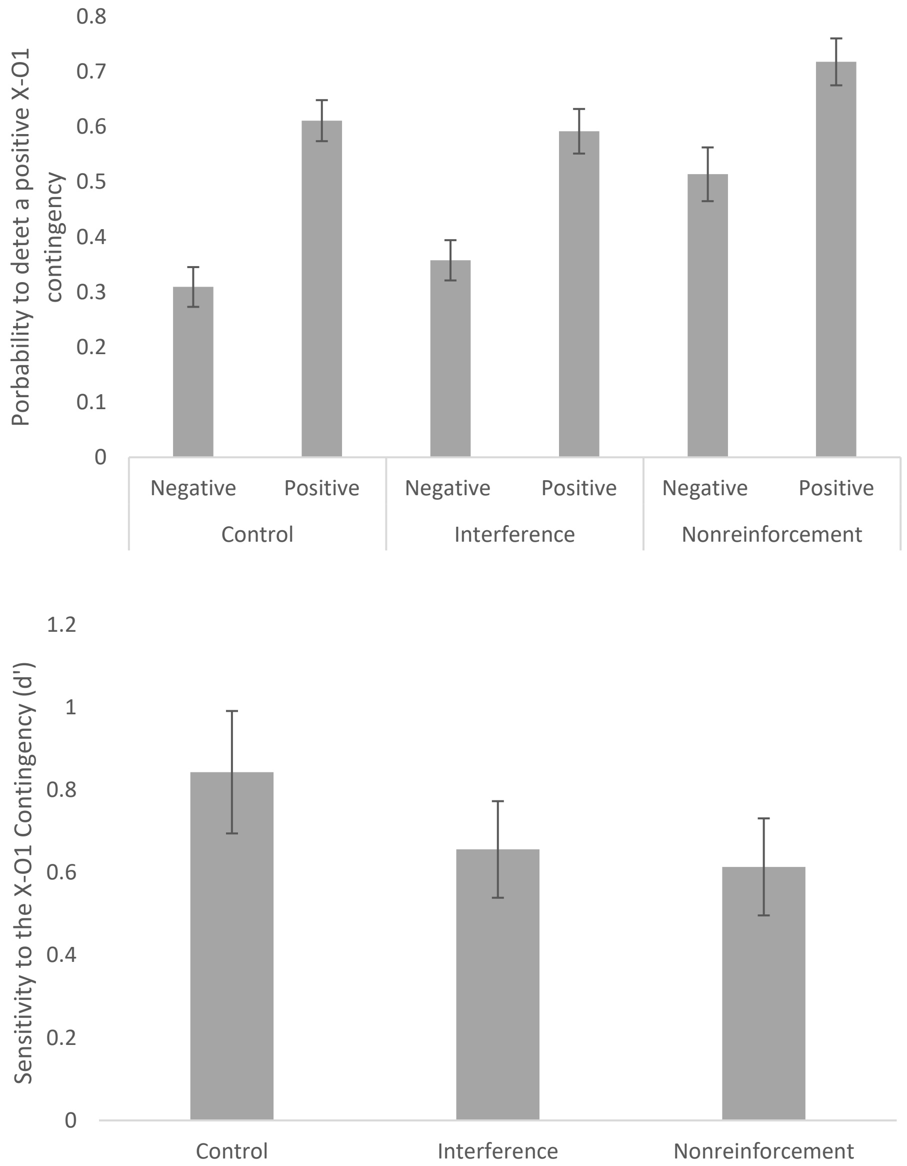

Figure 2. Results from Experiment 2. Top: Probability of detecting a positive X-O1 contingency as a function of conditions. Bottom: Sensitivity to the X-O1 contingency as measured through *d'* as a function of condition. Within-subject error bars are 95% CIs.

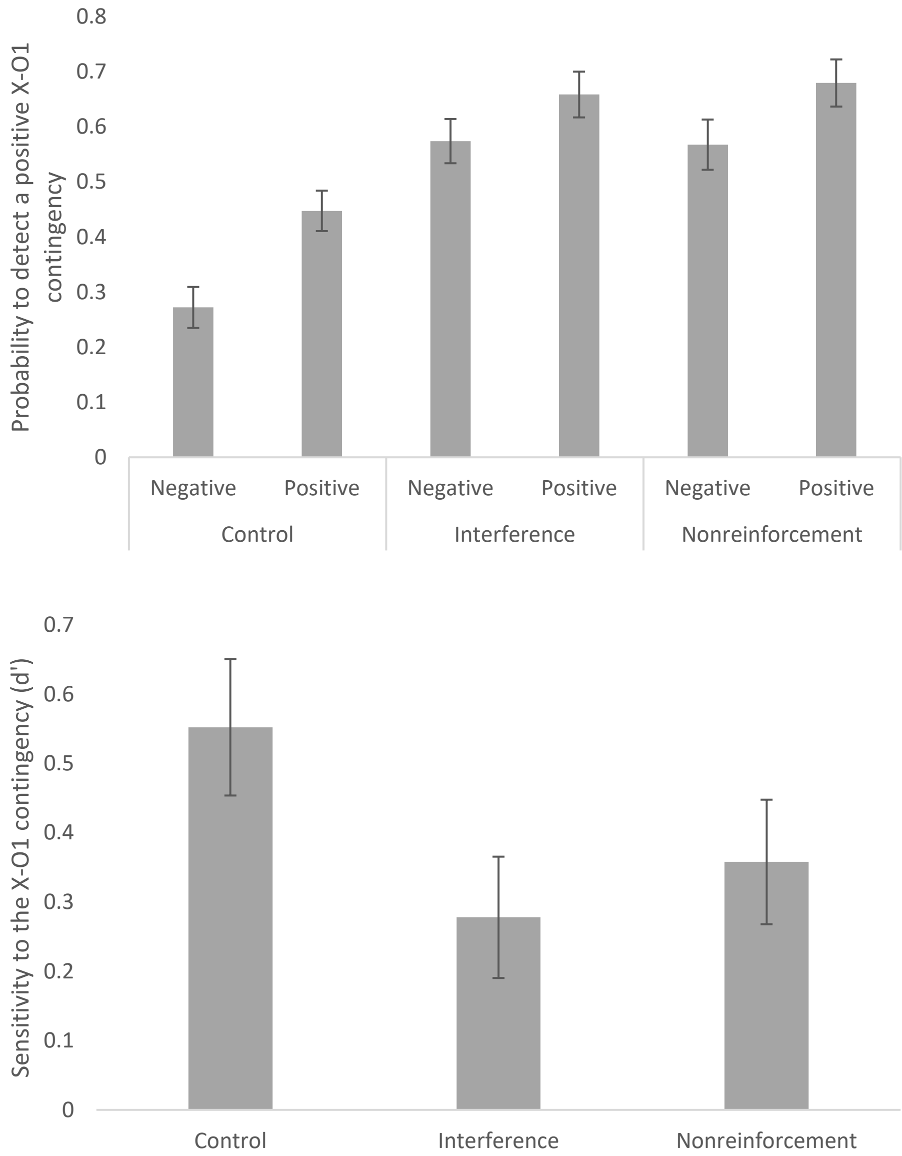

Figure 3. Results from Experiment 3. Top: Probability of detecting a positive contingency as a function of condition. Bottom: Sensitivity to the contingency as measured through *d'* as a function of condition. Within-subject error bars are 95% CIs.

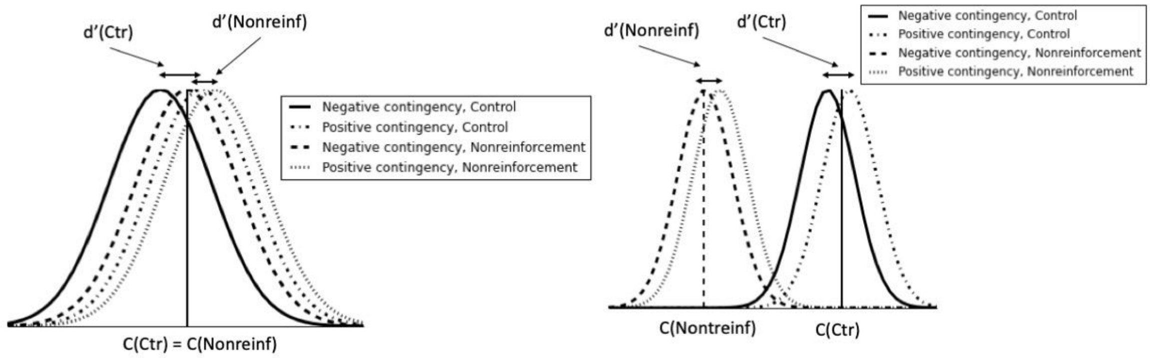

Figure 4. SDT models of Experiments 2 and 3. Left: The criterion is at the same location in both the Control and the Nonreinforcement conditions. Right: The criterion location in the Control conditions is different from the Nonreinforcement conditions. In both models, the value for *d'* and the location of the criterion are set so as to correspond to Experiment 2.