# Retardation of acquisition after conditioned inhibition and latent inhibition training in human causal learning

Peter F. Lovibond[1]

Julie Y.L. Chow[1]

Jessica C. Lee[2,1]

1 School of Psychology, University of New South Wales

2 School of Psychology, University of Sydney

Running head: Conditioned and latent inhibition

This study was funded by an Australian Research Council Discovery Project Grant [DP190103738] to Peter Lovibond. Jessica Lee was supported by a Discovery Early Career Researcher Award from the Australian Research Council [DE210100292].

Please address correspondence to p.lovibond@unsw.edu.au

## Abstract

Inhibitory stimuli are slow to acquire excitatory properties when paired with the outcome in a retardation test. However, this pattern is also seen after simple non-reinforced exposure: latent inhibition. It is commonly assumed that retardation would be stronger for a conditioned inhibitor than for a latent inhibitor, but there is surprisingly little empirical evidence comparing the two in either animals or humans. Thus, retardation after inhibitory training could in principle be attributable entirely to latent inhibition. We directly compared the speed of excitatory acquisition after conditioned inhibition and matched latent inhibition training in human causal learning. Conditioned inhibition training produced stronger transfer in a summation test, but the two conditions did not differ substantially in a retardation test. We offer two explanations for this dissociation. One is that learned predictiveness attenuated the latent inhibition that otherwise would have occurred during conditioned inhibition training, so that retardation in that condition was primarily due to inhibition. The second explanation is that inhibitory learning in these experiments was hierarchical in nature, similar to negative occasion-setting. By this account, the conditioned inhibitor was able to negatively modulate the test excitor in a summation test, but was no more retarded than a latent inhibitor in its ability to form a direct association with the outcome.


*Keywords:* conditioned inhibition, latent inhibition, retardation test, feature negative discrimination, modulation, occasion-setting

# Retardation of acquisition after conditioned inhibition and latent inhibition training in human causal learning

Pavlov's (1927) conditioned inhibition procedure, also known as feature negative training, is an effective method for establishing inhibition to a stimulus. In this procedure, trials on which one stimulus A is paired with an outcome (A+) are intermixed with trials on which a feature B is also present, and no outcome is given (AB-). Here, A predicts the outcome and acquires excitatory properties as shown by its ability to evoke anticipatory responses appropriate to the outcome, whereas B predicts the absence of the outcome and acquires inhibitory properties. Pavlov proposed the *summation test* as a primary method for assessing inhibition, by observing the ability of an inhibitory stimulus to suppress responding to a separately trained excitor. He also noted that inhibitors are slow to acquire excitatory properties when directly paired with the outcome, compared to a novel stimulus. Such a *retardation test* has become a second method for assessing inhibition (Konorski, 1948).

Rescorla (1969) argued that both summation and retardation tests are needed to demonstrate inhibition and rule out alternative processes such as external inhibition and changes in attention. In particular, he drew attention to the procedure of latent inhibition, in which a stimulus is repeatedly presented with no outcome in a pre-exposure phase, before being paired with the outcome in a second phase (Lubow & Moore, 1959; Lubow, 1965). Rescorla argued that latent inhibition "cannot properly be said to produce a conditioned inhibitor; . . . it may be better to consider it as due to some general tendency elicited by the stimulus such as failure to attend to the CS." (Rescorla, 1969, p. 87). The implication of both Pavlov's and Rescorla's treatment of inhibition is that retardation of acquisition after conditioned inhibition training is due to the stimulus having opposite properties to those produced by pairings, which have to be overcome before excitatory responding can be

observed. In turn, this analysis implies that the A+ trials in a conditioned inhibition procedure are critical for establishing inhibition and hence retardation. Pavlov argued there can be no inhibition without prior excitation, and Rescorla and Wagner (1972) formalised this idea in their prediction error model in which the outcome must be expected in order to produce negative prediction error and thereby inhibitory learning.

However, there is little empirical evidence that speaks to the importance of the A+ trials in generating retardation of excitatory learning to the feature after feature negative training. Indeed, the intermixed AB- trials can be thought of as instantiating a latent inhibition procedure, in that they involve exposure to stimulus B with no outcome. In principle, these trials alone could be responsible for the observed retardation of excitatory learning when B is subsequently paired with the outcome. We could only find one direct comparison of the speed of excitatory learning after conditioned inhibition and latent inhibition training, an animal study reported by Miguez et al. (2018, Experiment 1). This study showed an equivalent degree of retardation in the two conditions. However, in their procedure, latent inhibition training involved B- trials whereas conditioned inhibition training involved AB- trials. There is good evidence that latent inhibition is attenuated by a change of context between the pre-exposure phase and retardation phase (Channel & Hall, 1983; Lovibond et al., 1984; Gray et al., 2001). The absence of the partner stimulus A during the retardation phase in the conditioned inhibition condition in Miguez et al. (2018) constitutes a clear context change, making it hard to compare directly with the latent inhibition condition.

Accordingly, we set out to investigate whether conditioned inhibition produces greater retardation of excitatory learning than latent inhibition, using a procedure that controls for context change. Specifically, we used a compound stimulus in the latent inhibition condition, to match the conditioned inhibition procedure, and then paired one element with the outcome in the retardation test. We implemented this design within an established procedure in our lab

for examining inhibitory learning in human causal learning, using the popular allergist task in which foods serve as stimuli and an allergic reaction as the outcome (Lee & Lovibond, 2021).

## Experiment 1

In this experiment we sought to match the training as closely as possible for a conditioned inhibitor B and a latently inhibited stimulus E, in a within-subject design (see Table 1). Conditioned inhibition was implemented with an A+/AB- feature negative discrimination. This procedure necessarily involves the presence of an excitatory stimulus (A) to provide an expectation of the outcome during AB- trials. Therefore, we included an additional stimulus D in the latent inhibition procedure to equate the two procedures, yielding DE- trials. We will refer to B and E as the target stimuli for conditioned and latent inhibition respectively, and A and D as partner stimuli. We also presented stimulus D on its own to match the A+ trials in the conditioned inhibition condition. However, we did not want D to generate any prediction error on DE- trials. Therefore, we elected to leave the associative status of stimulus D indeterminate, by presenting it without any feedback as to the presence or absence of the outcome. We have previously demonstrated that this "no feedback" procedure has little or no effect on the associative status of a stimulus (Lee, Le Pelley & Lovibond, 2022). Finally, we included a range of additional stimuli in order to ensure that participants experienced both single and compound stimuli with and without the outcome, and also to match the design of subsequent planned experiments.

The primary comparison of interest was the acquisition of positive predictive ratings to B and E in the Retardation phase when they were each paired with the outcome. This comparison directly tests to what extent retardation of excitatory acquisition to a pre-exposed stimulus depends on the prior pairing of its partner stimulus with the outcome, because the only difference between B and E was in the treatment of their partner stimuli A and D during

training. To our knowledge this design has not previously been used to compare conditioned and latent inhibition.

## Method

### *Participants*

One hundred participants (55 female, *M* age = 33.7, *SD* age = 12.5) completed the experiment on the Prolific online platform in exchange for monetary payment (20 min at £6GBP/hr). This sample size was based on an expected exclusion rate of 15-20% based on our previous research, with the goal of achieving >80% power for detecting a small (d=0.3-0.4) effect size for within-group comparisons.

### *Apparatus and Stimuli*

As in our previous research (e.g., Lee & Lovibond, 2021), the experimental stimuli (A-K) were selected randomly from a pool of 16 food pictures that included a verbal label (e.g., "chicken"). The allergic reaction outcome consisted of the text "Allergic Reaction!" accompanied by a sad face emoticon. No outcome consisted of the text "No allergic reaction". The experiment was programmed using the jspsych library (de Leeuw, 2015), hosted using JATOS (Lange et al., 2015) and run on participants' web browsers.

### *Procedure*

The project was approved by the University of New South Wales Human Research Ethics Advisory Panel C (approval number 3136), and participants provided online consent. Table 1 shows the sequence of phases. The design and procedure followed that used in our previous research (Lee & Lovibond, 2021; Lovibond & Lee, 2021). In brief, participants were asked to play the role of an allergist trying to work out which foods cause allergic reactions in a patient "Mr X". Each trial represented a meal eaten by the patient. Participants were presented with the text "Mr X eats" followed by pictures of either one or two foods with their verbal

labels. After 500ms, participants were asked to rate the likelihood that Mr X would show an allergic reaction on a visual analogue scale from “Definitely NO ALLERGIC REACTION” to “Definitely ALLERGIC REACTION”. The predictive ratings were recorded on a numerical scale from 0 to 100. When participants had made their rating, the prediction scale and prompt were replaced by either the allergic outcome or no outcome. After 2s, the stimuli and feedback disappeared and the 1-s blank inter-trial-interval (ITI) period commenced.

On “no feedback” trials, a blank image was presented instead of the outcome or no outcome screens, for the same duration. Participants were instructed prior to the experiment that on some trials they would not receive feedback about whether an allergic reaction had occurred or not. We have previously shown that such a procedure leaves the associative status of a cue unchanged in an extinction design (Lee et al., 2022).

Table 1. *Design of Experiment 1.*

| Training | Retardation | Causal ratings |
|---|---|---|
| A+ AB- | B+ | A B |
| C+ | | C |
| D DE- | E+ | D E |
| F- GH+ | F- GH+ | |
| I- IJ- | I- IJ- | I J K |

Note: Letters represent experimental stimuli (foods); + represents the allergic reaction outcome; – represents the absence of the outcome; blank represents no feedback. Within each phase, trial types were randomly intermixed.

*Instructions.* Participants were first shown a series of instruction screens which described the causal scenario and outlined what they would be asked to do. We also included a

brief instruction check which participants were required to pass before proceeding. For further details, please see Lee and Lovibond (2021) and Lee et al. (2022).

*Training phase.* As shown in Table 1, A+ and AB- trials were used to establish B as a conditioned inhibitor. C+ trials were included to match the design planned for the following experiment. DE- trials implemented non-reinforced pre-exposure to the latently inhibited stimulus E. The presence of stimulus D served to match the presence of stimulus A on AB- trials. Stimulus D was also presented outside of the DE compound, but in contrast to the A+ trials, D was followed by no feedback. F- and GH+ trials were included to ensure that both single and compound stimuli were associated with the allergic outcome and no outcome. Finally, I- and IJ- trials were included in an attempt to reduce generalization from the A+/AB- discrimination to the D/DE- discrimination, by demonstrating that it was possible for both stimuli in a non-predictive compound to be non-predictive when presented alone. The order of trials in the Training phase was randomised, with the restriction that the same trial type could not be presented on successive trials. Each trial type was presented 6 times (twice within each of 3 blocks of 16 trials). The left-to-right presentation of the compound stimuli was counterbalanced within each block.

*Retardation phase*. This phase was our primary test of whether the two pre-exposure procedures would slow down subsequent excitatory learning. Accordingly, stimuli B and E were each presented alone and followed by the allergic reaction. F-, GH+, I- and IJ- trials were included to maintain continuity with the Training phase. There were again 6 trials of each type, with trial order randomised within blocks. We did not include a novel control stimulus that was also paired with the outcome in this experiment, in order to keep the number of new contingencies to a minimum, and also to prevent the overall rate base rate of the allergic reaction from becoming too high, which could potentially obscure any retardation effects. However, stimulus C provided a similar reference point as it was a novel stimulus that had

been paired with the outcome in the Training phase, and we included it in our analysis of the causal ratings.

*Causal ratings*. After the Retardation phase, participants completed causal ratings as an additional measure of the impact of the B+ and E+ trials. They were presented with all of the individual stimuli except for the filler stimuli F, G and H, plus a novel stimulus K. They were asked to rate "to what extent each food tended to prevent or cause an allergic reaction" on a visual analogue scale. The scale ranged from "Strongly PREVENTED ALLERGIC REACTION" to "Strongly CAUSED ALLERGIC REACTION", with "No effect" labelled in the middle of the scale. The causal ratings were recorded on a numerical scale from -100 to +100. As in our previous work (e.g., Lee & Lovibond, 2021), we also asked participants if they had written anything down during the experiment.

*Data Analysis*

We analysed the data with planned contrasts using the PSY package (Bird, Hadzi-Pavlovic & Isaac, 2000). The primary analysis concerned predictive ratings across trials in the Retardation phase. Within-participant contrasts compared stimulus B to stimulus E, as well as linear and quadratic trend across the 6 trials and the interactions between the B-E contrast and each of the trend contrasts. We also report the results of post hoc exploratory contrasts based on the observed data. Exploratory tests were not corrected for capitalisation on chance, so significant results should be treated with caution.

A similar analysis approach was used for the Training phase and for the causal ratings. For each contrast of interest, we report the p value as well as the corresponding standardized 95% confidence interval (CI). The mean of the CI represents the standardised effect size (Bird, 2004).

*Transparency and openness*

Data from all experiments in this paper are available on the Open Science Framework (https://osf.io/hftqm/). Stimuli and program code can be accessed by contacting the corresponding author p.lovibond@unsw.edu.au. The experiments were not preregistered.

## Results

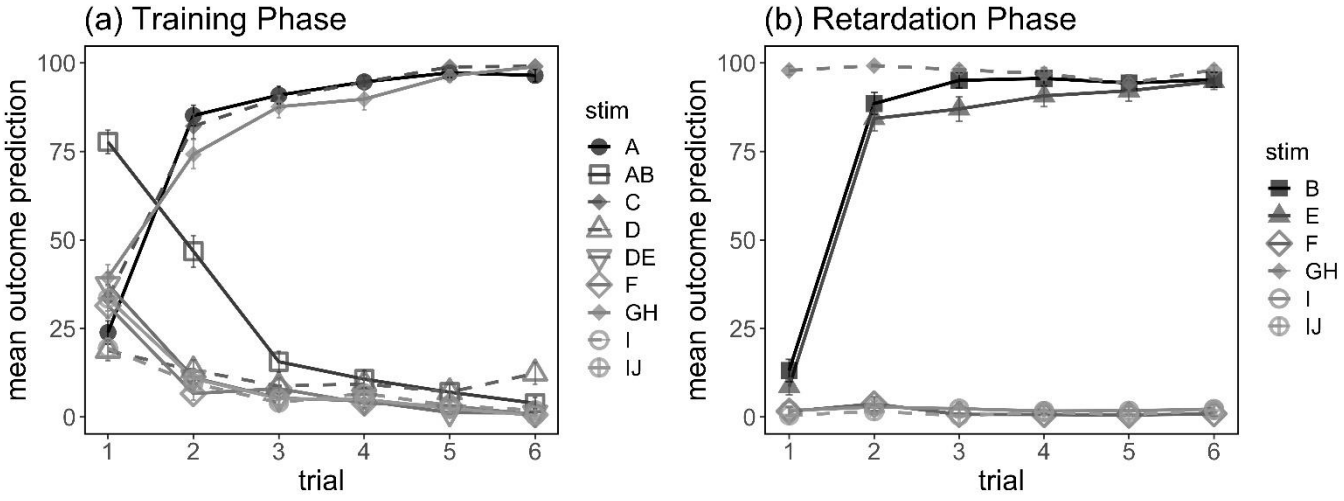
### *Exclusion Criteria*

The exclusion criteria were the same as in Lovibond et al. (2022). Participants were excluded if they a) reported having written anything down during the experiment (n=6), or b) failed the acquisition criterion which was an average rating > 75 for the stimuli that predicted the outcome (A+, C+, GH+) plus an average rating < 25 for the stimuli that predicted no outcome (AB-, DE-, F-, I-, IJ-) in the Training phase (n=7), or c) failed the instruction check more than twice (n=8). After applying all three criteria, 82 participants remained.

### *Training phase*

Figure 1a shows the mean outcome prediction ratings across trials in the Training phase. Participants rapidly acquired differential responding to stimuli that predicted the outcome (A, C and GH) versus those that did not. This resulted in a significant main effect for the contrast that compared predictive and non-predictive stimuli ($F(1,81) = 3344.7$, $p<.001$, 95% CI = 3.01, 3.22), as well as an interaction between this contrast and linear trend over trials ($F(1,81) = 1441.1$, $p<.001$, 95% CI = 1.39, 1.54). Ratings on the first AB- trial were higher than to the other stimuli, due to approximately half of the participants having previously received an A+ trial. This pattern led to a significant main effect for the contrast comparing AB- to the other two non-reinforced compounds (DE- and IJ-) averaged over trials ($F(1,81) = 96.9$, $p<.001$, 95% CI = 061, 0.92), as well as an interaction between this contrast and linear trend over trials ($F(1,81) = 121.0$, $p<.001$, 95% CI = -0.86, -0.60). By the end of the Training phase, ratings were similarly low for all the non-reinforced trial types.

Figure 1. Mean outcome prediction ratings (+/– 1 SE) during each phase of Experiment 1.




*Retardation phase*

Figure 1b shows the mean outcome prediction ratings across trials in the Retardation phase. Participants maintained low ratings to F, I and IJ, and high ratings to GH, throughout the phase. Statistical analysis focused on the B+ and E+ trials. Both stimuli showed rapid acquisition from trial 1 to trial 2, and further slower acquisition on subsequent trials, to a similar asymptote. The main effect for the B-E comparison, averaged over trials, was just significant in the direction of higher ratings for the conditioned inhibitor B than for the latently inhibited stimulus E ($F(1,81)=4.32$, $p=.041$, 95% CI = 0.01, 0.47). That is, retardation of learning seemed to be slightly greater for E than B. This comparison did not interact with either linear trend over trials ($F(1,81)=1.47$, $p=.23$, 95% CI = -0.25, 0.06) or quadratic trend over trials ($F(1,81)=1.80$, $p=.18$, 95% CI = -0.24, 0.05). We also tested two exploratory simple effect contrasts. The first of these compared the first trial of B with the first trial of E, to test whether the stimuli had different starting values. The second compared the second trial of B with the second trial of E, to test whether the stimuli differed after initial acquisition. Both

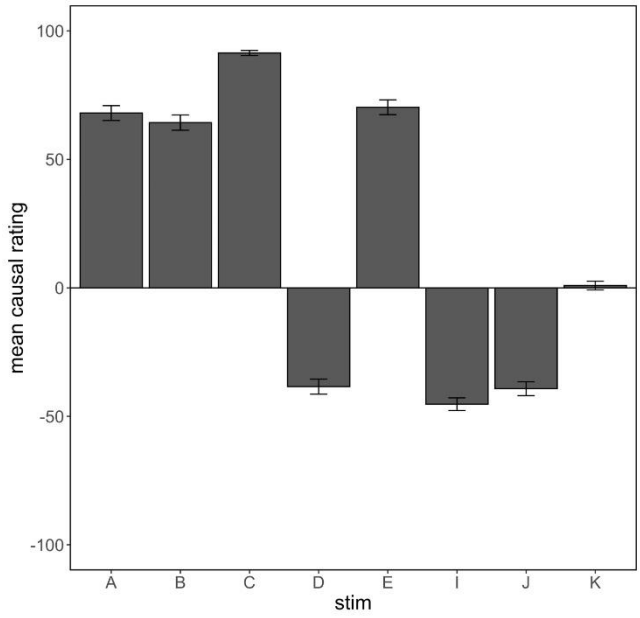
were non-significant ($F(1,81)=1.63$, $p=0.21$, 95% CI = -0.15, 0.68; and $F(1,81)=1.09$, $p=0.30$, 95% CI = -0.23, 0.73, respectively).

*Causal rating test*

Figure 2 shows the mean causal ratings for each stimulus. Ratings for the latently inhibited stimulus E did not differ significantly from the conditioned inhibitor B ($F(1,81)=1.04$, $p=0.31$, 95% CI = -0.39, 0.12). Both B and E were rated as less causal than stimulus C, which had also been paired with the outcome (in the Training phase) but without any prior history of pre-exposure ($F(1,81)=25.8$, $p<.001$, 95% CI = 0.32, 0.73). This pattern occurred despite the fact that on the last trial of the Retardation phase, which was immediately prior to the causal ratings, the final outcome predictions for B (95.3) and E (94.7; see Figure 1b) were relatively close to those on the final training trial for C in the Training phase (99.0; see Figure 1a).

An unexpected finding was that causal ratings for stimulus C were higher than for stimulus A, despite both stimuli having been consistently followed by the outcome when presented alone. We tested this difference with an exploratory contrast, which was significant ($F(1,81)=14.9$, $p<.001$, 95% CI = -0.77, -0.24). Although stimulus A may have been subject to some extinction on AB- trials, the final predictive rating for stimulus A in the Training phase (96.5) was very similar to the final rating for C (99.0). Furthermore, in previous experiments with a similar design we have not observed any substantial differences between A and C in causal ratings (e.g., Lee & Lovibond, 2021, Lovibond & Lee, 2021). Therefore, the lower causal ratings for stimulus A in this experiment appear to have been due to the pairing of its partner stimulus B with the outcome in the Retardation phase, perhaps creating some uncertainty about A. Finally, the stimuli that had not been paired with the outcome during the experiment (D, I and J) were all rated as mildly preventive, and their mean ratings were accordingly lower than for the novel stimulus K ($F(1,81)=74.8$, $p<.001$, 95% CI = 0.70, 1.13).

Figure 2. *Mean causal ratings (+/– 1 SE) for stimuli in Experiment 1. Positive ratings indicate causation, negative ratings indicate prevention, and zero indicates no effect.*




## Discussion

Our primary test of retardation of excitatory learning after conditioned inhibition and latent inhibition training was from the predictive ratings in the Retardation phase. These ratings showed a small but significant advantage in learning (less retardation) for the conditioned inhibitor B compared to the latently inhibited stimulus E. However, the 95% confidence interval showed that the true effect size for this difference could plausibly be as low as 0.01. The two stimuli did not differ at asymptote. Finally, causal ratings for the target stimuli B and E were lower than for stimulus C, which had not been pre-exposed before being paired with the outcome (in the Training phase). This difference is consistent with an overall retardation effect for B and E. However, the two pre-exposed stimuli did not differ significantly. Overall, then, the results from Experiment 1 suggest that the degree of retardation observed after conditioned inhibition and latent inhibition training, controlling for the presence of partner stimuli, is very

similar. That is, the presence of the outcome on A trials but not D trials did not lead to greater retardation of excitatory learning to the accompanying target stimulus on non-reinforced compound trials (AB- and DE-).

## Experiment 2

Given the similar rate of acquisition we observed to a conditioned inhibitor B and a latent inhibitor E in Experiment 1, we decided to conduct Experiment 2 to check that our training had in fact generated stronger conditioned inhibition to B than E. If not, then the lack of difference between these stimuli in a retardation test would be unsurprising. We therefore gave the same training as in Experiment 1, but in the second phase we combined stimuli B and E with the excitor C in a summation test. In our previous work with similar designs (e.g., Lee & Lovibond, 2021), we have used one element of a non-reinforced compound as a control stimulus (as suggested by Karazinov & Boakes, 2004). In this experiment, rather than add a further compound, we decided to use one of the existing stimuli, I, as a control for the summation test (i.e., to compare CB and CE with CI). We did so on the basis that stimulus I was the most conservative stimulus available, having been presented alone and in combination with J, without being paired with the outcome.

### Method

#### *Participants*

One hundred participants (43 female, *M* age = 26.2, *SD* age = 7.7) completed the experiment on the Prolific online platform in exchange for monetary payment (20 min at £6GBP/hr). The exclusion criteria were the same as for Experiment 1.

#### *Apparatus and Stimuli*

The apparatus and stimuli were the same as in Experiment 1.

*Procedure*

The Training phase for Experiment 2 was identical to that of Experiment 1. Instead of a retardation test, participants were then given a summation test in which the critical stimuli B and E were each combined with the separately trained test excitor C. The control stimulus I was also tested in compound with C. In addition, we tested all trial types from the Training phase (except for the filler trials F and GH), as well as the individual elements from compounds, and a novel stimulus K. Each trial type was presented once, in randomized order, within each of two successive blocks. No outcome feedback was provided during the Summation test phase. All other procedural details were the same as in Experiment 1.

*Data Analysis*

We followed the same analysis strategy as in Experiment 1. For the Summation phase analysis, we averaged ratings over the two presentations of each trial type.

Table 2. *Design of Experiment 2.*

| Training | Summation | Causal ratings |
|---|---|---|
| A+ AB- | C CB CE CI | A B |
| C+ | AB A B | C |
| D DE- | DE D E | D E |
| F- GH+ | | |
| I- IJ- | I J IJ K | I J K |

Notation as per Table 1.

## Results

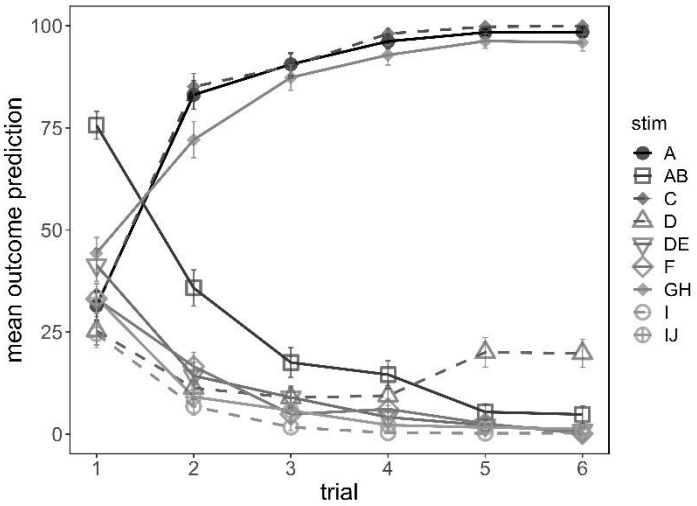
*Exclusion Criteria*

Nine participants failed the write check, 5 failed the acquisition criterion and 4 failed the instruction check. After all exclusions had been applied, 83 participants remained.

*Training Phase*

The training data were very similar to Experiment 1, as shown in Figure 3. Acquisition was demonstrated by a significant difference between the predictive and non-predictive stimuli, averaged over trials ($F(1,82)=3285.7$, $p<.001$, 95% CI = 2.81, 3.01) as well as an interaction between this contrast and linear trend over trials ($F(1,82)=977.0$, $p<.001$, 95% CI = 1.25, 1.42). Predictive ratings on AB- trials were again higher than on DE and IJ trials, averaged over trials, $F(1,82)= 61.7$, $p<.001$, 95% CI =0.47, 0.79, and took longer to decline to a similar asymptote, $F(1,82)= 78.7$, $p<.001$, 95% CI =-0.66, -0.42.

Figure 3. *Mean outcome prediction ratings (+/– 1 SE) during the Training phase in Experiment 2.*

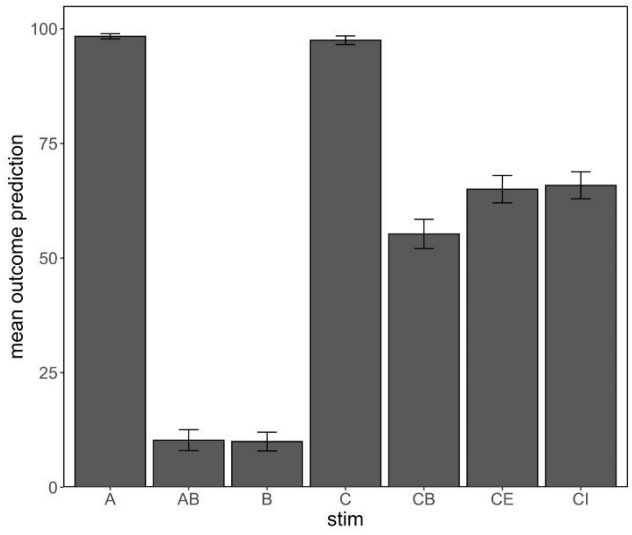



*Summation test*

The Summation phase predictive ratings are shown in Figure 4. The critical contrast of interest compared CB with CE. This contrast was significant, $F(1,82)=14.3$, $p<.001$, 95%CI=-

0.75, -0.23, confirming that the putative inhibitory stimulus B was better able to suppress predictive ratings to a separately trained causal stimulus C than the matched latent inhibitor E. Exploratory comparisons showed that CB differed significantly from the control compound CI, $F(1,82)=15.1$, $p<.001$, 95%CI=-0.81, -0.26, whereas CE did not, $F<1$. These results confirmed that the conditioned inhibitor B passed a summation test whereas the latent inhibitor E did not.

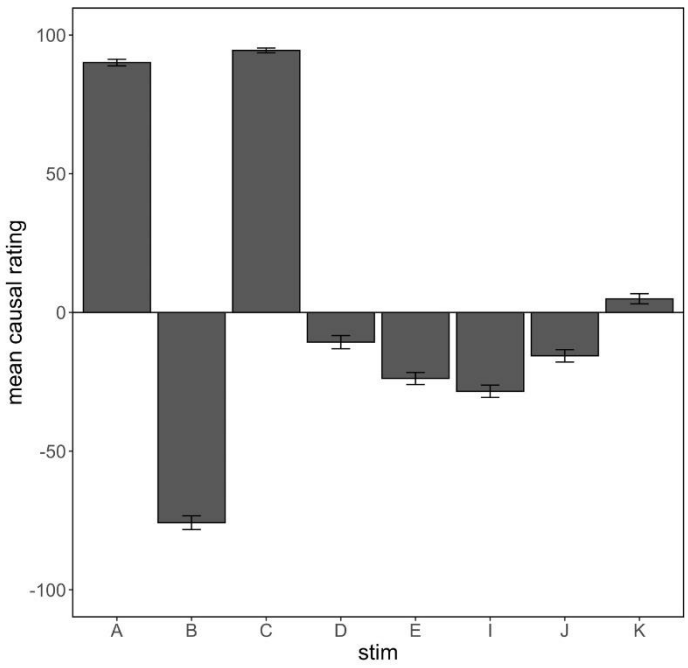
Figure 4. *Mean outcome prediction ratings (+/– 1 SE) during the Summation test phase in Experiment 2.*




*Causal Rating Test*

Figure 5 shows mean causal ratings for the test stimuli in Experiment 2. The primary finding here was that causal ratings to the inhibitory stimulus B were much lower (more preventive) than for the latent inhibitor E, $F(1,82)=61.4$, $p<.001$, 95%CI=-1.79, -1.07. This finding confirms that participants perceived B to be a moderately strong inhibitor, relative to E. Unlike in Experiment 1, causal ratings to A were only slightly lower than those to C, $F(1,82)=6.6$, $p=.01$, 95%CI=-0.21, -0.03. This pattern suggests that the larger difference seen in Experiment 1 was due to the pairing of A's partner stimulus B with the outcome in the preceding retardation test.

Figure 5. *Mean causal ratings (+/– 1 SE of the mean) for test stimuli in Experiment 2.*




## Discussion

After the same training procedure as in Experiment 1, the summation test confirmed that the conditioned inhibitor B was better able to suppress predictive ratings to C than the latent inhibitor E. Although the magnitude of summation was not large, it was highly significant and the effect size was moderate (Cohen's d= 0.49). As noted earlier, summation in human causal learning tasks is typically only modest (e.g., Karazinov & Boakes, 2004). The causal rating test confirmed that participants judged B to be strongly preventive (mean rating = -75.8), and significantly more so than E (mean rating = -23.8). The effect size (Cohen's d) for this difference was 1.43. Thus, Experiment 2 makes it clear that the training procedure used in Experiments 1 and 2 generated much stronger inhibitory learning to B than to E. Hence, the lack of any difference in retardation of excitatory learning between stimuli B and E in Experiment 1 cannot be attributed to a failure to establish inhibitory learning to B.

## Experiment 3

In this experiment, we wished to replicate and test the generality of the pattern seen in Experiment 1 where a conditioned inhibitor and a latent inhibitor showed similar rates of acquisition when each was paired directly with the outcome. We elected to use a between-group design, for two reasons. First, although a within-subject design has the advantage of controlling the level of associative strength gained by the context, it also sets up the possibility of interactions between the stimuli of interest, perhaps leading to more similar responding to them. Second, a between-group design is simpler for participants to learn, and is more similar to the majority of previous research in the literature, which has investigated retardation after either conditioned inhibition or latent inhibition training separately.

### Method

#### *Participants*

One hundred and seventy-five participants (71 female, *M* age = 28.2, *SD* age = 10.0) completed the experiment on the Prolific online platform in exchange for monetary payment (20 min at £6GBP/hr). Before exclusions, 84 participants were randomly allocated to the Conditioned Inhibition group and 91 were randomly allocated to the Latent Inhibition group. The exclusion criteria were the same as for Experiments 1 and 2.

#### *Apparatus and Stimuli*

The apparatus and stimuli were the same as in Experiments 1 and 2.

#### *Procedure*

As shown in Table 3, the general design of Experiment 3 was similar to that of Experiment 1. However, conditioned inhibition and latent inhibition training were applied to independent groups of participants, and we used the same letters A and B to represent the partner stimuli and target stimuli respectively. The Conditioned Inhibition group received A+

and AB- trials whereas the Latent Inhibition group received A and AB- trials, where stimulus A was followed by no feedback in the same way as stimulus D in Experiment 1. Each group also received training with similar trial types as in Experiment 1 (C+, DE-, F-, GH+). In order to balance experience with the various outcomes in the two groups, we introduced an additional filler stimulus X, which was followed by no feedback in the Conditioned Inhibition group and by the allergic outcome in the Latent inhibition group.

In the Retardation phase, both groups received 6 pairings of B with the outcome. Because of the simpler design of this experiment, we were able to also pair a novel stimulus I with the outcome to provide a within-phase control for acquisition to B. Finally, we included D-, DE-, F- and GH+ trials in order to provide continuity with the Training phase and also to prevent the overall outcome rate from being too high. In the Causal ratings test, we asked for cause-prevent ratings for all stimuli except for the filler stimuli G and H. All other procedural details were the same as in Experiments 1 and 2.

Table 3. *Design of Experiment 3.*

| Group | Training | Retardation | Causal ratings |
|---|---|---|---|
| Conditioned Inhibition | A+ AB- | | |
| | C+ DE- | B+ | A B C |
| | F- GH+ X | D- DE- | D E |
| Latent Inhibition | A AB- | F- GH+ | F |
| | C+ DE- | I+ | I J |
| | F- GH+ X+ | | |

Notation as per Table 1.

*Data Analysis*

We followed a similar analysis strategy to Experiments 1 and 2, with the addition of a between-group contrast to compare the two groups, and we also tested all interactions between group and repeated measures contrasts. The critical comparison in this study is the predictive ratings to B in the Retardation phase for the Conditioned Inhibition vs Latent Inhibition group.

## Results

*Exclusion Criteria*

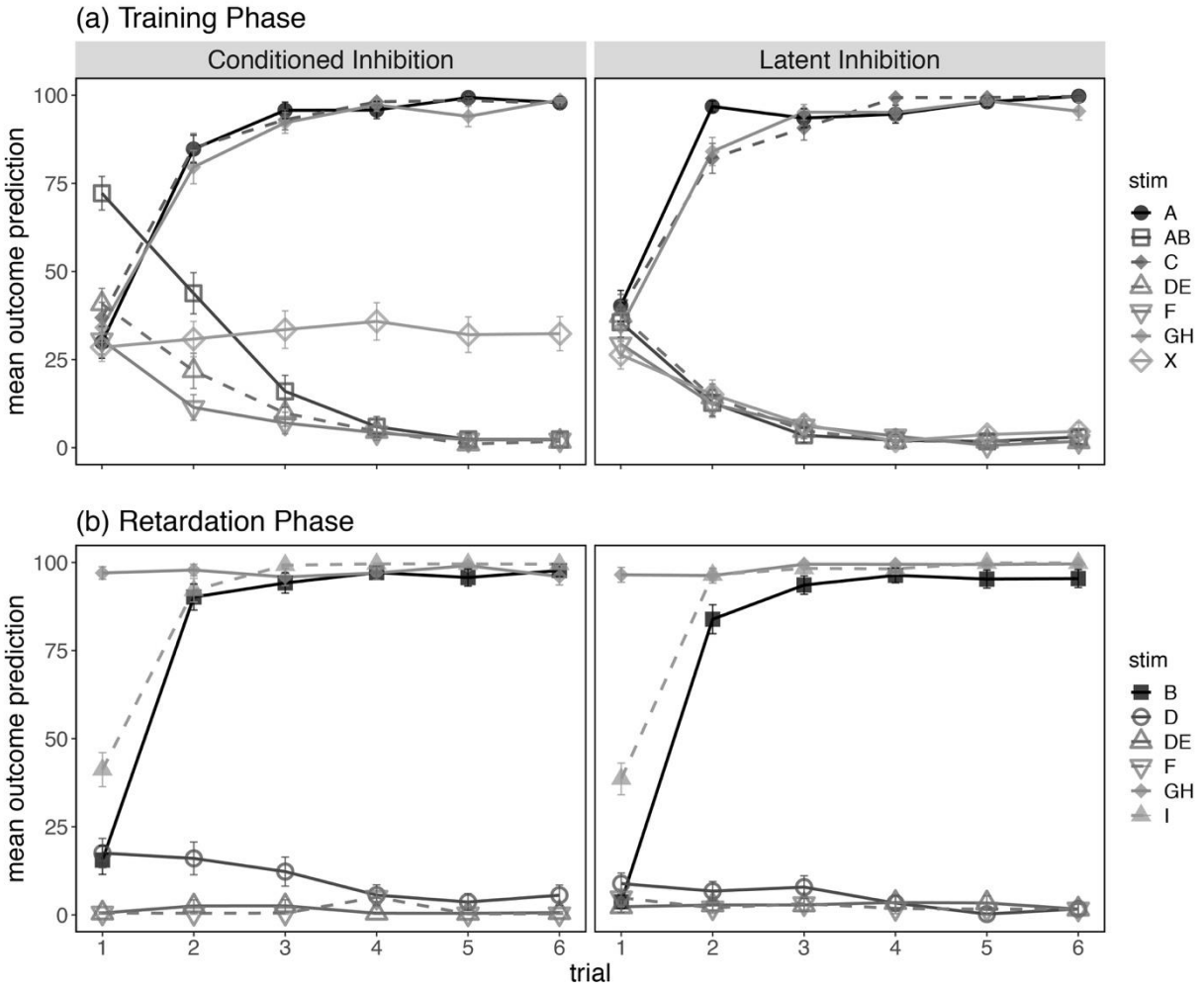
Thirty participants failed the write check, 17 failed the acquisition criterion and 10 failed the instruction check. After all exclusions had been applied, 125 participants remained (58 in the Conditioned Inhibition group and 67 in the Latent Inhibition group).

*Training Phase*

The training data were similar to the previous two experiments, as shown in the top row of Figure 6. Acquisition was demonstrated by a significant difference between the reinforced stimuli (C+ and GH+) and the non-reinforced stimuli (DE- and F-) that were common to the two groups, averaged over trials ($F(1,123)=3555.7$, $p<.001$, 95% CI =2.90, 3.10), as well as an interaction between this contrast and linear trend over trials ($F(1,123)=918.9$, $p<.001$, 95% CI = 1.20, 1.36). Neither of these contrasts interacted with the group comparison ($Fs<1$), confirming comparable acquisition in the two groups. In the Conditioned Inhibition group, predictive ratings on AB- trials were initially higher than on DE trials. However, in the Latent Inhibition group, there was very little difference between these two trial types. This pattern led to an interaction between the group comparison and the AB vs DE comparison ($F(1,123)=16.6$, $p<.001$, 95% CI =0.23, 0.67), as well as a further (i.e., triple) interaction with linear trend over trials ($F(1,123)=22.5$, $p<.001$, 95% CI =-0.78, -0.32).

Finally, in this experiment we had the opportunity to observe predictive ratings to a stimulus that had only ever been followed by no feedback, namely stimulus X in the Conditioned Inhibition group (in the previous experiments, stimuli followed by no feedback had also appeared in an explicitly non-reinforced compound). We ran an exploratory contrast to test for any linear trend in ratings to this stimulus over trials. This contrast was not significant ($F<1$), confirming that the no feedback procedure was successful in maintaining the initial level of outcome prediction to stimulus X (around 30%) in the Conditioned Inhibition group.

Figure 6. *Mean outcome prediction ratings (+/– 1 SE) during the Training phase (top row) and Retardation phase (bottom row) of Experiment 3. The left column shows data for the Conditioned Inhibition group and the right column shows data for the Latent Inhibition group.*



*Retardation phase*

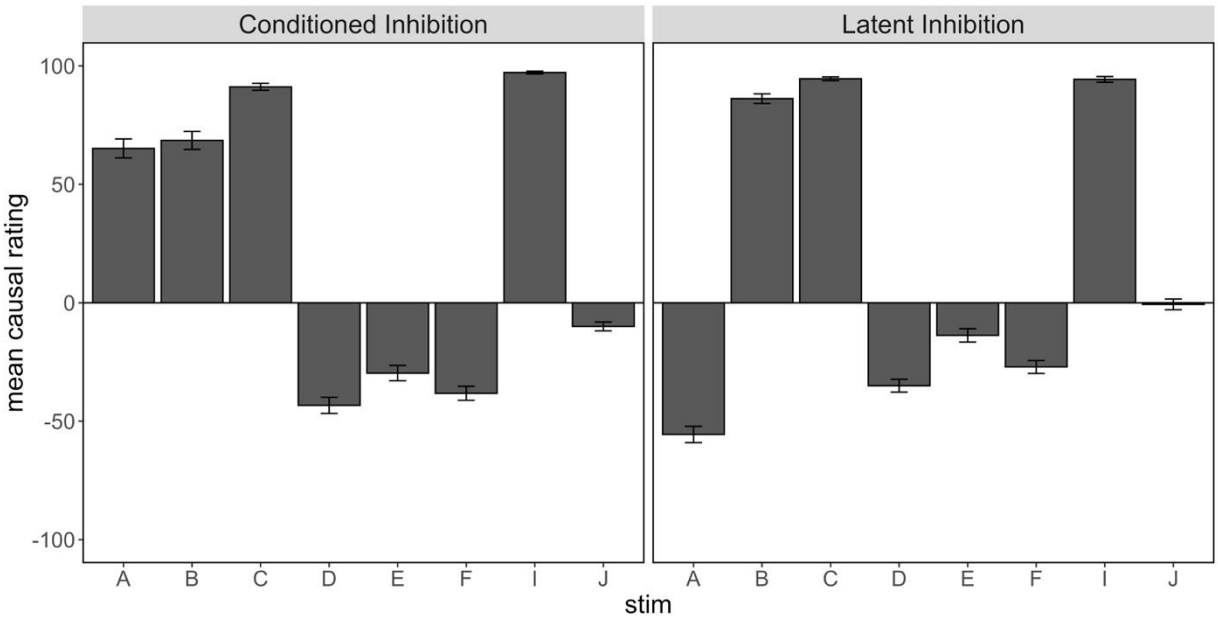
The bottom row of Figure 6 shows the mean outcome prediction ratings for the two groups across trials in the Retardation phase. In this experiment, the predictive rating for B on its first presentation was higher in the Conditioned Inhibition group than in the Latent Inhibition group ($F(1,123)=7.60$, $p=.007$, 95% CI=0.19, 1.17). The speed of acquisition after trial 1 appeared similar for the two groups. Averaged over trials and groups, predictive ratings were higher for the control stimulus I than for the target stimulus B ($F(1,123)=41.1$, $p<.001$, 95% CI=-0.65, -0.35), but this contrast did not interact with groups ($F(1,123)=1.89$, $p=0.17$, 95% CI=-0.09, 0.52). The B vs I comparison did interact with both linear trend over trials ($F(1,123)=41.9$, $p<.001$, 95% CI=0.33, 0.62), and quadratic trend over trials ($F(1,123)=29.1$, $p<.001$, 95% CI=-0.50, -0.23), reflecting convergence between these stimuli across the phase. However, non-significant triple interactions with groups showed that neither of the above effects differed between groups ($F(1,123)=1.92$, $p=0.17$, 95% CI=-0.49, 0.09; and $F(1,123)=1.19$, $p=0.28$, 95% CI=-0.12, 0.42, respectively). Exploratory contrasts failed to show any difference between the groups in predictive ratings to stimulus B on either the second trial ($F(123)=1.26$, $p=0.27$, 95% CI=-0.28, 1.00) or the final trial, $F<1$.

*Causal rating test*

Figure 7 shows the mean causal ratings for individual stimuli. Averaged over groups, ratings were lower for stimulus B than for stimulus I, the novel control stimulus that had been paired with the outcome during the Retardation phase ($F(1,123)=20.38$, $p<.001$, 95% CI=-0.64, -0.25). This difference is consistent with an overall retardation effect due to the Training phase manipulations. Furthermore, in this experiment, the B-I comparison interacted with groups ($F(1,123)=6.33$, $p=0.013$, 95% CI=-0.88, -0.11), indicating lower ratings to B (more retardation) in the Conditioned Inhibition group than in the Latent Inhibition group. An exploratory contrast that directly compared causal ratings to B between the two groups also

just reached significance in the same direction (F(1,123)=4.49, p=.036, 95%CI=-0.82, -0.03). Finally, as in Experiment 1 but not Experiment 2, causal ratings to A were significantly lower than those to C in the Conditioned Inhibition group, F(1,123)=8.7, p=.004, 95%CI=-0.98, -0.19, consistent with our earlier interpretation that this effect was due to the pairing of its partner cue B with the outcome in the Retardation phase.

Figure 7. *Mean causal ratings (+/– 1 SE of the mean) for test stimuli in Experiment 3 for the Conditioned Inhibition group (left panel) and the Latent Inhibition group (right panel).*




## Discussion

The retardation test in Experiment 3 showed a similar pattern of acquisition of positive predictive ratings to the target stimulus B in the two groups. Ratings to B were higher on the first trial in the Conditioned Inhibition group compared to the Latent Inhibition group, an effect we had not observed in Experiment 1. However, there was little evidence of any group difference in acquisition after that initial point. Predictive ratings to B were lower than to the novel control stimulus I, averaged over trials, which was consistent with an overall retardation effect. This effect was primarily due to the first trial where the novel stimulus elicited higher ratings than the target stimulus B. However, there was no difference between the two groups in

the magnitude of this effect. As in Experiment 1, the causal ratings also showed an overall retardation effect for the target stimulus B relative to the novel stimulus I. However, in this experiment, causal ratings for B were lower in the Conditioned Inhibition group, implying greater retardation of learning in the prior phase compared to the Latent Inhibition group.

## General Discussion

In this project, we compared two compound pre-exposure manipulations, conditioned inhibition and latent inhibition, with respect to their ability to slow learning to the target stimulus in a retardation test. The two manipulations were designed to match each other in terms of exposure to the target stimulus, but differ in terms of the treatment of the partner stimulus when presented outside the compound. The partner stimulus was followed by the outcome in the case of conditioned inhibition but followed by no feedback in the case of latent inhibition. Experiment 2 confirmed that this procedural difference conferred stronger inhibitory properties to the target stimulus in the conditioned inhibition condition, as demonstrated in both a summation test and causal ratings.

Our primary test of retardation of excitatory learning was trial by trial acquisition of predictive ratings in a retardation phase where each target stimulus was directly paired with the outcome. Experiment 1 showed similar acquisition curves for the two target stimuli in a within-subject design. The two stimuli did differ on later trials in the direction of lower ratings (more retardation) for the latent inhibition stimulus, but the effect size was small (0.24) and it was only marginally significant. Experiment 3 similarly showed parallel acquisition curves after conditioned inhibition and latent inhibition training in a between-group design, with no significant differences between the groups after the first trial. That experiment also provided some evidence for retardation in both groups relative to a novel control, albeit largely mediated by lower starting predictions for the pre-exposed stimuli relative to the novel stimulus.

Importantly, in both experiments we found no evidence of greater retardation of learning on the outcome prediction measure after conditioned inhibition compared to latent inhibition training.

We also collected causal ratings for each stimulus at the end of each experiment, as is common in the food allergist task. These ratings can be considered another measure of retardation, in the sense that they reflect learning that occurred across the experiment. This measure showed an overall retardation effect, with lower causal ratings for both the conditioned inhibition and latent inhibition stimuli compared to control stimuli that had been paired with the outcome the same number of times but without prior pre-exposure (stimulus C in Experiment 1 and stimulus I in Experiment 3). This difference could also be seen as a measurement artefact whereby the post-experimental rating procedure encouraged participants to consider all of their experiences during the experiment, albeit with somewhat greater weighting to the more recent reinforced trials (Collins & Shanks, 2002). This explanation could account for why causal ratings for the target stimuli were lower than control stimuli, despite the fact that immediately prior to the causal ratings, participants gave predictive ratings to the target stimuli that were near ceiling (trial 6 of the Retardation phase). This finding is also consistent with a multiple memory trace approach to dealing with inconsistent information similar to that put forward by Bouton (2004) in relation to extinction. In both cases, participants have two competing memories of the target cues, one involving the outcome and the other involving no outcome (see also Westbrook and Bouton, 2010). In the case of the present experiments, it seems that trial-by-trial predictive ratings and post-experimental causal ratings tap into different aspects of these memories, with predictive ratings sensitive to recency and causal ratings involving integration across the experiment.

It is therefore possible to interpret the causal ratings as providing evidence of retardation of excitatory learning due to the prior non-reinforced pre-exposure manipulations. On this measure, Experiment 1 showed no difference between the conditioned inhibition and

latent inhibition stimuli, whereas Experiment 3 found a significant difference in the direction of slower learning (more retardation) in the conditioned inhibition condition. Thus the causal rating measure in Experiment 3 provides some evidence for greater retardation after conditioned inhibition than after latent inhibition in a between-group design. It is unclear why the two experiments showed different outcomes on the causal rating measure, given our attempts to match both the conditions of pre-exposure and experience with the allergic outcome in the two groups in Experiment 3. Our only speculation here is that the within-subject design of Experiment 1 may have encouraged participants to treat the two target stimuli as more similar to each other due to their shared history of compound pre-exposure.

Overall, the primary trial-by-trial measure of acquisition in the retardation test showed very similar acquisition curves after conditioned inhibition and latent inhibition training, with a small effect in favour of greater retardation for the latent inhibition stimulus in one experiment. The causal rating measure showed a null effect in a within-subject design and a modest effect in favour of greater retardation for the conditioned inhibition condition in a between-group design. These mixed findings stand in contrast to the clear evidence for stronger inhibitory learning in the conditioned inhibition condition in Experiment 2, on boththe summation test and causal ratings. Accordingly, we conclude that overall, there was relatively little evidence for a stronger retardation effect in the conditioned inhibition condition compared to the latent inhibition condition.

At face value, this finding is problematic for the idea that retardation after conditioned inhibition is a more robust phenomenon than retardation after latent inhibition. It suggests that the negative prediction error on compound trials arising from reinforcing the partner stimulus in the conditioned inhibition procedure does not confer any greater retardation of subsequent excitatory learning compared to the compound latent inhibition procedure we used as a control. One way to explain how negative prediction error in the conditioned inhibition condition

generated greater evidence of inhibition in a summation test but not in a retardation test is to assume that the inhibitory learning, at least with the current procedure, was hierarchical in nature. In other words, participants may have learned a causal structure in which the target feature B modulates the ability of the training excitor A to cause the outcome, rather than learning that B directly prevents the outcome. This structure is analogous to negative occasion-setting in the associative literature (e.g., Fraser & Holland, 2019). An important feature of occasion-setting is that the modulatory properties of a stimulus appear to be at least partly dissociable from its own associability with the outcome. In the present experiments, this property would allow the inhibitory target stimulus to modulate the test excitor in a summation test, without impacting on its speed of acquisition in a retardation test, above and beyond the latent inhibition inherent in the feature negative design.

The idea that feature negative training might give rise to modulatory learning is consistent with our previous work on the nature of inhibitory learning in humans (e.g., Lee & Lovibond, 2021). Using open-ended and forced choice self-report measures, we have found that the majority of participants can be classified as having learned one of three causal structures after feature negative training: configural (memorising stimulus combinations and their outcomes), modulation (as above) or prevention (the feature prevents the outcome). Of these, modulation is typically the largest category. In a summation test, the degree of transfer (i.e., suppression of responding to the test excitor) is greatest in prevention participants and weakest in configural participants, with modulation participants in between. We initially treated these three subgroups as qualitatively distinct, but our more recent work suggests that the modulation and prevention subgroups may differ quantitatively rather than qualitatively. Specifically, we have suggested that the degree of transfer seen in a summation test is a function of participants' willingness to generalize the inhibitory properties of the feature from the training excitor to the test excitor. We have provided evidence that transfer in a summation

test does indeed follow principles of generalization (Chow et al., 2022; see also Bonardi et al., 2017), and we have shown from distributional analysis of summation data that prevention participants overlap with the "willing to generalize" end of the modulation participants' distribution (Lee & Lovibond, 2021). In other words, preventers may simply be modulators who are more willing to generalize. Finally, we have shown that transfer in a summation test is stronger when the test excitor has itself previously been modulated (in line with the occasion-setting literature), and critically, this advantage is seen in prevention participants, not just modulators (Lovibond & Lee, 2021). We still consider that configural participants are qualitatively different from the other two subgroups because their strategy does not involve inferring the properties of individual stimuli, but our working hypothesis is that the remaining participants may all be learning a modulatory causal structure.

The conclusion that modulatory learning may explain the present results rests on the assumption that our compound procedure produced the same loss of associability in the latent inhibition condition (where the partner stimulus was presented without feedback) as in the conditioned inhibition condition (where the partner stimulus was reinforced) – as predicted for example by the theory of Pearce and Hall (1980). However, it is possible to argue that this assumption is incorrect in the present design. Indeed, attentional theories of learning (e.g., Mackintosh, 1975; see Le Pelley et al., 2016) specifically assert that stimuli that have been learned to be predictive maintain associability and hence show diminished latent inhibition. In the conditioned inhibition design, the target stimulus B has predictive value - it signals that its partner stimulus A which otherwise signals the outcome will not do so when B is present. If attention is maintained to B, it could attenuate latent inhibition and thereby reduce its contribution to retardation. This idea seems plausible given the robust evidence for learned predictiveness maintaining the associability of stimuli. However, we note that in order to explain the similarity in degree of retardation seen in the conditioned inhibition and latent

inhibition conditions, the loss of latent inhibition due to learned predictiveness would have to match the degree of retardation arising from the inhibitory properties of stimulus B itself.

One limitation of the present project is that acquisition during the Retardation phase in each experiment was very rapid. This raises the possibility that the lack of differences in retardation between the conditioned inhibition and latent inhibition procedures was due to insensitivity of the task, at least for the trial-by-trial predictive measure. Another difficulty was that the novel control stimulus in Experiment 3 had a much higher starting value on the predictive measure, making it difficult to interpret the magnitude of subsequent changes. One way to reduce the speed of acquisition in future research would be to use partial reinforcement during training, although this would not deal with the starting value difference, which is a common feature of intentional learning tasks such as the allergist task. An alternative would be to examine retardation in an incidental learning task. Such a task might provide evidence for a form of latent inhibition (failure to learn a contingency) that is more similar to that seen in animal research, with potentially quite different properties (see Forrest et al., 2018).

A second potential limitation of our design from the perspective of latent inhibition is that the outcome was presented throughout the Training phase. It could therefore be argued that our procedure is better characterised as learned irrelevance than latent inhibition, where pre-exposure typically occurs before the outcome has been experienced (Baker et al., 2003; Byrom et al., 2018). However, our principal aim was to implement latent inhibition in a design that isolates the role of prediction error, rather than to make specific conclusions about latent inhibition per se. It would be difficult to match these two conditions without presenting the outcome (although see Urcelay et al., 2008, for a possible approach to this issue).

*Summary and Conclusions*

In the present experiments, we observed that conditioned inhibition training (A+/AB-) yielded stronger evidence of inhibition in a summation test than matched compound latent inhibition training (D / DE-). By contrast, there was mixed evidence for a difference between these two pre-exposure procedures in a retardation test. While this result could be due to lack of sensitivity of the present task for assessing retardation, it is also possible that retardation after inhibitory training is entirely due to latent inhibition arising from the non-reinforced compound trials intrinsic to this procedure. Our preferred explanation for this pattern is that inhibitory learning in this task is modulatory in nature, whereby the inhibitory stimulus is able to modulate a separately trained excitor in a summation test, but is no more impeded than a latently inhibited stimulus in its ability to predict the outcome in a retardation test. By this account, the retardation test may be less diagnostic of inhibitory learning than the summation test. An alternative explanation is that the predictive utility of the inhibitor maintained its associability and thereby attenuated the latent inhibition that would otherwise accrue to it. By this account, retardation in the compound control condition was due to latent inhibition, but retardation in the conditioned inhibition condition was instead due to the difficulty in turning an inhibitor into an excitor. We hope the design we have developed, which aims to match latent inhibition to conditioned inhibition as closely as possible, will help tease apart these explanations in future research.

## References

Baker, A.G., Murphy, R.A., & Mehta, R. (2003). Learned irrelevance and retrospective correlation learning. *Quarterly Journal of Experimental Psychology*, *56*, 90–101. https://doi.org/10.1080/02724990244000197

Bird, K. D., Hadzi-Pavlovic, D. & Isaac, A. P. (2000). *PSY* [Computer software]. Available from http://www.psy.unsw.edu.au/research/psy.htm

Bird, K.D. (2004). *Analysis of variance via confidence intervals*. London: Sage Publications. https://dx.doi.org/10.4135/9781849208598

Bonardi, C., Robinson, J., & Jennings, D. (2017). Can existing associative principles explain occasion setting? Some old ideas and some new data. *Behavioural Processes, 137*, 5-18. https://doi.org/10.1016/j.beproc.2016.07.007

Bouton, M.E. (2004). Context and behavioral processes in extinction. *Learning & Memory*, *11*, 485-494. https://doi.org/ 10.1101/lm.78804

Byrom, N.C., Msetfi, R.M. & Murphy, R.A. (2018). Human latent inhibition: Problems with the stimulus exposure effect. *Psychonomic Bulletin & Review*, *25*, 2102–2118. https://doi.org/10.3758/s13423-018-1455-4

Channell, S. & Hall, G. )1083). Contextual effects in latent inhibition with an appetitive conditioning procedure. *Animal Learning & Behavior*, *11*, 67–74. https://doi.org/10.3758/BF03212309

Chow, J.Y-L., Lee, J.C. & Lovibond, P.F. (in press). Inhibitory summation as a form of generalization. *Journal of Experimental Psychology: Animal Learning and Cognition.* https://doi.org/10.1037/xan0000320

Collins, D. J., & Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition*, *30*, 1138–1147. https://doi.org/10.3758/BF03194331

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioural experiments in a web browser. *Behaviour Research Methods, 47,* 1-12. https://doi.org/10.3758/s13428-014-0458-y

Forrest, D.R.L., Mather, M. and Harris, J.A. (2018). Unmasking latent inhibition in humans. *Quarterly Journal of Experimental Psychology, 71,* 380–395. https://doi.org/10.1080/17470218.2016.1249894

Fraser, K. M., & Holland, P. C. (2019). Occasion Setting. *Behavioral Neuroscience, 133,* 145-175. https://doi.org/10.1037/bne0000306

Gray, N.S., Williams, J., Fernandez, M., Ruddle, R.A., Good, M.A. & Snowden R.J. (2001). Context dependent latent inhibition in adult humans. *Quarterly Journal of Experimental Psychology, 54,* 233-45. https://doi.org/10.1080/02724990143000027

Karazinov, D.M. & Boakes, R.A. (2004). Learning about stimuli that prevent an outcome: Conditioned inhibition and differential inhibition in human predictive learning. *Quarterly Journal of Experimental Psychology, 57,* 153-178. https://doi.org/10.1080/02724990344000033

Konorski, J. (1948). Conditioned reflexes and neuron organization. Cambridge: Cambridge University Press.

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLoS ONE, 10(6)*: e0130834. https://doi.org/10.1371/journal.pone.0130834

Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N. & Wills, A. J. (2016). Attention and associative learning in humans: An integrative review. *Psychological Bulletin, 142,* 1111-1140. https://doi.org/10.1037/bul0000064

Lee, J.C., Le Pelley, M.E. & Lovibond, P.F. (2022). Nonreactive testing: Evaluating the effect of withholding feedback in predictive learning. *Journal of Experimental Psychology: Animal Learning and Cognition*, *48*, 17-28. https://doi.org/10.1037/xan0000311

Lee, J.C. & Lovibond, P.F. (2021). Individual differences in causal structures inferred during feature negative learning. *Quarterly Journal of Experimental Psychology, 74(1),* 150-165. https://doi.org/10.1177/1747021820959286

Lovibond, P.F., Chow, J.Y.L., Tobler, C. & Lee, J.C. (in press, accepted 15/4/2022). Reversal of inhibition by no-modulation training but not by extinction in human causal learning. *Journal of Experimental Psychology: Animal Learning and Cognition*. https://doi.org/10.1037/xan0000328

Lovibond, P.F. & Lee, J.C. (2021). Inhibitory causal structures in serial and simultaneous feature negative learning. *Quarterly Journal of Experimental Psychology*, *74*, 2165-2181. https://doi.org/10.1177/17470218211022252

Lovibond, P. F., Preston, G. C., & Mackintosh, N. J. (1984). Context specificity of conditioning, extinction, and latent inhibition. J*ournal of Experimental Psychology: Animal Behavior Processes*, *10*, 360-375. https://doi.org/10.1037/0097-7403.10.3.360

Lubow, R. E. (1965). Latent inhibition: Effects of frequency of nonreinforced pre-exposure of the CS. *Journal of Comparative and Physiological Psychology, 60*(3), 454-457. https://doi.org/10.1037/h0022576

Lubow, R. E., & Moore, A. U. (1959). Latent inhibition: The effect of nonreinforced pre-exposure to the conditional stimulus. *Journal of Comparative and Physiological Psychology, 52*(4), 415-419. https://doi.org/10.1037/h0046700

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276– 298. https://doi.org/10.1037/h0076778

Miguez, G., McConnell, B., Polack, C. W., & Miller, R. R. (2018). Proactive interference by stimuli presented without outcomes: Differences in context specificity of latent inhibition and conditioned inhibition. *Learning & Behavior, 46*(3), 265-280. https://doi.org/10.3758/s13420-017-0306-x

Pavlov, I. P. (1927). *Conditioned Reflexes*. London, UK: Oxford University Press.

Pearce, J. M. & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532–552.

Rescorla, R.A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin, 72,* 77–81. https://doi.org/10.1037/h0027760

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black & W.F. Prokasy (Eds.), *Classical conditioning II: Current research and theory, 2*, 64-99.

Urcelay, G.P., Perelmuter, O. & Miller, R.R. (2008). Pavlovian backward conditioned inhibition in humans: Summation and retardation tests. *Behavioral Processes, 77*, 299-305. https://doi.org/10.1016/j.beproc.2007.07.003

Westbrook, R.F. & Bouton, M.E. (2010). Latent inhibition and extinction: Their signature phenomena and the role of prediction error. In Lubow, R.E & Weiner, I. (Eds) Latent Inhibition: Cognition, Neuroscience and Applications to Schizophrenia, pp. 23-39. New York: Cambridge University Press. 2010. https://doi.org/10.1017/CBO9780511730184.003