AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Animal Learning and Cognition

© 2025 The Author(s)
ISSN: 2329-8456

2025, Vol. 51, No. 1, 1–12
https://doi.org/10.1037/xan0000390

Symmetrical "Super Learning": Enhancing Causal Learning Using a Bidirectional Probabilistic Outcome

Santiago Castiello[1, 2, 3], Gabriella FitzGerald[1], Georgina M. Aisbitt[1], A. G. Baker[4], and Robin A. Murphy[1]
[1] Department of Experimental Psychology, University of Oxford
[2] Department of Psychiatry, Yale University
[3] Centro Universitario de Ciencias de la Salud, Universidad de Gudalajara
[4] Department of Psychology, McGill University

In a learning environment, with multiple predictive cues for a single outcome, cues interfere with or enhance each other during the acquisition process (e.g., Baker et al., 1993). Previous experiments have focused on cues that signal the presence or absence of binary outcomes. This introduces a perceptual and perhaps motivational asymmetry between excitatory and inhibitory learning. Here, using a bidirectional outcome, we asked whether learning about both generative (incremental positive outcome) and preventative (incremental negative outcome) causal cues show similar enhancement effects in opposite directions. In three experiments with humans using predictive learning tasks, participants ($N = 133$) were exposed to probabilistic predictive cues for opposite polarity events. Generative cues caused an increase in outcome likelihood, while preventative cues decreased it. An analysis of explicit predictive ratings found evidence for symmetrical learning and enhanced learning for both generative and preventative cues. The results are discussed in relation to super learning, an effect derived from theories of competitive learning based on error correction and theories of contrasting probability estimates.

Keywords: associative learning, bidirectional outcomes, super learning, potentiation, predictive learning task

Organisms learn about contingent regularities in their environments from the probabilistic relations between the cues and outcomes they experience (Baker et al., 1996). These regularities provide evidence for generating appropriate behavioral responses that map to the causal structure of their environment (Baker et al., 2005; Murphy et al., 2008). The regularities may be either positive (generative causes) in which the outcome is more likely to follow the cause, or they might be negative (preventive causes) in which the outcome is less likely to follow the cause.

In associative learning terminology, these relationships are called excitatory or inhibitory (Baker & Mackintosh, 1977). In terms of information or statistics, the two types of correlation excitatory (positive) and inhibitory (negative) are equally informative (e.g., Castiello et al., 2022; Murphy et al., 2022). One might expect therefore, that they would be equally likely to be learned about, indeed the processes might be expected to be symmetrical (Baker & Mackintosh, 1977), but this does not always seem to be the case (Chow et al., 2023; White, 2006) We sought to investigate possible symmetrical causal human learning between generative and preventative contingencies while equating for the perceptual differences in the cues and learned relations. By symmetrical, we refer to learning how generative and preventative contingencies may follow a similar learning trajectory, based on a similar absolute value of magnitude. It is possible that acquisition of both types of contingency would proceed at different rates. But if

Andrew R. Delamater served as action editor.
Santiago Castiello https://orcid.org/0000-0002-3672-1366
Robin A. Murphy https://orcid.org/0000-0002-8763-5062
The present studies were presented at the Associative Learning Symposium 2023 at Gregynog, United Kingdom (Castiello et al., 2023). Corpus Christi College, University of Oxford, supported Robin A. Murphy and Santiago Castiello de Obeso. Universidad de Guadalajara funded Santiago Castiello de Obeso's PhD with an institutional scholarship (V/2018/1476 and V/2021/989). The authors have no conflicts of interest to declare. A preprint is available in PsyArXiv: https://osf.io/preprints/psyarxiv/7yu23.

Open Access funding provided by University of Oxford: This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0; https://creativecommons.org/licenses/by/4.0). This license permits copying and redistributing the work in any medium or format, as well as adapting the material for any purpose, even commercially.

Conceptualization: Georgina M. Aisbitt, Robin A. Murphy, Santiago Castiello. Data curation: Santiago Castiello, Georgina M. Aisbitt. Formal analysis: Santiago Castiello. Funding acquisition: Robin A. Murphy, Santiago Castiello. Investigation: Santiago Castiello, Georgina M. Aisbitt, Gabriella FitzGerald. Methodology: Santiago Castiello, Georgina M. Aisbitt, Robin A. Murphy. Project administration: Robin A. Murphy, A. G. Baker. Resources: Robin A. Murphy, Santiago Castiello. Software: Santiago Castiello. Supervision: Robin A. Murphy, A. G. Baker. Validation: Georgina M. Aisbitt, Santiago Castiello, Robin A. Murphy. Visualization: Santiago Castiello. Writing–original draft: Santiago Castiello, Georgina M. Aisbitt, Robin A. Murphy. Writing–review and editing: Santiago Castiello, Gabriella FitzGerald, A. G. Baker, Robin A. Murphy.

Santiago Castiello contributed equally to conceptualization, visualization, formal analysis, and writing–original draft. Gabriella FitzGerarld contributed equally to conceptualization and writing–original draft and served in supporting role for visualization and formal analysis. Georgina M. Aisbitt contributed equally to conceptualization and writing–original draft. Robin A. Murphy served as lead for project administration and contributed equally to writing–original draft, visualization, and formal analysis. A. G. Baker served as lead for project administration and contributed equally for visualization and formal analysis.

Correspondence concerning this article should be addressed to Robin A. Murphy, Department of Experimental Psychology, University of Oxford, New Radcliffe House, Radcliffe Observatory Quarter, Oxford, OX2 6GG, United Kingdom. Email: robin.murphy@psy.ox.ac.uk

generative and preventive learning are symmetric with respect to a 0 axis, we might expect the learning to reflect a sort of mirror image.

There is often more than one possible cause of any outcome, and these causes or cues can interact physically and psychologically (Baker et al., 1993; Castiello et al., 2021; Vallée-Tourangeau et al., 1998). For example, the perceived causal power of individual cues seems to add up to the causal power of the compound of the cues. Alternatively learning about one cause can be interfered with by the presence of another valid cause of the same outcome. Under certain conditions, this leads to cue competition phenomena such as blocking (e.g., Kamin, 1968) or overshadowing (e.g., Ojeda-Aguilar et al., 2023) where causes compete for causal power and the presence of one cause reduces learning about the other. Associative models posit that cue-outcome relations are learned via an error correction process, applied to a link between the representations of the cue and outcome (e.g., Shanks & Dickinson, 1987; Vallée-Tourangeau et al., 1998; Wasserman et al., 1993). Descriptively, these models predict that generative, preventative, or zero covariation between cue and outcome will lead to excitatory, inhibitory, or weak links respectively (Rescorla, 1968; Sosa, 2024; Vallée-Tourangeau et al., 1998; Wasserman et al., 1993).

Much of the theoretical work on learning and its empirical base derives from animal experiments in which cues signal binary outcomes, i.e., the presence or absence of an event (e.g., animals either receive food or do not). However, with this form of binary outcome, there are perceptual and motivational asymmetries that may impact learning (White, 2006). While generative cues signal the presence of an outcome, preventative cues usually signal the absence of an outcome. One obvious asymmetry is that preventative learning about the predicted absence of an event requires the animal to learn that nothing happening has a specific meaning (Castiello et al., 2022)—as opposed to other times when nothing happening is an actual meaningless event. Learning that important events happen must precede learning that important events can be prevented from happening (Baker, 1974). From the perspective of learning about food occurrences, the normal background is food scarcity (the rat is hungry) so the presence of a food pellet is an important motivating event whereas food absence cannot motivate in the same way—at least when food is not otherwise expected. To address the asymmetry implied by binary outcomes, we moved away from cues signaling the absence or presence of an event. Instead, we studied predictors of increased or decreased likelihood of an outcome (i.e., increased and decreased store sales).

Cue competition effects, such as blocking and overshadowing, are well established (Dickinson et al., 1984) and are complemented by the finding that a strong preventative cue (negative predictor, i.e., inhibitory learning; Sosa, 2024) can not only weaken learning about other less effective negative predictors but also enhance learning about a new generative predictor. This latter effect is referred to as super-conditioning in animals or super learning in humans (Darredeau et al., 2009; Urushihara & Miller, 2017).

In super-learning experiments, a negative predictor of an outcome enhances learning about a novel generative predictor (Aitken et al., 2000; Darredeau et al., 2009; Urushihara & Miller, 2017). Learning this type of relations has been investigated using both deterministic (e.g., Baetu et al., 2005) and probabilistic relationships involving binary outcomes (e.g., Baker et al., 1993). Deterministic relations involve a cue or cause (C) that always signals either the presence or the absence of the outcome (O), for example, $p(O|C) = 1$ or 0. Nondeterministic, thus probabilistic, relations may be more analogous to real-world uncertainty in which cues signal an outcome with some probability but are imperfect predictors, for example, $p(O|C) \neq 1$ or 0.

In real-world situations, there is usually a degree of likelihood with which outcomes occur, thus, the real world is probabilistic. Some cues predict an increase in the likelihood of an outcome and others a decrease. Here, we frame our task using a probabilistic outcome that may increase or decrease in likelihood in the presence of a set of cues (presented alone or in the compound). This way both generative and preventative predictors have a similar and symmetrical weight.

Current Experiments

To test for symmetry between generative and preventative super learning, we established either a generative or preventative cue in the first phase and subsequently paired it in training with a preventative or generative cues of opposite polarity. The scenario in these experiments used either an increase or a decrease in store sales as the positive or negative outcome against an implied typical sales figure. Participants needed to learn which items seemed to cause an increase in sales and which caused a decrease in sales. The design of the experiments is shown in Table 1. The critical or target cues are presented as capital letters (e.g., *A* and *B*), while the extra control or nontarget cues are presented as lower case *u* (e.g., *v*, *w*, *x*, *y*, *z*). There were two trial types, with cues either presented individually (e.g., *u*) or together (e.g., *Au*). The probability of the positive outcome following each trial type is presented in parentheses. A positive outcome was an increase in sales. The negative outcome was a decrease in sales. There were no trials in which the "ice cream sales" did not change.

Table 1
Designs: Experiments 1, 2, and 3

		Phase 2		
Super-learning treatment	Phase 1	Super	Control**a** (E1 and E3)	Control**b** (E2 and E3)
Excitatory/generative	*Au* (.125)	*CA* (.875)	*Ew* (.875)	*Gy* (.875)
	u (.875)	*A* (.125)	*w* (.875)	*y* (.125)
Inhibitory/preventative	*Bv* (.875)	*DB* (.125)	*Fx* (.125)	*Hz* (.125)
	v (.125)	*B* (.875)	*x* (.125)	*z* (.875)

Note. Super = super-learning condition. Control**a** (E1 and E3) = control for Experiments 1 and 3. Control**b** (E2 and E3) = control for Experiments 2 and 3. Each cell with letters represents a trial type. Every letter is a cue type and the number inside the parenthesis is the probability of an increase in sales. *A* = inhibitor; *B* = excitor; *C* = super-excitatory cue; *D* = super-inhibitory cue; *E* and *F* = control cues, E1 and E3; *G* and *H* = control cues, E1 and E3; *u*, *v*, *w*, *x*, *y*, and *z* = nontarget or extra cues.

In the first phase, we pretrained cues either as strong predictors, $p(O|C) = 7/8 = .875$, of an outcome (increased ice cream sales), a generative cause, or with a weak likelihood of the outcome increase, $p(O|C) = 1/8 = .125$, a preventative cause. For these contingencies, the probability of a decrease was the complement of the probability of an increase (e.g., for *u* the probability of an increase was .875 and a decrease was .125). In the second phase, participants continued to experience these cues along with novel target cues (*C* and *D*). We were interested in whether the presence of the preestablished generative predictor (*B*; excitor) for an outcome would facilitate learning about a new cue that signaled a low likelihood that sales would increase (a preventative predictor; *D*) and whether the presence of an established preventative predictor (*A*; inhibitor) during training would enhance learning about a generative predictor (*C*). We assessed learning by asking participants to rate the relationship between cues and the outcome on a scale ranging from −100 to 100, where 0 meant that the cue did not provide information about the occurrence of an outcome.

Experiment 1

In Experiment 1 we used the design described in Table 1 with a go/no-go procedure: which involved a pretraining phase with generative and preventative cues, and in the second phase, the pretrained cues were presented in a compound with other cues and trained in the opposite direction to the pretraining contingency. To anticipate the findings, we demonstrated symmetrical learning and then used two further experiments to rule out potential explanations for our findings. Experiment 2 replicated the effect and incorporated an additional control condition. Experiment 3 replicated and tested the generality of the result using both control conditions and two-alternative forced-choice task rather than the go/no-go procedure.

Method

Participants

We recruited an opportunistic sample of university students. Those with self-identified color blindness were excluded. Forty-seven individuals (23 female, 24 male; age $M = 21.43$ years, $SD = 2.46$) participated for course credit or payment. Ethical approval for the study was granted by the Oxford University Research Ethics Committee (Reference MSD-IDREC-C2-2013-008). The sample size was based on a medium effect size and power of .80, estimated from our previous work in the lab on similar contingency learning although with no direct comparison with symmetrical super learning, we considered a similar effect from Experiment 1 in Darredeau et al. (2009) with a target sample of 48.

Design

Table 1 illustrates the design of the fully within-subject experiment. There were two phases during which participants were trained with a selection of the 16 different stimulus configurations. In Experiment 1, 12 trial types or stimulus configurations were used. The control cues *Ew*(.875) and *Fx*(.125) used in Experiment 1 were designed to make sure that the super compound cues *CA*(.875) and *DB*(.125) had the same absolute reinforcement schedule. The control condition used in Experiment 2 controls for the relative value or validity of the cues within the compound, for example, the experimental condition was *CA*(.875) and *A*(.125), and the control condition was *Gy*(.875) and *y*(.125). Experiment 3 contained both control treatments.

Half of the stimulus configurations consisted of individual elements (cues) in a cover story involving the sales of an ice cream shop. Participants were instructed to discover how sales depended on the "daily special" flavors on offer, for example, *u*(.875) from Table 1. The other half were compounds of one of the individual elements with another flavor, for example, *Au*(.125). Note that the number in parenthesis represents the proportion of trials that were paired with the outcome of increased sales. Thus, participants were tasked with six cue discriminations each designed to make the unique flavor in that discrimination into either a generative cause (*B* in *Bv* vs. *v*) or preventative cause (e.g., *A* in *Au* vs. *u*). The exceptions to this were the control cues used in Experiments 1 and 3 (see Table 1) which were pseudo-discriminations in which both the flavor compounds and individual flavors were followed by the same likelihood of an increase in sales, *Ew*(.825), *w*(.825) and *Fx*(.175), *x*(.175).

Capital letters indicate the target flavors (e.g., *C*) and lower-case letters indicate flavors that might have been expected to facilitate or interfere with learning of the target flavors. In both phases of Experiment 1 we balanced the number of individual and compound trials as predictors of increased and decreased likelihood of the outcome. Participants could differentiate the flavors based on their color, which was indicated by the ice cream flavors and an associated name (e.g., chocolate chip), and we used a set of 12 flavors in two configurations. The roles the flavors played in these two sets were randomized and the sets were counterbalanced so that half the participants received each set. Thus, half of the participants received one color-contingency set and the other half a second different set.

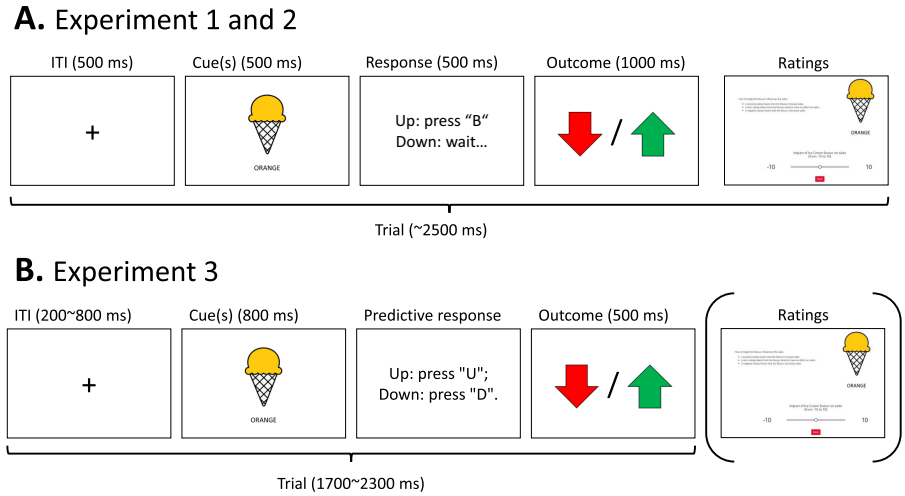
Procedure

The learning scenario was embedded in a go/no-go task requiring participants to learn when to respond or withhold a response based on the antecedent cues. There were eight presentations of each of the four trial types in Phase 1, and four presentations of each of the eight trial types in Phase 2. Ratings were recorded after each trial. The trial types were repeated in blocks where each trial type was presented once in random order in each block.

Prior to the presentation of the task, participants read a cover story in which they were told to imagine that they were monitoring the daily sales of an ice cream shop. Their goal was to evaluate whether the "daily special" flavor(s) resulted in increased or decreased sales (see also Aisbitt & Murphy, 2016). Each colored cue represented a flavor (both the color of the "scoop" and the flavor name were shown, see Figure 1A). Each trial began with the presentation of the daily special(s); either a single element (one ice cream flavor) or a compound (two flavors) for 500 ms. During that 500 ms, the participants were instructed to press the "B" key on the computer keyboard if they predicted that sales would increase with these flavors. If they believed sales would decrease, they were asked to withhold a response. Following either the pressing of "B" or the participant waiting out the 500 ms presentation, one of two feedback arrows was presented for 1,000 ms. This was to indicate the actual sales result (a green arrow pointing up indicated an increase; a red arrow pointing down indicated a decrease). Additionally, increased sales were accompanied by a high-frequency tone and decreased sales by a low-frequency tone, operationalizing outcome polarity.

After the response period, participants were shown some of the trained cues to rate as singletons (participants did not rate the compound cues) on a scale from −100 (negative) to +100 (positive). Participants responded to the question "How strongly does this flavor

Figure 1
Trial Structure



Note. (A) Experiments 1 and 2: ITI, cues' presentation, responses, outcome, and ratings. (B) Experiment 3: ITI, cues, responses, outcomes, and for some blocks of trials: contingency ratings (see Table 1). The rating screen in B is in parentheses because they were not required at the end of each trial or each block, but only at the end of the blocks: first, second, fourth, eighth, and 16th. ITI = inter-trial interval. See the online article for the color version of this figure.

influence the sales?" and were instructed as follows: (a) "A positive rating means that the flavor increases sales," (b) "A zero rating means that the flavor tends to have no effect on sales," and (c) "A negative rating means that the flavor decreases sales." After single-element presentations, they were asked to rate the cue they had just seen. On compound trials, they were asked to only rate the element that was unique to the compound. For example, after seeing cue *w* they would be asked to rate cue *w*. After seeing *Ew*, however, they would only be asked to rate flavor *E*. We also collected confidence ratings and response error rates but these provided no additional insights into learning in this experiment, and so are not included in the following analysis.

Statistical Analysis

In Phase 1, the dependent variables were the ratings of the four cues *A*, *B*, *u*, and *v* per block of trials. We conducted an analysis of variance (ANOVA), with causal direction (generative vs. preventative; 2), super-learning relevance, (*A* and *B*) versus extra cues (*u* and *v*); 2, and blocks of trials (eight) as within-subject factors. Having established whether participants learned about the cues in Phase 1, the ability of these cues to influence learning about other cues was tested in Phase 2. Participants quickly learned the new discriminations in Phase 2 so we report only the ratings for the first four blocks of trials.

In Phase 2, to compare the similarity of the effects between the generative and preventative cues, we ran a repeated measures ANOVA with the individual ratings of all preventative cues multiplied by −1 (i.e., the sign of the ratings was switched). If the generative and preventative curves were to be mirror images of one another, this operation should make them "fall on top" of one another. These ANOVAs investigated the following effects, with two causal directions, two cue types (super learning vs. control cues), and four trials. An α of .05 was used throughout and effect sizes (partial eta-squared, η_p^2) and 95% confidence intervals (CIs) are presented. Follow-up analyses used separate error terms rather than pooled error terms. We used Bonferroni corrections for post hoc tests.

Bayes Factor: Testing Symmetry Explicitly

Given that the symmetry effect requires accepting the null hypothesis that there is no effect of causal direction in the −1 transformed data, we tested whether the symmetry parameter (causal direction [generative/preventative]) in the ANOVA provided a better model by using a Bayes factor (BF) ratio (Morey & Rouder, 2021) against a reduced model without the symmetric parameter assumption. Thus, we compared the full model that assumes asymmetrical learning (β_3), that is, a different effect for each causal direction:

$$\text{adj. rating} \sim \beta_0 + \beta_1 \text{ blocks} + \beta_2 \text{ target} + \beta_3 \text{ direction} + \text{interactions}, \quad (1)$$

against the reduced model which assumes symmetry, i.e., without the causal direction parameter (β_3), as follows:

$$\text{adj. rating} \sim \beta_0 + \beta_1 \text{ blocks} + \beta_2 \text{ target} + \text{interactions}. \quad (2)$$

For the BF, the full model (Equation 1) was used as the denominator and the reduced model (Equation 2) was used as the numerator. A BF ratio > 1 supports symmetry, otherwise, BF < 1 supports asymmetry.

Software and Code

All analysis and visualizations were performed using R (R Core Team, 2021). Data processing packages were reshape2, hmisc, devtools, and dplyr (Harrell, 2021; Wickham, 2007; Wickham et al., 2021, 2022), and data visualization, ggplot2 (Wickham, 2016), and ggpubr (Kassambara, 2020). The BFs were implemented with the R package BayesFactor (Morey & Rouder, 2021). R scripts are available upon request.

Transparency and Openness Statement

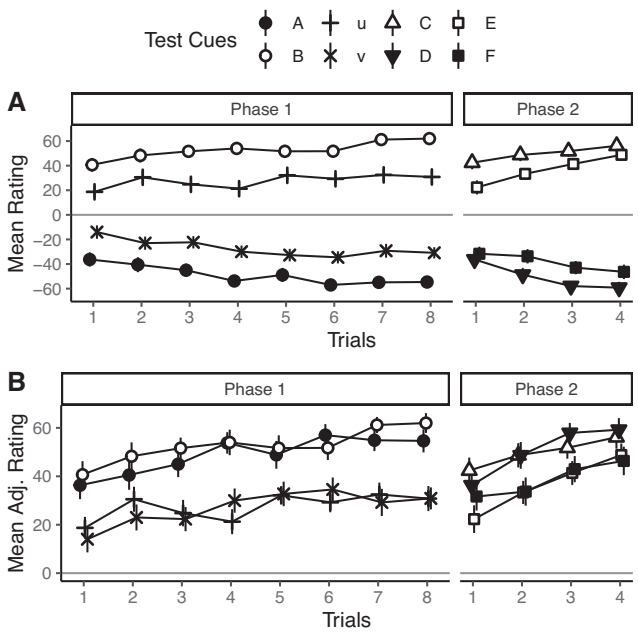
We cite all R packages that we use. The processed data and scripts available https://github.com/santiagocdo/symmetricalSuperLearning. R scripts to conduct cleaning, analysis, visualization, and simulations are available upon request. The code used to run Experiments 1 and 2 was performed in e-prime and Experiment 3 was coded in gorilla.sc and all are available upon request. We used APA style format across the whole document; sample sizes were based on previous work (Aisbitt, 2016). We did not preregister the predictions, but we would like to note that all predictions regarding super learning are planned predictions derived from the formal associative model we present. We did not preregister the conducted analysis. Experiment 2 replicates conditions from Experiment 1 and Experiment 3 systematically replicates parts of both Experiment 1 and Experiment 2 thus, all critical comparisons are replicated at least once.

Results

Phase 1: Acquisition

Figure 2A shows the raw means of the causal estimates with the standard error of the mean for Experiment 1. The judgments are generally consistent with the objective contingencies; *A* was rated negatively and *B* positively with evidence of learning across trials.

Figure 2
Experiment 1



Note. (A) Mean rating as a function of blocks, Phase 1 shown in the left panel and Phase 2 in the right panel. (B) Mean adjusted (adj.) ratings as a function of blocks, and phases in left and right panels. Adjusted ratings were scored by multiplying all inhibitory or preventative cues' ratings by −1. *A* cue is an inhibitor, *B* an excitor, *u* and *v* are extra cues, and *C* and *D* are the super-learned cues (super excitor *C* was paired with the inhibitor *A*; and super inhibitor *D* was paired with the excitor *B*). *E* and *F* are the control cues. See Table 1 for more details. The error bar represents the standard error of the mean. Adj. = adjusted.

This serves as a basic indication that the participants learned the compound/element discriminations in Phase 1. Figure 2B displays transformed data (preventative ratings multiplied by −1) and suggests that the learning was of similar absolute magnitude for both generative and preventative learning. ANOVAs confirmed these observations. There was a significant main effect of trials, $F(7, 322) = 7.16$, $p < .001$, $\eta_p^2 = .13$, 95% CI [.06, .19], and of target cues (*A* and *B* vs. extra cues), $F(1, 46) = 88.02$, $p < .001$, $\eta_p^2 = .66$, 95% CI [.49, .76]. None of the interactions nor the main effect for causal direction were reliable ($Fs < 1.42$). This is consistent with the explanation that learning of initial discriminations was similar for generative and preventative causes.

Phase 2: Symmetric Learning

Participants rated the super-learning cues (*C* and *D*) strongly for both preventative and generative cases (see Figure 2, Phase 2). These cues were also rated more strongly positive and negative compared with the control cues (*E* and *F*) which shared the same training but had not been compounded with a cue of opposite polarity in Phase 1. The repeated measures ANOVA show two reliable main effects: super learning versus control cues, $F(1, 46) = 11.38$, $p < .001$, $\eta_p^2 = .198$, 95% CI [.03, .39], and Blocks, $F(1, 138) = 17.5$, $p < .001$, $\eta_p^2 = .276$, 95% CI [.15, .38], but no effect of causal direction, $F(1, 46) = 0.179$, $p = .67$. No other interactions were significant, suggesting a symmetrical generative and preventative super-learning effect.

Finally, we tested the symmetry model against the asymmetry model and we found a BF[symmetry/asymmetry] > 100, suggesting strong evidence in favor of symmetrical super learning.

Discussion

In Experiment 1, consistent with our symmetrical super-learning hypothesis, opposite outcome polarity super-learning cues were learned similarly. Participants also rated the super-learning cues more positively or negatively than their respective control cues although both types of cues were associated in the compound with the same probability of a change in sales. An associative learning model (e.g., Rescorla & Wagner, 1972) explains super learning because the Phase 1 trained cues enhance learning about cues in Phase 2 because of an enhanced error correction signal. The pretrained cues in the compound with the Phase 2 cues create a larger prediction error when they are paired in the compound with the target cues. However, despite the evidence that the differences between the super learning and control cues were consistent from the beginning of Phase 2 and were not simply a function of training during Phase 2, as suggested by an earlier reviewer of this experiment that it is possible the differences reflect a suppression of estimates of the control cues (*E* and *F*) rather than an enhancement of the estimates of the experimental cues (*C* and *D*).

This alternative explanation that the apparent enhanced learning about *C* and *D* might involve reduced learning about the control cues *E* and *F* relies on the selective learning phenomenon called relative validity (Baker et al., 1993; Murphy et al., 2001a, 2001b; Wagner et al., 1968). This effect demonstrates that a target cue's acquisition of associative strength is weakened by its co-occurrence with a cue that is more predictive of an outcome. For example, cue *w* in the *Ew* versus *w* discrimination predicts more increased sales than *E* and thus might reduce judgments of the strength of *E,* because *w* is

relatively more valid. Therefore, the relative validity effect might have increased the likelihood of finding a significant difference between the generative super-learning cue (*C*) and the comparison control cue (*E*) that had little to do with enhanced learning of *C* but everything to do with a reduction in the learning of the control cue *E*. The same is true of the preventative super-learning cue and its comparison cue. It should be noted, however, that since no cues in the control condition are pretrained, the effect of the more valid cues on the less valid cues requires training to acquire the different validities. The effect therefore might be expected to be weak until the more valid cue acquires its associative strength. Our test phase involved few trials so the effect of relative validity would not be expected to be large. Nonetheless, it is worth pursuing the robustness and replicability of our super-learning effect. Experiment 2 includes a different control condition that controls for relative validity effects.

Experiment 2

Our interpretation of the results of Experiment 1 suggests that super-learning emerges from pretraining elements in the generative and preventative compound cues, rather than emerging as a function of training in Phase 2 alone. In Experiment 2 we performed a further test of this by equating the validity of the compound-trained cues in the second phase. In this experiment, *C* and *G* are both paired with cues that receive training with an outcome in the compound but not when presented alone, that is, *CA*(.875)/*A*(.125) and *Gy*(.875)/*y*(.125). Similarly, *D* and *H* are both presented with cues with the same contingency during Phase 2, that is, *DB*(.125)/*B*(.875) and *Hz*(.125)/*z*(.875). Consequently, the super-learning target cues (*C* and *D*) and control cues (*G* and *H*) receive the same training in Phase 2 (Table 1), the only difference is a lack of pretraining for the nontarget cues in the control condition. However, now the control cues (*G* and *H*) are predicted to acquire more associative strength than the control condition used in Experiment 1. While the symmetry is expected to be similar to that found in Experiment 1, the Rescorla–Wagner (RW) model (1972) predicts that the magnitude of any super-learning effect will be weaker in Experiment 2.

Method

Participants

Thirty-four participants took part in the study (24 female, 10 male; age $M = 21.67$ years, $SD = 1.96$). The recruitment process and ethical approval were the same as in Experiment 1.

Procedure

The trial structure, cover story, and method of stimulus presentations were as in Experiment 1. Changes to the design are outlined above and appear in Table 1, in the column headed "Super" and "Control**b** (E2 and E3)."

Statistical Analysis

As in the previous experiment, we conducted two ANOVAs, one for each phase. Then we used a BF to test evidence supporting the symmetry model.

Results

Phase 1: Acquisition

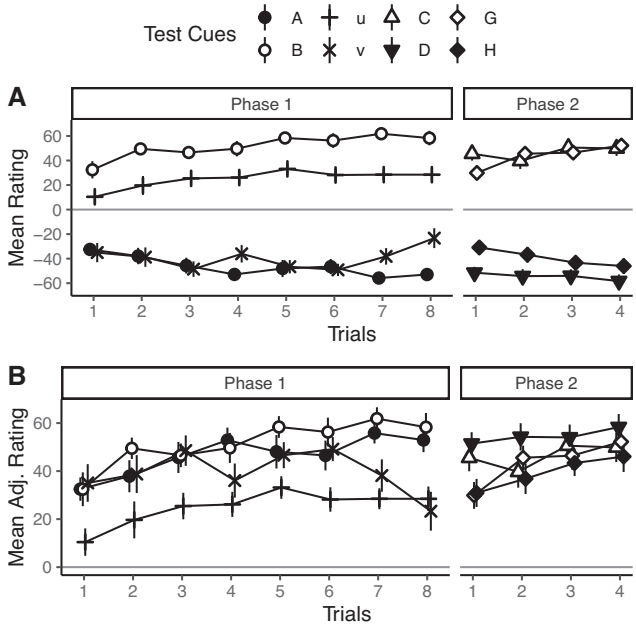
Figure 3 shows that the acquisition of the generative discrimination (*B* and *u*) proceeded in much the same manner as Experiment 1. The preventative discrimination (*A* and *v*) proceeded more slowly, however, by the end of Phase 1 it was of the same magnitude as the generative discrimination.

The statistical analyses conducted on the −1 transform data from Figure 3B were consistent with this interpretation. As in Experiment 1 there was a reliable main effect for target (*A* and *B*) versus extra (*u* and *v*) cues, $F(1, 33) = 69.06$, $p < .001$, $\eta_p^2 = .69$, 95% CI [.48, .79], and for trials, $F(7, 231) = 4.38$, $p < .001$, $\eta_p^2 = .12$, 95% CI [.03, .18]. However, unlike Experiment 1, the causal direction main effect was also reliable, $F(1, 33) = 4.18$, $p = .049$, $\eta_p^2 = .11$, 95% CI [.00, .33]. In addition, the interaction between causal direction and super learning was also reliable, $F(1, 33) = 10.83$, $p = .002$, $\eta_p^2 = .25$, 95% CI [.04, .47], suggesting that the generative and predictive learning in Phase 1 proceeded at different rates, however, by the end of Phase 1 similar levels of discrimination were found.

Phase 2: Symmetric Super Learning

As we found in Experiment 1 there was a super-learning effect and it was symmetrical, Figure 3 appears to show that the effect

Figure 3
Experiment 2



Note. (A) Mean rating as a function of blocks, Phase 1 shown in the left panel and Phase 2 in the right panel. (B) Mean adjusted (Adj.) ratings as a function of blocks, and phases in left and right panels. Adjusted ratings were scored by multiplying all inhibitory or preventative cues' ratings by −1. *A* cue is an inhibitor, *B* an excitor, *u* and *v* are extra cues, and *C* and *D* are the super-learned cues (super excitor *C* was paired with the inhibitor *A*; and super inhibitor *D* was paired with the excitor *B*). *G* and *H* are the control cues. See Table 1 for more details. The error bar represents the standard error of the mean. Adj. = adjusted.

appeared on only the first block of trials. However, the ANOVA carried out on these data does not justify this limited conclusion. In the analysis (Figure 3B) only the main effects of trials, $F(3, 99) = 4.17$, $p = .008$, $\eta_p^2 = .11$, 95% CI [.01, .22], and super learning versus controls were reliable, $F(1, 33) = 6.77$, $p = .014$, $\eta_p^2 = .17$, 95% CI [.01, .40]. Neither the main effect of causal direction nor the interactions were reliable. Only the interaction between trials and super learning had a p value below .1, $F(3, 99) = 2.239$, $p = .088$, $\eta_p^2 = .06$, 95% CI [.00, .16]. This omnibus analysis allows us to conclude that we found both generative and preventative super learning and these effects did not differ in magnitude thus replicating and extending the results of Experiment 1.

Finally, like Experiment 1, we tested the Bayesian symmetry model (without the causal direction parameter) against the asymmetry model (all factors as in ANOVA). We found a BF[symmetry/asymmetry] ratio > 100, suggesting strong evidence in favor of symmetrical super learning.

Discussion

The results of Experiment 2 provide evidence that the enhanced learning effects seen in Experiment 1 are at least partially dependent on the training the cues receive in both Phase 1 and Phase 2 and are not simply an example of relative validity causing a reduction in the learning to the control cues in Phase 2. One possibility for why the positive super learning only appears in the first trial is that the preventative discrimination from Phase 1, upon which the super-learning relies, seemed to be learned more slowly and was, perhaps, weaker, although by the end of Phase 1, no statistical differences in ratings were observed. Also, the positive super learning is dependent on the continued preventative training of the cue used to facilitate positive enhanced learning, $u(.875)$. In contrast, enhanced preventative learning was seen throughout Phase 2 in both experiments, suggesting that the enhanced negative judgments relied on the training with the generative cue from Phase 1, and not the continued generative training of the cue used to facilitate negative enhanced learning, $v(.125)$. This difference may suggest that a preventative (or inhibitory) association may extinguish more rapidly than the generative (or excitatory) association (see also Baetu & Baker, 2010; Sosa, 2024).

Experiment 1 had a clearer symmetric super-learning effect, while the effect was weaker in Experiment 2, maybe because of a not fully learned generative nor preventative relation in Phase 1. Thus, more Phase 1 training might be expected to observe super learning. Therefore, in Experiment 3 we combined the controls from Experiments 1 and 2 and added more training trials in Phase 1.

Furthermore, the rationale of these experiments is to look for parallels between preventive and generative causes, but the task used in these two experiments itself has both temporal and response asymmetries. In a go/no-go task, one option is paired with an overt response, and the other with withholding that response. Moreover, there is a difference in latencies between withholding the response and making the overt response (see also Shenoy & Yu, 2012). The trial moves on following the overt response whereas the participants wait out the 500 ms stimulus presentation when it is not made. To increase the comparability, in Experiment 3, a two-alternative forced-choice task was used in which the participants were asked to make an overt response to predict both increases and decreases in sales.

Experiment 3

In Experiment 3: (a) we used twice as many trials in Phase 1; (b) we incorporated both controls from the previous experiments; and (c) the task was not embedded in a go/no-go paradigm, but rather participants were asked to provide an answer to whether the shop's sales increased or decreased that day by using the "U" and "D" keys, respectively, in a two-alternative forced-choice paradigm (Figure 1B). We made these changes to avoid a possible lack of discrimination in Phase 1, to compare both controls employed in the previous experiments in a within-subject design, and to more carefully equate exposure time spent at the predictive response stage of the task. In this way, feedback was given at the same time point after generative and preventative trials, rather than preventative predictions requiring that the participant wait out the full 500 ms.

Method

Participants

Fifty-two participants took part in the study (28 female, 21 male, and three who chose not to disclose their gender; age $M = 27$ years, $SD = 12.9$). The recruitment process and ethical approval were the same as described previously. We used the target sample size of 48 consistent with E1 with 10% extra to account for online recruitment losses.

Procedure

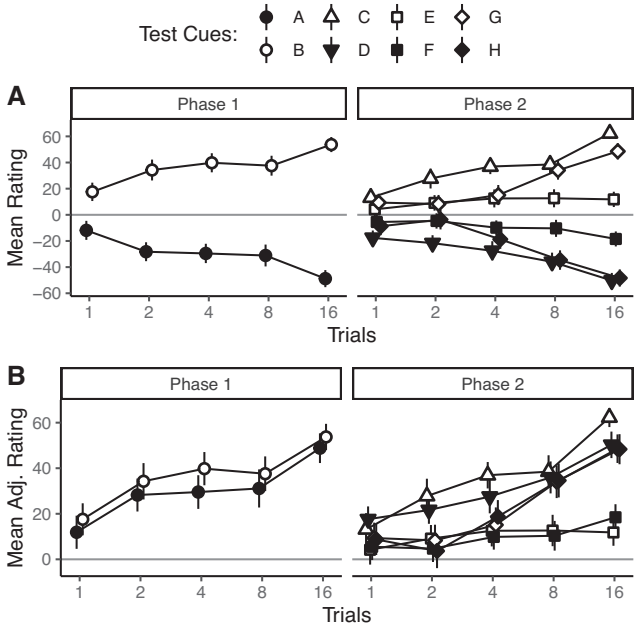
The trial structure, cover story, and stimulus presentations were the same as in the previous experiments. Changes to the design are outlined above and appear in Table 1. The experiment was run via the online platform gorilla.sc (Anwyl-Irvine et al., 2020). Another difference is that ratings were not collected at the end of each trial but rather following trials: 1, 2, 4, 8, and 16 (for both Phases 1 and 2). Each block of trials consisted of the fully random presentation of all trial types within that phase and ratings were collected for all trials at the end of that block of trials. For example, for Phase 1, the first block of trials consisted of four trial types: Av, v, Bu, and u. After these, participants provided ratings for A and B in random order following each of Trials 2, 4, 8, or 16. In Phase 2, each block of trials consisted of 12 trial types, and participants provided ratings after Trials 1, 2, 4, 8, and 16. For each participant, we used this procedure to add more learning trials with a reduced session length and to provide more sensitivity early in learning, when changes should be more rapid.

Results and Discussion

The raw ratings and adjusted ratings for Phase 1 training are displayed in Figure 4A (raw) and 4B (adjusted). As for previous experiments, we analyzed the adjusted rating values (i.e., negative/preventative cues multiplied by -1). Generative and preventative cues (A and B) were discriminated in Phase 1 and became strongly positive and negative. The statistical analysis corroborates this impression (Figure 4B). In the ANOVA of the -1 transformed scores, only the main effect of trials was reliable, $F(4, 204) = 5.6$, $p < .001$, $\eta_p^2 = .1$, 95% CI [.02, .17], neither the effect for causal direction nor the interaction was reliable (Fs < 2.6).

In Phase 2, cues were paired in the compound with the target super-learning cues (C and D) in the compounds: CA and DB. We

Figure 4
Experiment 3



Note. (A) Mean rating as a function of blocks, Phase 1 shown in the left panel and Phase 2 in the right panel. (B) Mean adjusted (Adj.) ratings as a function of blocks, and phases in left and right panels. Adjusted ratings were scored by multiplying all inhibitory or preventative cues' ratings by −1. *A* cue is an inhibitor, *B* an excitor, *u* and *v* are extra cues, and *C* and *D* are the super-learned cues (super excitor *C* was paired with the inhibitor *A*; and super inhibitor *D* was paired with the excitor *B*). *E*, *F*, *G*, and *H* are the control cues. See Table 1 for details. The error bar represents the standard error of the mean. Adj. = adjusted.

observed that in early trials the super-learning cues showed stronger acquisition for both the generative (*C*) and preventative (*D*) contingencies than compared with either control condition. The full ANOVA had three factors: five trial ratings, two causal directions (generative/preventative), and three super learning and control conditions (super|control**a** E1|control**b** E2; see Table 1). This analysis revealed two main effects, trials, $F(4, 204) = 20.02$, $p < .001$, $\eta_p^2 = .28$, 95% CI [.18, .37], and conditions, $F(2, 102) = 18.3$, $p < .001$, $\eta_p^2 = .26$, 95% CI [.12, .39]. There was a reliable interaction between trials and conditions, $F(8, 408) = 5.22$, $p < .001$, $\eta_p^2 = .09$, 95% CI [.03, .14]. Neither other main effects nor interactions were significant ($Fs < 0.5$). The lack of a causal direction effect supports the hypothesis that generative and preventative super learning are symmetrical.

The previous interaction implies an overall difference between at least one of the three conditions. To explore this interaction further, we ran follow-up ANOVAs for each trial. We found that the difference between the two causal directions was never significant, thus generative and preventative training does not differ in absolute magnitude. There were no differences on Trial 1 but there were on all subsequent trials, second, $F(2, 51) = 6.17$, $p < .01$, fourth, $F(2, 51) = 6.54$, $p < .01$, eighth, $F(2, 51) = 10.13$, $p < .001$, and 16, $F(2, 51) = 35.55$, $p < .001$. No interactions between causal direction and conditions were found ($Fs < 2.46$).

To test whether the control conditions (as in Experiments 1 and 2) differed from the target super-learned cues, we ran two separate follow-up ANOVAs, that is, 5 trials × 2 causal direction (generative/preventative) × 2 condition (super vs. control**a** or control**b**). The first was between the super-learned cues and control cues as in Experiment 1, and the second was between super-learned cues and control as in Experiment 2. The results are as follows.

The first ANOVA, comparing super learning versus control**a** conditions, revealed the same significant effects as the previous analysis. Two main effects, trials, $F(4, 204) = 11.62$, $p < .001$, $\eta_p^2 = .19$, 95% CI [.09, .27], and super learning, $F(1, 51) = 25.06$, $p < .001$, $\eta_p^2 = .33$, 95% CI [.13, .50], and their interaction, $F(4, 204) = 6.95$, $p < .001$, $\eta_p^2 = .12$, 95% CI [.04, .20], differed reliably. No other effects, $Fs < 0.97$. The second ANOVA, comparing the conditions super learning versus control**b** only revealed two main effects, trials, $F(4, 204) = 20.92$, $p < .001$, $\eta_p^2 = .29$, 95% CI [.18, .38], and super learning, $F(1, 51) = 11.00$, $p < .01$, $\eta_p^2 = .18$, 95% CI [.03, .36], but no other main or interaction effects, $Fs < 1.7$.

The interaction for the first ANOVA between trials and super learning (*C* and *D*) versus control**b** (*E* and *F*) suggests a super-learning effect. The main effect of the second ANOVA suggests that even with 16 trials in Phase 2, if Phase 1 is long enough symmetric super learning is found. Overall, Experiment 3 supports super learning for both positive and negative ratings; hence, symmetrical.

Finally, as in previous experiments, to test that the lack of causal direction effects in this experiment implies that generative and preventative learning enhancement is symmetrical, we calculated a BF. We compared a symmetry model—no causal direction parameter—against the asymmetry model—including a causal direction parameter—and obtained a BF supporting the symmetry hypothesis, BF (symmetry/asymmetry) > 100.

General Discussion

We present three experiments that used a causal learning scenario involving a modified go/no-go task (Experiments 1 and 2) and a two-alternative choice task (Experiment 3) to assess generative and preventative causal associative learning. We investigated whether prior experience could enhance learning about other cues and whether these effects were symmetrical for generative and preventative cues. The results of all three experiments support the hypothesis that generative and preventative causal learning is symmetric. Throughout, participants' ratings reflected an ability to form generative and preventative associations as expressed by causal ratings.

The results show how pairing novel cues with established positive and negative cues influences learning about these novel cues. We used an enhanced or super-learning procedure (Rescorla, 1971), which provides an indirect means of assessing generative and preventative learning. Rather than a direct test of learning, this learning is demonstrated by the ability to enhance learning to a secondary cue. This method differentiates inhibition, the active suppression of excitatory learning (see Sosa, 2024), from the failure to learn about a stimulus. Across the experiments, we replicated findings of generative enhanced learning in humans (Lovibond et al., 1988) and showed that this super-learning effect can be induced using a go/no-go paradigm (Experiments 1 and 2) and in a two-choice forced task (Experiment 3). The facilitated or super-learning performance is a particularly important prediction of models of active inhibition (Rescorla, 1971; Williams & McDevitt, 2002).

Positive super learning is the enhanced generative learning about a novel cue (C) that occurs when that novel cue is paired with an inhibitor or preventative cue (A). Models of learning based on error correction like the RW model argue that the acquisition of associative strength reflects the error term: $(\lambda - \sum V)$, where λ represents the value of the actual outcome, and $\sum V$ represents the value of the expected outcome. According to this, if the expectation or associative strength is negative as occurs in inhibitory learning; (e.g., $\sum V = -1$), but the outcome does occur ($\lambda = 1$), the error term will be large, $(\lambda - \sum V) = 1 - (-1) = 2$, compared to a novel cue that generates no expectations ($\sum V = 0$) when it is paired with the positive outcome ($\lambda = 1$). Indeed, this effect has previously been reported in both animals (Rescorla, 1971) and humans (Urushihara & Miller, 2017; Williams & McDevitt, 2002). Furthermore, this is the first demonstration of enhanced preventative learning, in which a previously trained excitatory or generative cue can facilitate preventative learning about a novel cue (D) when the two are subsequently trained in a preventative compound (DB). This finding has implications for the theoretical conceptualization of inhibition because it is consistent with the symmetry in enhanced learning predicted by Rescorla and Wagner (1972). As a result, these findings suggest that there may be a symmetry of generative predictors (i.e., excitors) and preventative predictors (i.e., inhibitors), and it is for future research to explore the extent of this possibility.

Furthermore, this work extends the analysis of selective learning effects with bidirectional outcomes (Baetu & Baker, 2010; Melchers et al., 2006). There was reliable reasonably symmetrical generative and preventative super learning compared to the control stimuli in each experiment. This expectation was predicted by the RW model (see below for an account of the predictions). The reader will recall that there was concern that relative validity might contribute to a perceived super-learning effect found in Experiment 1 by reducing the value of the control cues. Experiment 2 eliminated this relative validity interpretation as the sole cause of the critical difference. A direct comparison of the control treatments is a test of the question of whether relative validity influenced the judgments of the control cues. We conducted Experiment 3 to test both controls in a within-subjects design. Thus, there is little evidence that our symmetrical super-learning effects are an artifact of absolute (control**a** from Experiment 1) or relative (control**b** from Experiment 2) validity effects between the two experiments.

Notes on Learning Theory and Simulations

The implications for using a probabilistic bipolar rather than a binary (0,1) outcome for a theory of learning, such as the RW model, are worth elaborating. Evidence suggests that symmetry is more likely to emerge with bidirectional outcomes (Chow et al., 2023). Thus, in our current work and following simulations we used a bidirectional outcome (i.e., + and −; see Murphy et al., 2011). If an outcome is a Bernoulli probabilistic event (i.e., it can have a λ value of 1 or 0), λ can be expressed as: $\lambda = p(\text{O}|\text{C}) = p = .875$. For the simulations described here, we replaced λ with the asymptotic value given its probability of that outcome in both positive and negative forms. The standard method of modeling the reward (λ) in the RW model involves trial-by-trial sampling. In this way λ takes the value of 1 or 0 depending upon the presence or absence of the outcome (i.e., $\lambda_{\text{probabilistic}}$). Formally this can be represented as a Bernoulli distribution, Be[p]; with a parameter p determining the success, and the sampling outcome is the result of a Bernoulli experiment with a binary outcome:

$$\lambda_{\text{probabilistic}} \sim \text{Be}[p(\text{O})], \tag{3}$$

where $\sim$ indicates that $\lambda_{\text{probabilistic}}$ is distributed as the right side of the equation; and Be[p(O)] is a Bernoulli distribution with probability of success or "increased sales" equal to the probability of outcomes, that is, p(O). Next, we simulated the effects with bidirectional outcomes using the RW model, and without using the mean of multiple random runs. We note too that these simulations match those generated by the standard trial-by-trial simulation process (e.g., Murphy et al., 2001a).

In our current experiments, there were two possible outcomes states: "increased sales" and "decreased sales." We assumed that the change in sales is one dimension (i.e., sales change), hence, our model used one outcome element with two symmetric and opposite values encoded with λ. We coded "increased sales" with $\lambda = 1$, and "decreased sales" with $\lambda = -1$. In addition to encoding the outcome bilaterally, we conducted the simulation using an analytic solution for the probabilistic nature of the task rather than using the mean of many runs of simulations with probabilistic λ (Equation 3). Thus, here we fixed the λ for each trial type by using a probability weighing of λ approach:

$$\lambda_{\text{fixed}} = [p(\text{O}) \times 2] - 1, \tag{4}$$

where p(O) is the probability of "increased sales" (in Table 1 represented as .875 or .125). For example, in the RW model calculation for a trial with 0.875 (e.g., trial type E) probability of increased sales, then the value of λ_{fixed} is:

$$\lambda_{\text{fixed}} = [p(\text{O}) \times 2] - 1 = [0.875 \times 2] - 1 = 0.75, \tag{5}$$

and for a trial with a 0.125 (e.g., trial type I) probability of increased sales:

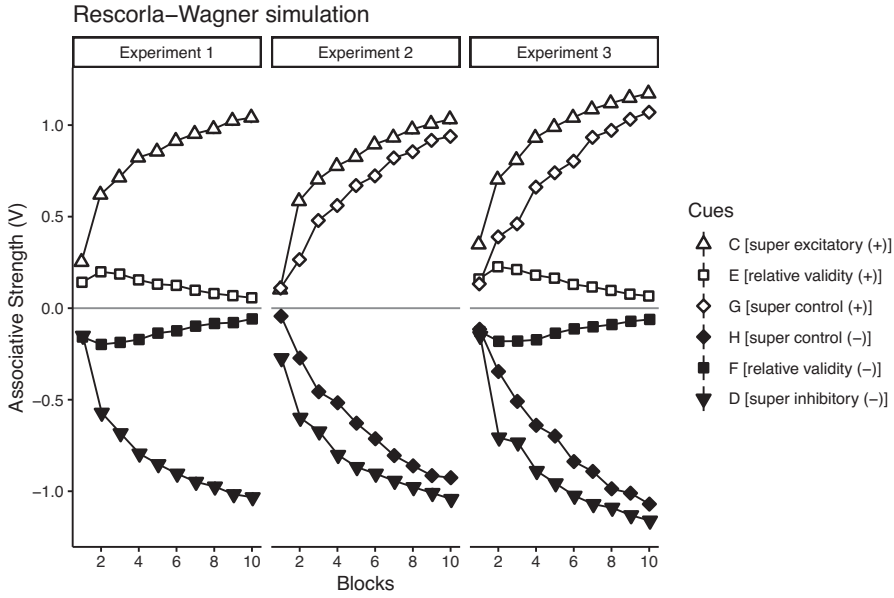
$$\lambda_{\text{fixed}} = [p(\text{O}) \times 2] - 1 = [0.125 \times 2] - 1 = -0.75. \tag{6}$$

This probability-weighted λ_{fixed} value captures the probabilistic and symmetric aspects of the task and provides an alternative (and computationally efficient) method for solving the model for probabilistic outcomes.

In the next section, we used the above λ_{fixed} rule to model the three experiments. The simulations shown in Figure 5 illustrate how the associative strength for the generative (B) and preventative (A) predictors in an initial learning phase (not shown) enhanced subsequent learning from novel cues, that is, generative (C) and preventative (D) super learning. Figure 5 shows the associative strength of individual cues in a second phase in which two pairs of compound target cues (CA and DB) were simultaneously trained with the cues trained in Phase 1 and compared with control cues (E and F; and G and H) that were paired with novel cues (w and x, and y and z). In Phase 2 the two control cues (E and F) receive the same training experience as the super-learning cues (C and D), the difference between them is the previous experience of the compounding cues (u and v). The strength of associations for C and D would be expected to develop faster in the presence of pretrained opposite polarity cues (see Figure 5). For the target cues shown in the right panel (Experiment 3), the predictions illustrate that during the initial training trials, C and D acquire generative and preventative associative strength but notice that while the control cues initially acquire

Figure 5

The Three Panels Show Three RW Model Simulations for Each Experiment



Note. The presented panels are Phase 2 of a symmetrical super-learning procedure. The parameters for the Rescorla and Wagner (1972) model are $\alpha = .2$ and $\beta = 1$, λ values described in the General Discussion section. Each block corresponds to two trial types and each line with shape corresponds to the associative strength of a particular cue (see Table 1). RW model = Rescorla–Wagner model.

associations with the outcome, *E* and *F* are ultimately blocked from acquiring associative strength (Baker et al., 1996). It is worth noting that in Experiment 1 cues *E* and *F*, trained with (*w* and *x*) respectively, were novel in Phase 2 but received consistent training as predictors for increased or decreased outcomes. This has two potential implications (a) the associative strengths of *w* and *x* might generalize to our control cues (via within compound associations) thereby enhancing them relative to the target cues and undermining our attempt to find a super-learning effect, or (b) *w* and *x* may come to strongly overshadow or block learning about the control cues thereby seemingly enhancing our evidence for super learning. Given the findings and these comparisons, there is good evidence for these predictions from the simulations.

Conclusions

To conclude, we sought to determine whether we could induce preventative causal learning (i.e., associative inhibition, *AX*−, *X*+), and to test whether this type of preventative learning (or inhibition) could be used to induce enhanced learning in a symmetrical manner to that found with positive or generative causal cues and ascertain whether these effects were symmetrical about generative and preventative cues. All three experiments support these hypotheses, demonstrating that participants can acquire preventative associations and can use these associations to induce enhanced learning effects. Also, these effects are symmetrical for generative cues, leading to the facilitation of enhanced preventative learning, and generative cues, leading to the facilitation of enhanced generative learning. We found these effects with explicit judgments and anticipate that this learning could explain differences in action rates or instrumental choice, although there was no indication in these experiments that response error rates reflected these differences. Finally, the potential to magnify learning through these enhancement procedures, extends the range of effects that an associative model based on cue competition can be applied.

References

Aisbitt, G. M. (2016). *An associative account of impaired inhibition in psychopathy development of the psychopathy attention theory* [Unpublished DPhil]. University of Oxford.

Aisbitt, G. M., & Murphy, R. A. (2016). An application of a theory of attention (Mackintosh, 1975) to psychopathy: Variability in the associability of stimuli. In J. B. Trobalon & V. D. Chamizo (Eds.), *Associative learning and cognition: Homage to Professor N. J. Mackintosh* (pp. 89–108). Edicions de la Universitat de Barcelona. https://dialnet.unirioja.es/servlet/articulo?codigo=6893078

Aitken, M. R. F., Larkin, M. J. W., & Dickinson, A. (2000). Super-learning of causal judgements. *The Quarterly Journal of Experimental Psychology B*, *53*(1), 59–81. https://doi.org/10.1080/027249900392995

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Baetu, I., & Baker, A. G. (2010). Extinction and blocking of conditioned inhibition in human causal learning. *Learning & Behavior*, *38*(4), 394–407. https://doi.org/10.3758/LB.38.4.394

Baetu, I., Baker, A. G., Darredeau, C., & Murphy, R. A. (2005). A comparative approach to cue competition with one and two strong predictors. *Learning & Behavior*, *33*(2), 160–171. https://doi.org/10.3758/BF03196060

Baker, A. G. (1974). Conditioned inhibition is not the symmetrical opposite of conditioned excitation: A test of the Rescorla–Wagner model. *Learning*

and Motivation, *5*(3), 369–379. https://doi.org/10.1016/0023-9690(74)90018-6

Baker, A. G., & Mackintosh, N. J. (1977). Excitatory and inhibitory conditioning following uncorrelated presentations of CS and UCS. *Animal Learning & Behavior*, *5*(3), 315–319. https://doi.org/10.3758/BF03209246/METRICS

Baker, A. G., Mercier, P., Vallée-Tourangeau, F., Frank, R., & Pan, M. (1993). Selective associations and causality judgments: Presence of a strong causal factor may reduce judgments of a weaker one. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 414–432. https://doi.org/10.1037/0278-7393.19.2.414

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

Baker, A. G., Murphy, R. A., & Vallée-Tourangeau, F. (1996). Associative and normative models of causal induction: Reacting to versus understanding cause. In D. R. Shanks, K. Holyoak, & D. L. Medin (Eds.), *Causal learning* (pp. 1–45). Academic Press. https://psycnet.apa.org/record/2003-00368-001

Castiello, S., FitzGerarld, G., Aisbitt, G. M., Baker, A. G., & Murphy, R. A. (2023). *Symmetrical "super learning": Enhancing learning in both directions using a bidirectional probabilistic outcome*. Associative learning symposium 2023, Oral Presentation at Gregynog, United Kingdom.

Castiello, S., Miller, R. R., Witnauer, J. E., Alcaide, D. M., Fung, E., Pitliya, R. J., Morrissey, D. K. C., & Murphy, R. A. (2022). Benefiting from trial spacing without the cost of prolonged training: Frequency, not duration, of trials with absent stimuli enhances perceived contingency. *Journal of Experimental Psychology: General*, *151*(8), 1772–1792. https://doi.org/10.1037/xge0001166

Castiello, S., Zhang, W., & Delamater, A. R. (2021). The retrosplenial cortex as a possible "sensory integration" area: A neural network modeling approach of the differential outcomes effect in negative patterning. *Neurobiology of Learning and Memory*, *185*, Article 107527. https://doi.org/10.1016/j.nlm.2021.107527

Chow, J. Y. L., Lee, J. C., & Lovibond, P. F. (2023). Inhibitory learning with bidirectional outcomes: Prevention learning or causal learning in the opposite direction? *Journal of Cognition*, *6*(1), Article 19. https://doi.org/10.5334/JOC.266

Darredeau, C., Baetu, I., Baker, A. G., & Murphy, R. A. (2009). Competition between multiple causes of a single outcome in causal reasoning. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*(1), 1–14. https://doi.org/10.1037/A0012699

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

Harrell, F. E., Jr. (2021). *Hmisc: Harrell miscellaneous*. https://CRAN.R-project.org/package=Hmisc

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Aversive stimulation* (pp. 9–31). University of Miami Press.

Kassambara, A. (2020). *ggpubr: "ggplot2" Based Publication Ready Plots*. https://CRAN.R-project.org/package=ggpubr

Lovibond, P. F., Siddle, D. A. T., & Bond, N. (1988). Insensitivity to stimulus validity in human Pavlovian conditioning. *The Quarterly Journal of Experimental Psychology*, *40B*(4), 377–410. https://doi.org/10.1080/14640748808402331

Melchers, K. G., Wolff, S., & Lachnit, H. (2006). Extinction of conditioned inhibition through nonreinforced presentation of the inhibitor. *Psychonomic Bulletin & Review*, *13*(4), 662–667. https://doi.org/10.3758/BF03193978

Morey, R. D., & Rouder, J. N. (2021). *BayesFactor: Computation of Bayes factors for common designs*. https://CRAN.R-project.org/package=BayesFactor

Murphy, R. A., Baker, A. G., & Fouquet, N. (2001a). Relative validity effects with either one or two more valid cues in Pavlovian and instrumental conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*(1), 59–67. https://doi.org/10.1037/0097-7403.27.1.59

Murphy, R. A., Baker, A. G., & Fouquet, N. (2001b). Relative validity of contextual and discrete cues. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*(2), 137–152. https://doi.org/10.1037/0097-7403.27.2.137

Murphy, R. A., Mondragón, E., & Murphy, V. A. (2008). Rule learning by rats. *Science*, *319*(5871), 1849–1851. https://doi.org/10.1126/science.1151564

Murphy, R. A., Schmeer, S., Vallée-Tourangeau, F., Mondragón, E., & Hilton, D. (2011). Making the illusory correlation effect appear and then disappear: The effects of increased learning. *Quarterly Journal of Experimental Psychology*, *64*(1), 24–40. https://doi.org/10.1080/17470218.2010.493615

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 a free lunch! *Journal of Experimental Psychology: General*, *151*(1), 41–64. https://doi.org/10.1037/xge0000910

Ojeda-Aguilar, Y. L., Burgos, J. E., García-Leal, O., & Buriticá, J. (2023). Ensombrecimiento posterior al preentrenamiento de cada EC en redes neurales artificiales y ratas [Postpretraining overshadowing of each CSs in artificial neural networks and rats]. *Acta Comportamentalia*, *31*(3), 441–465. https://doi.org/10.32870/ac.v31i3.86449

R Core Team. (2021). *R A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

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), 1–5. https://doi.org/10.1037/h0025984

Rescorla, R. A. (1971). Variation in the effectiveness of reinforcement and nonreinforcement following prior inhibitory conditioning. *Learning and Motivation*, *2*(2), 113–123. https://doi.org/10.1016/0023-9690(71)90002-6

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 (Ed.), *Classical conditioning II* (pp. 64–99). Appleton-Century-Crofts.

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), *Psychology of learning and motivation: Advances in research and theory* (Vol. 21, No. C, pp. 229–261). Academic Press. https://doi.org/10.1016/S0079-7421(08)60030-4

Shenoy, P., & Yu, A. J. (2012). *Strategic impatience in Go/NoGo versus forced-choice decision-making*. Advances in Neural Information Processing Systems (Vol. 25, pp. 2132–2140). https://papers.nips.cc/paper_files/paper/2012/hash/6d70cb65d15211726dcce4c0e971e21c-Abstract.html

Sosa, R. (2024). Conditioned inhibition, inhibitory learning, response inhibition, and inhibitory control: Outlining a conceptual clarification. *Psychological Review*, *131*(1), 138–173. https://doi.org/10.1037/rev0000405

Urushihara, K., & Miller, R. R. (2017). Causal superlearning arising from interactions among cues. *Journal of Experimental Psychology: Animal Learning and Cognition*, *43*(2), 183–196. https://doi.org/10.1037/XAN0000137

Vallée-Tourangeau, F., Murphy, R. A., Drew, S., & Baker, A. G. (1998). Judging the importance of constant and variable candidate causes: A test of the power PC theory. *The Quarterly Journal of Experimental Psychology Section A*, *51*(1), 65–84. https://doi.org/10.1080/713755745

Wagner, A. R., Logan, F. A., & Haberlandt, K. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology*, *76*(2, Pt.1), 171–180. https://doi.org/10.1037/h0025414

Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: Role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(1), 174–188. https://doi.org/10.1037/0278-7393.19.1.174

White, P. A. (2006). The causal asymmetry. *Psychological Review*, *113*(1), 132–147. https://doi.org/10.1037/0033-295X.113.1.132

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12), 1–20. https://doi.org/10.18637/jss.v021.i12

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., & Müller, K. (2022). *dplyr: A Grammar of data manipulation*. https://CRAN.R-project.org/package=dplyr

Wickham, H., Hester, J., Chang, W., & Bryan, J. (2021). devtools: Tools to make developing R packages easier. In *CRAN* (R package Version 2.4.3).

Williams, B. A., & McDevitt, M. A. (2002). Inhibition and superconditioning. *Psychological Science*, *13*(5), 454–459. https://doi.org/10.1111/1467-9280.00480

Received March 25, 2019
Revision received November 21, 2024
Accepted November 21, 2024 ■

Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you "social psychology" is not sufficient—you would need to specify "social cognition" or "attitude change" as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

APA now has an online video course that provides guidance in reviewing manuscripts. To learn more about the course and to access the video, visit http://www.apa.org/pubs/journals/resources/review-manuscript-ce-video.aspx.