Journal of Experimental Psychology: Animal Learning and Cognition
2015, Vol. 41, No. 1, 18–31

© 2014 American Psychological Association
2329-8456/15/$12.00 http://dx.doi.org/10.1037/xan0000047

Automaticity and Cognitive Control in the Learned Predictiveness Effect

Lauren T. Shone, Irina M. Harris, and Evan J. Livesey
University of Sydney

In novel contexts, learning is biased toward cues previously experienced as predictive compared with cues previously experienced as nonpredictive. This is known as learned predictiveness. A recent finding has shown that instructions issued about the causal status of cues influences the expression of learned predictiveness, suggesting that controlled, volitional processes play a role in this effect. Three experiments are reported further investigating the effects of instructional manipulations on learned predictiveness. Experiment 1 confirms the influence of inferential processes, extending previous work to suggest that instructions affect associative memory as well as causal reasoning. Experiments 2 and 3 used a procedure designed to tease apart inferential and automatic contributions to the bias by presenting instructed causes that were previously predictive and previously nonpredictive. The results demonstrate that the prior predictiveness of cues influences subsequent learning over and above the effect of explicit instruction. However, it appears that the relationship between explicit instruction and predictive history is interactive rather than additive. Potential explanations for this interactivity are discussed.

Keywords: learned predictiveness, stimulus associability, reasoning, automaticity

An important question facing research in associative learning is the nature of the relationship between learning and attention. The idea that behavior is biased not only by features of the world that hold intrinsic salience (Wolfe & Horowitz, 2004), but also by their acquired characteristics (Le Pelley, 2010), is critical in describing how a subset of available information in the environment is selected for action in a given situation. Accordingly, many models of associative learning (e.g., Kruschke, 2001; Mackintosh, 1975; Pearce & Hall, 1980) accept that the stimulus selection necessary for the acquisition of associative relationships is influenced by some form of selective attention. Such models share the same basic assumption that the attention devoted to a stimulus is flexible and governed by its past utility in predicting events.

For example, in his original model, Mackintosh (1975) proposed that the attention devoted to a cue increases if that cue is a more effective predictor of an observed outcome relative to other cues available at the same time. Formally, this is incorporated within model predictions as an interdependent relationship between stimulus associability and associative strength. Here, associability, or α, describes the property of a stimulus that determines the rate at which it will enter into associations. Associability changes according to the associative strength between a cue and the outcome with which it is paired. If that associative strength is high, that is, if a cue is a reliable predictor of the correct outcome, then associability will increase. If a cue is a poor predictor of the outcome, because its associative strength is relatively low, associability will decrease. Thus, the rate of learning provides an index of attentional change whereby attention to a stimulus is maintained and preferentially supports further learning provided it remains a good predictor relative to the other cues present at the same time.

Evidence suggesting that learned attention varies according to changes in associability comes from experiments in which past predictive utility biases learning in a novel situation. A robust example, first reported by Le Pelley and McLaren (2003; see also Lochmann & Wills, 2003), is the learned predictiveness effect. The basic experimental design used to demonstrate this effect is shown in Table 1. In the original demonstration, participants completed a causal learning task in which they were asked to play the role of a doctor in order to learn which of a variety of foods led to allergic reactions in a fictitious patient. Each trial in the first phase consisted of a compound of two food cues, leading to one of two allergic reactions. The critical manipulation was that each compound consisted of one perfectly predictive cue, represented by A–D, and one nonpredictive cue, W–Z. For example, Cue A consistently predicted the presence of Outcome 1 (O1), and therefore had perfect predictive utility. Alternatively, W had no predictive utility because it was equally predictive of both O1 and O2.[1]

In Phase II, a novel patient was introduced and again participants were required to learn the causal relationship between food cues and their respective allergic reactions. The same cues, in novel combinations, served to predict the occurrence of these

This article was published Online First October 20, 2014.

Lauren T. Shone, Irina M. Harris, and Evan J. Livesey, Department of Psychology, University of Sydney.

A portion of this work is reported in Shone and Livesey (2013), as part of the proceedings of the annual meeting of the Cognitive Science Society, 2013. This work was supported by an APA Scholarship awarded to Lauren T. Shone. Irina M. Harris was supported by an ARC Future Fellowship (FT0992123).

Correspondence concerning this article should be addressed to Lauren T. Shone, School of Psychology, University of Sydney, Sydney, 2006, Australia. E-mail: lsho0771@uni.sydney.edu.au

[1] It could be argued that describing W as nonpredictive is not entirely accurate, given that it predicts the presence of an outcome (either O1 or O2) on every trial. However, in line with prior literature on learned predictiveness, "nonpredictive" here indicates that a given cue is equally paired with both available outcomes during Phase I and therefore provides no benefit in making correct predictions during this phase. That is, participants would be expected to demonstrate chance performance during training if using the cue W to make predictions.

Table 1
A Basic Learned Predictiveness Design (e.g., Le Pelley & McLaren, 2003)

Phase I	Phase II	Test
AW–O1	AY–O3	AC
AX–O1	BZ–O4	WY
BW–O2	CW–O3	BD
BX–O2	DX–O4	XZ
CY–O2		
CZ–O2		
DY–O1		
DZ–O1		

Note. Letters (A–D and W–Z) refer to individual food cues. O1–O4 refer to four outcomes.

novel reactions. In this phase, both components of the compound discrimination were equally predictive. That is, both A and W shared the same objective relationship with O3 and O4, respectively. The critical difference between components was their status as a predictive or nonpredictive cue in the initial phase of learning. The test phase assessed learning by means of novel compounds. Half of these compounds consisted of previously predictive components that signaled the same outcome during Phase II (AC and BD), and the remainder consisted of previously nonpredictive components signaling the same outcome during Phase II (WY and XZ). Test ratings, as shown in Figure 1, revealed that previously predictive compounds were more easily paired with their correct outcomes, suggesting better learning for these cues.

This effect, consistently replicated across various scenarios (see Le Pelley, 2010, for a recent review), is consistent with models that predict attentional change according to the mechanisms of associative competition (e.g., Le Pelley, 2004; Mackintosh, 1975; Pearce & Mackintosh, 2010). According to such models, attention will be biased toward predictive cues A–D and away from nonpredictive cues W–Z throughout Phase I. This means that cues A–D will have an advantage when entering into new associations during the second phase, resulting in the observed preference in learning for these cues. Importantly, the hypothesized process does not rely on a deliberate attempt by the individual to control attention in a biased fashion according to the nature of the Phase I relationships. Instead, given the proposed relationship between associative history and associability, this explanation relies on the assumption that the attentional processes underlying the emergence of this bias are relatively automatic. However, this assumption has recently been questioned. Indeed, selective attention is associated with a variety of cognitive mechanisms (see Le Pelley, 2010; Wright & Ward, 2008, for reviews), raising the question of which processes are critically involved in the effect.

In demonstrations of learned predictiveness, there is often a high degree of conceptual similarity between the scenarios across the two phases of learning. One possibility, therefore, is that the effect is governed by a simple heuristic arising from inferential reasoning. It may be that participants make the explicit assumption that the predictive utility of Cues A–D will transfer across similar contexts resulting in the controlled, volitional selection of those cues throughout the second phase of learning (Mitchell, Griffiths, Seetoo, & Lovibond, 2012). Such an explanation accords well with claims that associative learning relies exclusively on propositional knowledge (e.g., Mitchell, de Houwer, & Lovibond, 2009).

Consistent with this possibility is evidence suggesting that learned predictiveness is susceptible to manipulation of inferred beliefs. In their Experiment 2, Mitchell, Griffiths, Seetoo, and Lovibond (2012) directly manipulated inferential beliefs across the two phases of a learned predictiveness design by way of instruction. At the onset of the second phase, participants in their continuity condition were explicitly instructed that the same cues would be relevant. Alternatively, those in the change condition were instructed the opposite, that previously predictive cues were now irrelevant. Critically, this condition revealed a complete reversal of the effect whereby more was learned about the relationship between previously irrelevant cues and the novel outcomes. That learned predictiveness is sensitive to variations in explicit reasoning suggests a role for controlled, volitional processes in explaining the effect.

There is, however, evidence to suggest that the presence of the inference alone is not sufficient to produce the learned predictiveness effect. For example, Le Pelley et al. (2010a) investigated the expression of learned predictiveness adopting a procedure in which the use of higher order reasoning was encouraged. In this demonstration, the critical relationships were explained to participants in written statements highlighting the summary of cue-outcome pairings as well as the frequency with which they occurred. Such a manipulation should indeed strengthen the ability of participants to engage in deductive reasoning, given the minimal demands on working memory. Interestingly, they failed to observe learned predictiveness under these conditions. The bias was only observed when the relevant information was presented in trial and error form across multiple trials. This is contrary to what would be expected if explicit causal attribution were the sole mechanism responsible for the effect.

Similarly, related attentional effects in predictive learning appear to be inconsistent with an explanation based solely on inferential processes. For example, in a two-stage causal learning paradigm similar to that of learned predictiveness, Le Pelley,

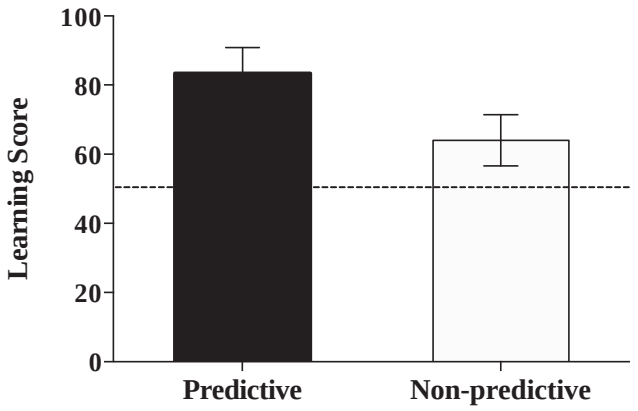


Figure 1. Learning scores reflect a linear transformation of the data reported by Le Pelley and McLaren (2003) in order to remain consistent with the measures used in the current study. Scores range from 0 to 100, with a score of 50 indicative of chance, as represented by the dotted line. These are shown for test compounds consisting of predictive components and test compounds consisting of nonpredictive components. Error bars represent standard error of the mean difference between predictive and nonpredictive cues.

Mitchell, and Johnson (2013) observed an attentional bias toward cues previously experienced as predictors of a high value outcome. However, instructions issued following training appeared to bias attention toward cues previously experienced as predictors of low value outcomes. Thus, opposing influences of training and instruction on learned attentional responses have been found. Further, effects such as the inverse base-rate effect show that attentional biases can lead to nonrational decisions, counter to what would be expected on the basis of inferential reasoning (Wills, Lavric, Hemmings, & Surrey, 2014). Taken together these findings raise the possibility that learned predictiveness reflects the operation of a combination of inferential and noninferential processes.

Although the design of Mitchell et al.'s (2012) reversal experiment confirms that the effect of instructed attention is strong under certain conditions, it is not well-equipped to test whether automatic processes also contribute to the learned predictiveness bias. At test, Mitchell et al. (2012) used a learning score measure that combines cue-outcome recall and the extent to which that same cue is judged to be causal. The use of this measure in learned predictiveness experiments is certainly not uncommon, nor is the use of causal ratings in associative learning research more generally. One can make a defensible argument that the strength with which a cue is associated with an outcome serves as an important source of evidence when making judgments about the causal relationship between them, especially when no further information is provided that might bear on this relationship. However, when a more compelling source of evidence, such as an explicit instruction, is present, there are strong reasons to question the link between the strength of learning and the strength of a causal judgment. Given that a set of cues was explicitly emphasized as important in each condition, it is not surprising that participants responded accordingly, providing higher causal ratings for cues instructed as likely to be causal. This might be viewed as nothing more than demand characteristics, or as a genuine form of rational causal attribution that does not necessarily reflect the strength of learning. For example, in the change condition, learning might still be weaker for W, a previously nonpredictive cue. Thus, at test, the association between W and the outcome with which it was paired, O3, would be weak. However, given the nature of the instruction, that is, that W is likely to be causal, participants might provide a higher rating in order to reflect the requirements of the task. By the same logic, a previously predictive cue such as A might be given a very low causal rating even if the participant remembers its associated outcome very well, because they have been instructed that the cue probably does not cause the effect.

Certainly, it is unsurprising that explicit instruction about causation affects causal reasoning. However, given the introduction of this instruction, one would assume the relationship between associative strength and causal inference to change. Indeed, the finding of Le Pelley, Mitchell, and Johnson (2013) suggests that instruction can affect causal learning, though importantly in a way that is dissociable from prior training. This suggests that associative history and instructed inference may operate as distinct bases of causal judgment. Given that the major theories that offer an explanation of the learned predictiveness effect (e.g., Mackintosh, 1975) are theories of associative learning and not causal reasoning, a different measurement approach is needed when instructional manipulations are used, one that assesses the strength of associative retrieval and the strength of causal judgment separately.

The aim of the present experiments was to test whether these sources of bias could be distinguished within learned predictiveness in order to further gauge the relative contribution of automatic and controlled processes in producing the effect. To this end, we explored instructional manipulations of learned predictiveness in which cue-outcome recall and causal attribution were assessed separately at test. Thus, the test phase consisted of two components across all experiments. Cues were presented individually and rated on both a memory and causal inference question. The memory question probed the extent to which participants could accurately pair a given cue with its associated outcome, while the causal rating tested the extent to which a cue was believed to cause an allergic reaction, independent of the knowledge of the identity of that reaction. Given that instructional manipulations were administered throughout the study, one might expect a measure of causal inference to reflect that manipulation. That is, if you tell a participant that a specific food is likely to cause an allergic reaction, then their rating on a causal scale should reflect that instruction.

However, according to interpretations of learned predictiveness that rely on changes in processing according to associative strength, such as that offered by the Mackintosh (1975) model, the benefit for previously predictive cues in Phase II relies on the ability of those cues to become associated with specific outcomes. The recall measure, therefore, should provide an indication of that association. If a dissociation between these measures can be demonstrated, such that a residual bias from Phase I training can be observed in recall despite a direct correspondence between instruction and causal ratings, this would implicate automatic attentional processes in producing the learned predictiveness bias. Such measures have been used successfully in related causal learning paradigms examining the blocking effect (e.g., Mitchell, Lovibond, & Gan, 2005; Mitchell, Lovibond, Minard, & Lavis, 2006), but have not previously been applied to learned predictiveness.

Experiment 1 attempted to replicate the finding by Mitchell et al. (2012) making use of the same instructional manipulation, albeit with a different cover scenario. To anticipate, we failed to find a complete reversal, instead observing no difference across both test ratings according to predictiveness in the change condition. Thus, learned predictiveness was abolished instead of reversed. Subsequent experiments introduced an orthogonal manipulation between Phase I predictiveness and instruction. If automatic biases are evident then one might expect to see differences in recall between previously predictive and nonpredictive cues within the same instructional condition, despite causal ratings reflecting the instructional manipulation more directly.

Experiment 1

The aim of Experiment 1 was to replicate the instructed reversal of learned predictiveness reported by Mitchell et al. (2012). Participants completed the allergist causal learning task (Le Pelley & McLaren, 2003) according to the learned predictiveness training structure outlined in Table 1. In this task, participants were asked to observe the allergies of a fictitious patient in order to predict which foods were causing various allergic reactions. At the start of Phase II, participants were told that they would be observing a new patient, suffering from different allergies. Further, one group of participants, those in the continuity condition, were told that it was likely that both patients were allergic to the same foods. Those in

the change condition were instructed that the two patients likely suffered from allergies to different foods. In line with the findings of Mitchell et al. (2012), we anticipated that learning would be sensitive to the instructions issued at the start of Phase II. However, given the modifications to the test phase whereby associative memory and causal judgment were assessed independently at test, a different pattern of responding across these measures was expected. If the instructional manipulation indeed serves to control the extent to which participants reason about the causal nature of specific cues, then one would expect to see this inference reflected in the causal ratings. This measure, therefore, should show a complete reversal of learned predictiveness according to instruction. Alternatively, accuracy in the recall measure relies more specifically on the ability to identify the outcome with which a cue was paired. Although this should still be sensitive to the instructions given at the start of Phase II, demonstrating an overall benefit in recall for cues instructed as causal, an automatic bias may be evident in better memory for cue-outcome relationships for previously predictive cues (A–D) compared with previously nonpredictive cues (W–Z).

Method

Participants and apparatus. Forty-eight first year psychology students from the University of Sydney (27 female, mean age = 19) were tested individually for Experiment 1 in exchange for partial course credit. The experiment was conducted on Apple Mac Mini computers attached to a 17-in. CRT monitor, running software programmed in PsychToolbox for Matlab (Brainard, 1997; Pelli, 1997).

Stimuli. Eight photographic pictures of food items, consisting of: coffee, fish, lemon, cheese, eggs, garlic, bread, and peanuts, were selected as stimuli in the experiment. These were randomly allocated for each participant to serve as predictive (A–D) and nonpredictive (W–Z) cues. All cues measured 6.9° × 6.9° of visual angle at a viewing distance of approximately 70 cm and were presented in pairs in the upper half of the screen separated by 9.5°. Four allergic reactions were randomly allocated to serve as the outcomes (O1–O4). These were: headache, nausea, rash, and fever. Outcomes appeared in text format in the lower half of the screen.

Procedure. After being randomly allocated to either the continuity or change conditions, participants were informed that the task required them to take the role of an allergist in order to discover the allergens of a fictitious patient. They were told that on every trial they would observe two foods that the patient, Mr. X, had eaten. On being shown the foods, participants were required to predict which of two allergic reactions would occur. Each self-paced selection was immediately followed by feedback for the duration of 1 s stating whether the prediction was correct or incorrect, as well as providing the actual allergic reaction experienced.

Phase 1 consisted of the eight trial types shown in Table 1. Each of these was presented twice in each of the eight blocks of trials. Within each block, the order of trials was randomized, and the left and right position of the cues was counterbalanced.

At the start of Phase II, participants were told that they now had a new patient, Miss Y, and as before would be required to learn about which foods were causing which allergic reactions. They were then issued with one of the following sets of instructions:

> Mr. X and Miss Y [*are/are not*] allergic to the same foods. Therefore, it is highly [*likely/unlikely*] that the foods that controlled which reaction Mr. X suffered in the last phase will also influence which reactions Miss Y suffers in this phase.

The instructions in italics differentiate between the two conditions, such that those in the continuity group were told that the two patients shared the same allergens, and those in the change group were told that the patients were allergic to different foods. Participants completed eight blocks of trials, with each block consisting of two of the four component discriminations shown in Table 1. Again, the order of trials was randomized and the location of cues counterbalanced.

A test phase was administered immediately following Phase II. All cues were presented individually and in randomized order throughout this phase. Participants were first issued with the following set of instructions emphasizing the difference between the recall and causal ratings:

> Every food that Miss Y ate (during the second phase of the experiment) was only ever followed by one type of allergic reaction. However, this does not necessarily mean that the food caused the allergic reaction . . . Indicate which allergic reaction followed after the food was eaten by Miss Y. This is a bit like a memory test to see how well you recall which foods and which reactions went together . . . [Then] indicate to what degree you think this food caused the allergic reaction in Miss Y.

On each trial, participants were first required to indicate whether the cue had been paired with Outcome 3 or Outcome 4. This was done by making a rating on a linear analogue scale, labeled *definitely goes with* [Outcome 3] on the left anchor, and *definitely goes with* [Outcome 4] on the right anchor. The midpoint of the scale was made explicit with the label *no idea.* Once this rating had been made, another scale appeared asking participants to rate whether that cue was causal. This scale was labeled *definitely does not cause the reaction* on the left, and *definitely does cause the reaction* on the right. Each rating scale yielded a value out of 100.

Finally, a manipulation check was included to ensure participants had encoded the instructions at the start of Phase II. Participants were presented with both sets of instructions and required to report which of those applied to their patient. There were no exclusions on the basis of this check, all participants having correctly indicated the set of instructions corresponding to the condition they had completed.

Results

Given that the critical predictions throughout this article rely on sufficient acquisition of the initial component discriminations, an additional exclusion criterion was adopted whereby participants who failed to achieve 60% accuracy across the final quarter of the first phase were excluded from further analysis. This is consistent with previous learned predictiveness protocols (Le Pelley & McLaren, 2003) and was applied to all three experiments. No participants were excluded on the basis of this criterion in Experiment 1.

Phase I. In order to assess initial acquisition, the accuracy of predictions was averaged across the eight compound trial types for each block. Participant accuracy increased steadily across training. A mixed measures analysis of variance (ANOVA) with Block (1–8) and Condition (continuity vs. change) as factors revealed a significant main effect of Block, $F(7, 322) = 65.37$, $p < .001$, $\eta_p^2 = .59$, but no significant effect of Condition, and no Block × Group interaction, $Fs < 1$, providing no evidence of initial group differences. Phase I and Phase II learning curves are shown in Figure 2.

Phase II. Again, prediction accuracy was collapsed across trial types and increased steadily for both groups across Phase II. A mixed measures ANOVA with Training Block (1–8) and Group (continuity vs. change) as factors showed a significant effect of Block, $F(7, 322) = 81.48$, $p < .001$, $\eta_p^2 = .64$. Accuracy for the change group was slightly better than for the continuity group, though this difference failed to reach conventional levels of significance, $F(1, 46) = 3.94$, $p = .053$, $\eta_p^2 = .08$, and there was no significant Block × Group interaction, $F(7, 322) = 1.88$, $p = .07$, $\eta_p^2 = .04$.

Test phase.

Memory ratings. For each cue, the memory rating was recoded as a score out of 100, where 100 reflects the most confident possible choice of the correct outcome, and 0 reflects the most confident possible choice of the wrong outcome. Thus, 50 on this scale is representative of chance. Scores were averaged according to whether they were predictive (A–D) or nonpredictive (W–Z) in Phase I. These are shown for the continuity and change conditions in the upper panel of Figure 3.

Scores were subjected to a mixed measures ANOVA with Group (continuity vs. change) and Predictiveness (predictive vs. nonpredictive) as factors. There was no main effect of Group, $F < 1$, and the main effect of Predictiveness failed to reach conventional significance, $F(1, 46) = 3.49$, $p = .07$, $\eta_p^2 = .07$, despite a slight benefit in memory for predictive cues. However, as suggested by Figure 1, this resulted from a significant Group × Predictiveness interaction, $F(1, 46) = 8.79$, $p = .004$, $\eta_p^2 = .16$. This was explored using simple effects analysis, which revealed that memory scores for predictive cues were higher than nonpredictive cues in the continuity condition, $F(1, 23) = 11.51$, $p = .002$, $\eta_p^2 = .33$. The scores for the predictive and nonpredictive cues did not differ significantly in the change condition, $F < 1$.

Causal ratings. Causal ratings were averaged according to Phase I predictiveness and are shown in the lower panel of Figure 3 for the continuity and change conditions. A mixed measures ANOVA with Group (continuity vs. change) and Predictiveness (predictive vs. nonpredictive) as factors showed a significant effect of Predictiveness, $F(1, 46) = 6.22$, $p = .016$, $\eta_p^2 = .12$, whereby predictive cues were given significantly higher causal ratings overall, as well as a significant Group × Predictiveness interaction, $F(1, 46) = 14.96$, $p < .001$, $\eta_p^2 = .25$. The main effect of Group did not reach significance, $F < 1$. The simple effects analysis investigating the interaction revealed that the pattern of causal ratings mirrored that of the memory ratings. That is, the difference in causal ratings between predictive and nonpredictive cues was significant in the continuity condition, $F(1, 23) = 16.56$, $p < .001$, $\eta_p^2 = .42$. This difference was not statistically significant in the change condition, $F(1, 23) = 1.21$, $p = .28$, $\eta_p^2 = .05$.

Discussion

Experiment 1 provides a partial replication of Mitchell et al. (2012). Overall, a clear effect of instruction on learned predictiveness was observed on both associative memory and causal ratings. In the continuity condition, recall was better for cues instructed as important, that is, previously predictive cues, compared with nonpredictive cues. This was also observed in the causal ratings, whereby predictive cues were rated as more likely to be causal compared with nonpredictive cues. However, learned predictiveness was abolished rather than reversed in the critical change condition. That is, there was no difference in recall between previously predictive and previously nonpredictive cues, and no difference in the extent to which these cues were considered as causal of an allergic reaction when participants were told that nonpredictive cues were informative for the second phase.

These results further validate the influence of voluntary processes on learned predictiveness. However, they provide an important extension to those of Mitchell et al. (2012) by demonstrating that the instructed reversal changes the strength of associative memory as well as the extent to which cues are judged as causal. A fairly significant caveat, however, is that our reversal was incomplete in the change condition. On the basis of the current design, this lack of complete reversal may be attributable to several

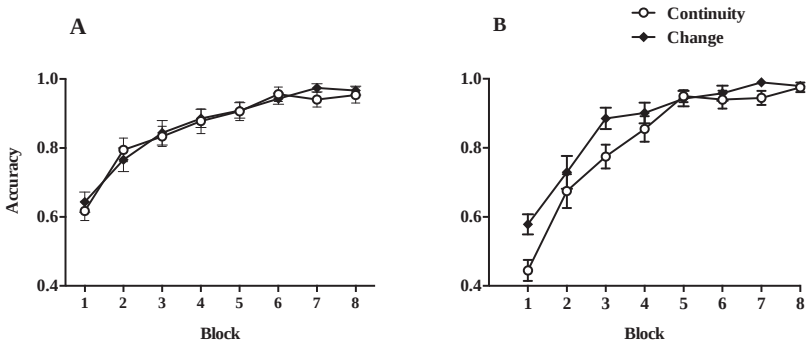


Figure 2. Mean accuracy during training for Phase I (A) and Phase II (B) for the continuity and change conditions in Experiment 1. Error bars indicate *SEM*.

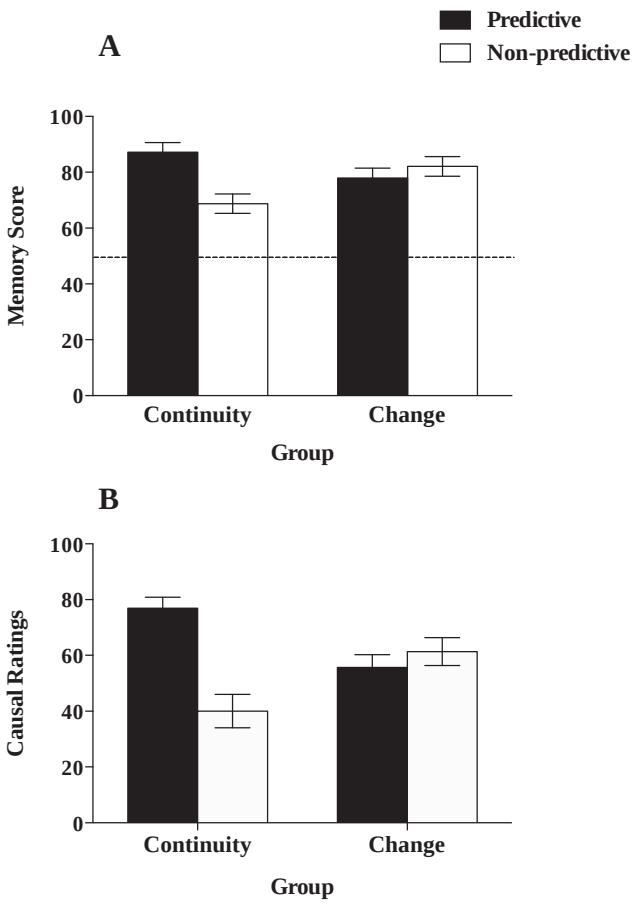


Figure 3. Memory scores (A) and causal ratings (B) for the continuity and change conditions for previously predictive and previously nonpredictive cues in Experiment 1. Those in the continuity condition were instructed that previously predictive cues were likely to be relevant, and those in the change condition were instructed that previously nonpredictive cues were likely to be relevant. Error bars represent standard error of the mean difference between predictive and nonpredictive cues. The dotted line in Panel A represents chance.

factors. One possibility is that the instructional manipulation increases difficulty in the change condition. If more is learned about the predictive cues in Phase I, these may still be needed for confirming the identity of cues that are important in Phase II. In order to know which cue is the allergen on any given trial, participants in this condition may use a strategy in which the known allergen from Phase I is identified and excluded. This is an additional process that is not necessary in the continuity condition. Alternatively, the results in the change condition might reflect competition between opposing inferential and automatic processes. If both of these processes are in operation, then an automatic attentional bias toward cues A–D conflicts with the instructed explicit inference favoring cues W–Z. Although these two explanations are not necessarily mutually exclusive, the question remains as to why they might be evident here and not in the results of Mitchell et al. (2012).

It is worth noting that this discrepancy arises from a different pattern of results in the change condition across the two experiments. In the change condition of Mitchell et al. (2012), learning was better for previously nonpredictive cues compared with previously predictive cues. In our results, there was no evidence for differences in learning across these two cue types. Given our null result in the critical change condition, one potentially informative analysis is a Bayes Factor (BF) calculation indicating whether the null hypothesis (no difference between predictive and nonpredictive cues) is more likely than the alternative hypotheses given the data. Employing a method suggested by Rouder, Speckman, Sun, Morey, and Iverson (2009), a BF was calculated for the *t* test comparing cue types in our change condition. As Rouder et al. suggest, we used the JZS prior that assumes a Cauchy distribution of effect sizes for the alternative hypothesis (see Rouder et al., 2009 for further detail). Odds of 3 to 1 in favor of the alternative (BF = 0.33) or 3 to 1 in favor of the null (BF = 3) are widely considered to constitute moderate evidence in favor of one hypothesis over the other. Our analysis yielded a BF of 4.7, suggesting the null hypothesis is 4.7 times more likely than the alternative, providing moderate evidence favoring the conclusion that no instructed reversal of the learned predictiveness effect was observed.

Further, there was no differentiation between recall and causal ratings. This is somewhat surprising in light of related causal learning research showing that causal judgments and prediction judgments are sensitive to different kinds of information about the relationship between a cue and an outcome (Vadillo & Matute, 2007; Vadillo, Miller, & Matute, 2005). However, although this suggests that different judgments at test are sensitive to different kinds of information about the relationship between a cue and an outcome, there are some important differences between these previous protocols and the one used here. For example, the prediction judgment employed by Vadillo et al. (2005) requires participants to predict the likelihood that an outcome will occur given a cue. Although this may rely on recall to some degree, it requires a judgment regarding probability of a future event. This is distinct from our recall measure which requires a judgment about the prior co-occurrence of a cue and an outcome, independent of what might happen if the cue were presented again in the future.

Given the introduction of an explicit instruction about causality, whether a dissociation between our measure of recall and causal attribution is anticipated by an inferential account of learned predictiveness, such as that favored by Mitchell et al. (2012), is unclear. While a more detailed discussion of the limitations of such an explanation will be examined in the General Discussion, given the absence of a complete reversal in either measure, it seems unlikely that learning is completely under the control of inferential reasoning. In order to assess stimulus associability independently of causal reasoning, a clear instruction is needed that separates knowledge of causality from Phase I bias. Because Experiment 1 established that a reversal instruction had a significant effect on the memory for cue-outcome pairings, Experiment 2 sought more definitive evidence of an automatic effect of Phase 1 predictive validity, that is, an effect that cannot be reasonably attributed to the instruction itself. The reversal design is not well-equipped to test this possibility because, as is evident in Experiment 1, even when resistance to reversal is observed it might be attributed to several mitigating factors, such as the reversal instruction just being inherently more complex.

Experiment 2

Experiment 2 was designed to further tease apart the involvement of inferential and automatic processes. This was achieved by introducing an orthogonal manipulation between Phase I predictiveness and instruction, the design of which is shown in Table 2. The first phase of training, in which participants learnt about the allergies of an initial patient, was identical to that of Experiment 1. Again, instructions about the causal status of foods as allergens for a new patient were issued before the start of the second phase. In Experiment 1 participants were either told that the same cues were likely to be relevant for the new patient, or that different cues were likely to be important, thus making use of a general instruction that relied on Phase I learning. Alternatively, in Experiment 2 all participants were told that the new patient was allergic to a list of four foods, the names of which were made explicit. Thus, for example, they were told that Miss Y is only allergic to fish, coffee, lemon, and cheese. The critical manipulation here is that two of those foods were predictive in the first phase, corresponding to Cues A and C, and the remaining two were nonpredictive cues from Phase I, corresponding to Cues X and Z. This means that in Experiment 2 there were four cues known to be causal, and four cues known to be noncauses. Of the known causes, two were predictive in Phase I (A and C), while two were nonpredictive (X and Z). Of the cues now known to be safe, two were predictive in Phase I (B and D), and two were nonpredictive in Phase I (W and Y). This orthogonal design therefore creates the condition in which an unambiguous instructional manipulation is present without removing the opportunity to observe an automatic influence of Phase I training, if indeed it is present.

If, as suggested by the findings of Experiment 1, controlled inferential processes are in operation, then a clear influence of instruction should be observed whereby causal ratings should reflect the instructions issued at the start of Phase II. Thus, by providing the exact identity of the allergens, participants should be able to accurately identify causal and noncausal cues on test. Again, it was anticipated that the memory ratings would show distinct results from those of the causal ratings. Given the explicit nature of the manipulation, recall should be better, overall, for cues instructed as causal. However, if an automatic attentional bias favoring predictive cues is also present, then a difference should be observed between instructed cues according to whether they were predictive (A and C) or nonpredictive (X and Z) in the first phase. Given the advantage conferred by predictive utility, this predicts better recall of cue–outcome relationships for A and C relative to X and Z. Similarly, a bias would be expected in favor of previously predictive cues for cues known to be noncausal. Thus, one might expect to observe better recall for B and D relative to W and Y. This result, despite the introduction of knowledge equating the causality of predictive and nonpredictive cues, would provide good evidence for an automatic residual bias favoring previously predictive information.

Table 2
Design of Experiment 2

Phase I	Phase II	Test
AW–O1	**A**Y–O3	A
AX–O1	B**Z**–O4	B
BW–O2	**C**W–O5	C
BX–O2	D**X**–O6	D
CY–O2		W
CZ–O2		X
DY–O1		Y
DZ–O1		Z

Note. Letters indicate individual cues. Bold underlined letters indicate cues instructed as informative for Phase II. O1–O6 refer to the six outcomes.

Method

Participants and apparatus. Twenty-six University of Sydney first year students (20 female, mean age = 19) participated in Experiment 2 for partial course credit. Apparatus remained as per Experiment 1.

Stimuli. Stimuli remained identical to those employed in Experiment 1with the exception of the introduction of two additional allergic reactions in order to account for additional outcomes included within the design. These were coughing and sweating.

Procedure. Phase I training and instructions remained the same as in Experiment 1. Following Phase I, participants were told that they were now observing the allergies of a new patient. Further, they were issued with an instruction explicitly stating which foods were allergens for the new patient. That is, they were shown the names of four foods, corresponding to Cues A, C, X, and Z and were informed that they would need to learn which of these corresponded to the various reactions that the patient was experiencing. Participants completed two blocks, each block consisting of two of the four trial types shown in Table 2. This change in procedure from Experiment 1, which employed longer Phase II training, was motivated by an attempt to avoid ceiling performance in memory, given the small number of cues and the explicit nature of the instructional manipulation. On each trial, participants were now required to predict which of four allergic reactions would occur. By employing four outcomes we ensured that previously predictive known causes A and C were paired with different outcomes, as were previously nonpredictive known causes, Z and X.

For each trial of the memory test, the four outcomes were displayed on screen as distinct alternatives beneath an individual cue. Participants were required to indicate which of these the cue had been paired with. Once that judgment had been completed, a rating scale appeared asking them to rate how confident they were in their response. The left anchor was labeled *not at all confident*, and the right anchor labeled *very confident*. This was followed by the appearance of the causal rating scale, which was identical to that used in Experiment 1.

Finally, the manipulation check required participants to report the instructed allergens of the second patient. Participants were excluded if they failed to correctly report all four allergens. Five participants were excluded on the basis of this check, as well as two participants who failed to reach the Phase I learning criterion, leaving 19 participants in the analysis. Remaining details of the method were as per Experiment 1.

Results

Phase I. In line with Experiment 1, prediction accuracy was averaged across the eight component discriminations and again increased steadily across training, as shown by Figure 4 A

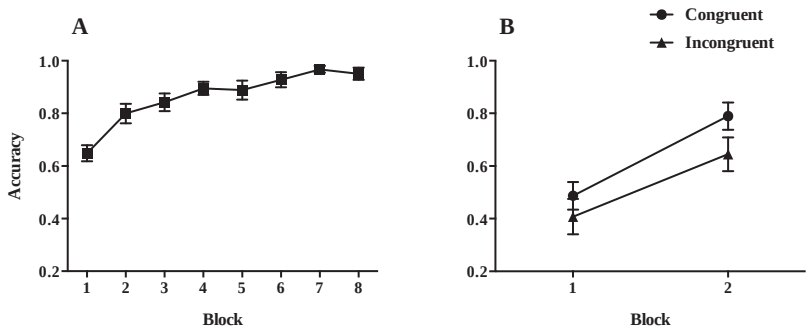


Figure 4. Panel A shows mean accuracy for Phase I training for Experiment 2. Panel B shows accuracy for Phase II for compounds with an instructed component that was previously predictive (congruent), and an instructed component that was previously nonpredictive (incongruent). Each block represents two presentations of the compound trials shown in Table 2. Error bars indicate standard error of the mean.

repeated-measures ANOVA showed a significant main effect of Block on accuracy, $F(7, 126) = 26.35$, $p < .001$, $\eta_p^2 = .59$.

Phase II. Accuracy for Phase II was averaged across compounds according to whether it contained an instructed component that was previously predictive (i.e., congruent: AY/CW) or an instructed component that was previously nonpredictive (i.e., incongruent: BZ/DX). This is shown in Figure 4. A repeated-measures ANOVA with Block (1–2) and Compound (congruent vs. noncongruent) showed a significant effect of Block on accuracy, $F(1, 18) = 39.29$, $p < .001$, $\eta_p^2 = .69$, and a main effect of Compound, $F(1, 18) = 7.25$, $p = .015$, $\eta_p^2 = .29$, such that accuracy was significantly higher for congruent compounds. The Block × Compound interaction failed to reach significance, $F < 1$.

Test.

Memory ratings. Accuracy for cue-outcome pairings was averaged according to whether cues were predictive or nonpredictive in the first phase. This is shown for cues instructed as causal and those known to be noncausal in Figure 5. One advantage of having four outcomes at test is that differences in accuracy are potentially easier to detect, allowing use of outcome accuracy as a measure (without conflating accuracy and confidence, as is the case with most learning scores in this literature). These were analyzed by means of a Predictiveness (predictive vs. nonpredictive) × Instruction (causal vs. noncausal) repeated-measures ANOVA. There was a significant main effect of Instruction, $F(1, 18) = 18.28$, $p < .001$, $\eta_p^2 = .5$. Although there was no main effect of Predictiveness, $F < 1$, Predictiveness and Instruction showed a significant interaction $F(1, 18) = 10.6$, $p = .004$, $\eta_p^2 = .37$, such that accuracy was higher for predictive than for nonpredictive cues instructed as causal, $F(1, 18) = 5.7$, $p = .028$, $\eta_p^2 = .24$, while for known non causes accuracy for nonpredictive cues was higher relative to predictive cues, $F(1, 18) = 6.4$, $p = .02$, $\eta_p^2 = .26$, as revealed by simple effects analysis.[2]

Causal ratings. Again, causal ratings were averaged according to Phase I predictiveness and instruction, as shown in Figure 5. A repeated-measures ANOVA with Predictiveness (predictive vs. nonpredictive) and Instruction (causal vs. noncausal) as factors showed a significant effect of Instruction, $F(1, 18) = 22.03$, $p < .001$, $\eta_p^2 = .55$. Cues instructed as causal were given higher causal ratings at test. There was no main effect of Predictiveness, and no Instruction × Predictiveness interaction, Fs < 1.

Discussion

The causal inference and cued-recall tests in Experiment 2 clearly diverge. Causal ratings closely followed the instructions issued at the start of Phase II: Regardless of the predictive status of cues in the first phase, cues instructed as allergens were attributed as causal at test, while those known to be safe were given low causal ratings. Recall of cue-outcome pairings, however, showed differences according to predictiveness within each of the instructional conditions. Consistent with predictions, the learned predictiveness effect was still evident among cues known to be allergenic. That is, more was learned about the previously predictive cues compared with previously nonpredictive cues, despite the explicit knowledge that both sets of cues were allergens.

Surprisingly, the opposite pattern emerged for cues that were known noncauses, whereby recall was better for cues that were nonpredictive in Phase I. This reversal raises the question of how Phase I training influences further learning under conditions of instructional manipulation. Although the result for known causes suggests a role for automatic bias, this does not appear to combine with inferential reasoning in an additive manner to direct learning.

One limitation for providing a coherent account of the observed reversal is an asymmetry in outcome equivalence for predictive and nonpredictive components across the two phases. For example, given that predictive Cues A and C were paired with distinct outcomes during Phase I, O1 and O2, respectively, one way in which learning about the cue-outcome relationships may have been facilitated for these cues in Phase II is if participants equate O1 with O3, and O2 with O5, that is, the outcomes with which A and C are subsequently paired. A more detailed analysis of the acquisition results will be considered in the General Discussion, but it is worth noting that acquisition was indeed better for compounds in which outcome equivalence might facilitate learning (i.e., AY and CW). However, given that these are also the compounds in which the instructed component was previously predictive, there is no way to distinguish the effects of outcome equiv-

[2] For consistency with the other experiments, it is worth noting that when the same analysis was conducted on a measure that combined recall accuracy with confidence scores, it showed a similar pattern of results, that is, a robust interaction between Phase I predictiveness and instruction.

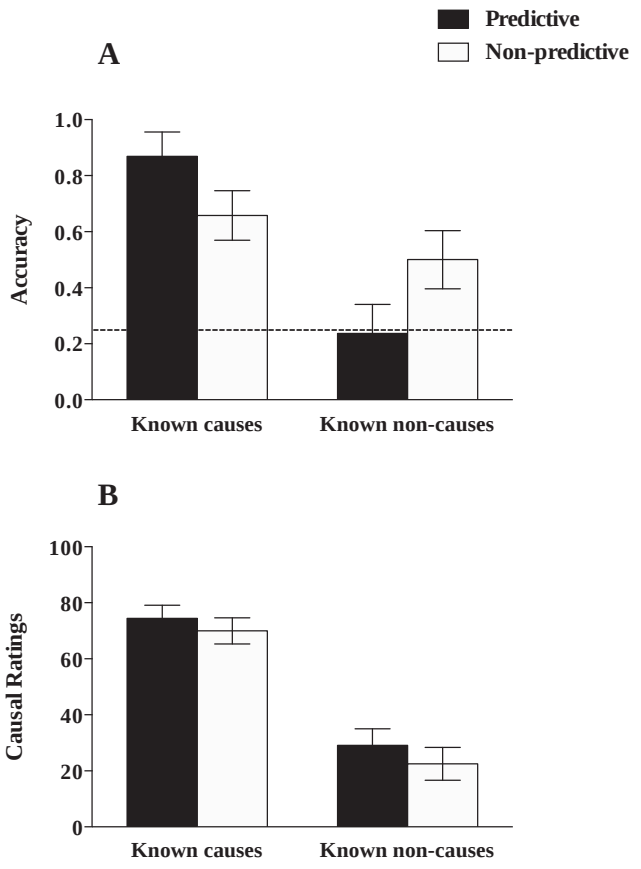


Figure 5. Accuracy for cue-outcome pairings (A) and causal ratings (B) averaged according to predictiveness, for known causes and known non-causes for Experiment 2. Error bars represent standard error of the mean difference between predictive and nonpredictive cues. The dotted line in Panel A represents chance.

alence from Phase I predictiveness on learning. If the results observed in Experiment 2 can be replicated under conditions in which no outcome equivalence is present for predictive cues across the two phases of learning, this would provide clearer evidence for an interaction between automatic and inferential processes in learned predictiveness under conditions of instructional manipulation.

Experiment 3

In Experiment 3 an extended design was employed, closely following that of Experiment 2. This is shown in Table 3. As per previous experiments, the initial phase of training followed the learned predictiveness procedure, although it employed additional cues that served to double the number of Phase I trial types. Again, at the start of Phase II, participants were informed which cues would be allergenic for the new patient. This instruction was issued at the category level whereby it was made explicit that the new patient was allergic to (and *only* allergic to) a specific category of foods. Therefore, all food cues in this experiment came from one of two categories. Half of the cues, those corresponding to A–D as well as W–Z, belonged to one category that was instructed as allergenic. The remaining cues, those corresponding to E–H and S–V, belonged to the "safe" category, thus each compound consisted of one component from either category. Given that half of the cues in each category were predictive in Phase I, and the remaining half nonpredictive, this manipulation creates the same conditions as those in Experiment 2. If the results observed in Experiment 2 are robust, then recall for previously predictive cues should be better than nonpredictive cues in the instructed category. For the category known to be safe, the reverse is expected whereby recall should be better for nonpredictive compared with predictive cues. This measure should no longer be sensitive to differences in outcome equivalence across the two phases of learning, which are negated in the extended design. For example, Cues A and D are predictive of O1 during the first phase. These same cues are equally likely to predict O3 and O4, respectively in Phase II. Thus, any strategy employed by participants to equate outcomes across training phases would yield incorrect predictions on half of the trials. Causal ratings are again expected to reflect Phase II instructions, showing high causal attribution to the category of foods instructed as allergenic, and low ratings for foods in the category known to be safe.

Method

Participants and apparatus. Twenty-seven first year students enrolled at the University of Sydney participated in the experiment (17 female, mean age = 20). Apparatus was the same as previous experiments.

Stimuli. Food stimuli for this experiment were drawn from the two categories of "fruit and vegetables," and "animal products such as meat, poultry, and dairy." The eight fruit and vegetable category items consisted of: lemon, apple, avocado, peas, banana, mushroom, strawberries, and broccoli. Items from the meat and animal products category included: fish, milk, eggs, steak, cheese, bacon, chicken, and yogurt. The four outcomes were as per Experiment 1, as well as remaining details of the presentation parameters.

Table 3
Design of Experiment 3

Phase I	Phase II	Test
AS–O1	**A**U–O3	A
AT–O1	**B**V–O4	B
BS–O2	**C**S–O3	C
BT–O2	**D**T–O4	E
CU–O2	E**Y**–O3	F
CV–O2	F**Z**–O4	G
DU–O1	G**W**–O3	H
DV–O1	H**X**–O4	S
EW–O1		T
EX–O1		U
FW–O2		V
FX–O2		W
GY–O2		X
GZ–O2		Y
HY–O1		Z
HZ–O1		

Note. Letters refer to individual food cues. Bold underlined letters indicate cues instructed as causal for Phase II. O1–O4 refer to the four outcomes in this experiment.

Procedure. The procedure for the initial phase of training was identical to that of the previous two experiments, though with an additional number of trial types in each block given the extended design. Following this, participants were again informed that they would be observing a new patient. They were told that this patient was only allergic to a specific category of food, corresponding to either fruit and vegetables, or animal products such as meat, poultry, and dairy. The eight compounds in this phase appeared twice per block, with participants completing two blocks each. The allocation of categories as allergens was counterbalanced across participants.

Aside from the additional cues presented at test, corresponding to each of the 16 cues included in the design, the structure of the test phase was identical to that of Experiment 1. This was followed by a manipulation check asking participants to report the category of food that the second patient was allergic to. There were no exclusions on the basis of this check, as all participants correctly reported the allergenic category. Three participants were excluded on the basis of the Phase I learning criterion, leaving 24 participants in the final analysis.

Results

Phase I. Initial acquisition proceeded as expected, with a repeated-measures ANOVA showing a significant effect of Block, $F(7, 161) = 50.57$, $p < .01$, $\eta_p^2 = .69$.

Phase II. Figure 6 shows accuracy for Phase I and Phase II, averaged across compounds according to whether they contained an instructed component that was previously predictive (congruent) or previously nonpredictive (incongruent). A repeated-measures ANOVA with Block (1–2) and Compound (congruent vs. incongruent) showed a significant effect of Block, $F(1, 23) = 29.3$, $p < .001$, $\eta_p^2 = .56$. Acquisition for congruent compounds was slightly better than incongruent compounds, although this difference did not reach a conventional level of significance, $F(1, 23) = 3.42$, $p = .07$, $\eta_p^2 = .13$, and there was no significant Block × Compound interaction, $F < 1$.

Test.

Memory test. Memory scores for each cue were calculated as per Experiment 1 and averaged according to whether they were predictive (A–H) or nonpredictive (S–Z) in the first phase. These are shown for cues instructed as causal, and those known to be noncausal in Figure 7. Scores were analyzed by way of a repeated-measures ANOVA with Predictiveness (predictive vs. nonpredictive) and Instruction (causal vs. noncausal) as factors. This revealed a main effect of Predictiveness, $F(1, 23) = 4.81$, $p = .039$, $\eta_p^2 = .17$, a main effect of Instruction, $F(1, 23) = 32.62$, $p < .001$, $\eta_p^2 = .59$, as well as a significant Predictiveness × Instruction interaction, $F(1, 23) = 9.16$, $p = .006$, $\eta_p^2 = .29$. A simple effects analysis investigating the interaction revealed that for instructed causes, memory scores were significantly higher for previously predictive compared to previously nonpredictive cues, $F(1, 23) = 8.36$, $p = .008$, $\eta_p^2 = .27$. For known noncauses, the benefit for nonpredictive cues was marginally higher compared with predictive cues, showing a similar, although less robust, trend to that observed in Experiment 2, $F(1, 23) = 3.55$, $p = .07$, $\eta_p^2 = .13$.

Causal ratings. Causal ratings, shown in Figure 7, were similarly analyzed by way of repeated-measures ANOVA, showing a significant main effect of Predictiveness, $F(1, 23) = 5.27$, $p = .03$, $\eta_p^2 = .19$, as well as a main effect of Instruction, $F(1, 23) = 64.52$, $p < .001$, $\eta_p^2 = .74$. The Predictiveness × Instruction interaction did not reach significance, $F(1, 23) = 1.11$, $p = .30$, $\eta_p^2 = .05$.

Discussion

Overall, the results of Experiment 3 provide a replication of those seen in Experiment 2, under conditions in which facilitation through outcome equivalence is impossible. For cues instructed as causal, better retention of the specific cue-outcome relationships was observed for predictive compared to nonpredictive cues. The difference between predictive and nonpredictive cues in the safe category did not reach conventional levels of statistical significance by a two-tailed test ($p = .07$). Nevertheless, the ordinal pattern of means was the same as Experiment 2 and, given the robust interaction between predictiveness and instruction, it is very clear that prior predictiveness does not have the same effect on learning about known noncauses as it does on learning about known causes. Again, causal ratings showed little evidence of an interaction between Phase I training and instruction, closely reflecting the instructional manipulation. Further, acquisition fol-

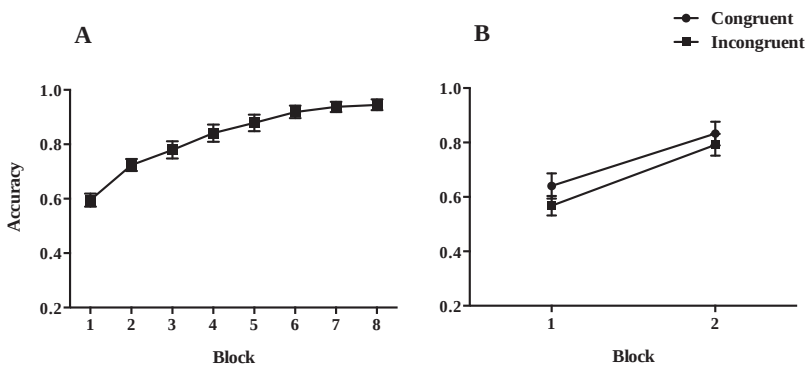


Figure 6. Panel A shows mean accuracy for Phase I training for Experiment 3. Panel B shows accuracy for Phase II for compounds with an instructed component that was previously predictive (congruent), and an instructed component that was previously nonpredictive (incongruent). Each block represents two presentations of the compound trials shown in Table 3. Error bars represent *SEM*.

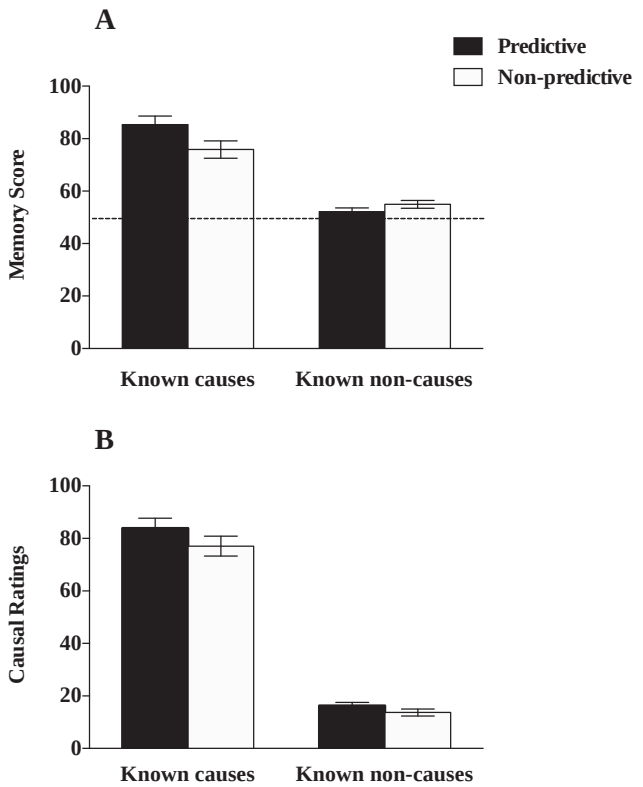


Figure 7. Memory scores (A) and causal ratings (B) averaged according to predictiveness, for known causes and known noncauses for Experiment 3. Error bars represent standard error of the mean difference between predictive and nonpredictive cues. The dotted line in Panel A represents chance.

lowed a similar, although less statistically robust ($p = .07$), pattern to that observed in Experiment 2, such that learning was better for compounds that contained an instructed component that was previously predictive.

This pattern of results suggests a reliable persistence of the influence of Phase I training in learned predictiveness over and above the influence of instructional manipulations. Although making use of a categorical instruction appears to have improved memorability of the instruction following Phase II, this may have allowed for differences in generalization between cues across Phase I and Phase II. As cues from the same category are likely to be more similar to one another compared with cues from the alternative category, the associability of an item from Phase I might generalize more readily to items of the same category in Phase II. However, this is highly unlikely to have a confounding influence on the results. Across both stages of learning, each compound consisted of one component from either category. However, each category consisted of cues that were predictive in Phase I, and cues that were nonpredictive in Phase I. Take for example the instructed components A and Y, from compounds AU and EY respectively. Both A and Y are taken from the same category, yet one was predictive in Phase I, that is A, and one was nonpredictive in Phase I, that is Y. If there was greater generalization between members of the same category, this would affect both A and Y. This means that any potential generalization of associability within a category is equated for predictive and nonpredictive components.

General Discussion

Three experiments investigated the effect of instructional manipulations on associative memory and causal reasoning in learned predictiveness. In all experiments, the instructional manipulation clearly influenced recall of cue-outcome associations. Overall, this was also the case for causal ratings. Although these closely followed the pattern of results observed for cue-outcome recall in Experiment 1, causal judgments in Experiments 2 and 3 diverged from our measure of associative memory. In line with the findings of Mitchell et al. (2012), Experiment 1 confirmed that reversal instructions influence the learned predictiveness bias, and provided an important extension by demonstrating that this influence is not isolated to causal reasoning, but also includes associative memory. In the continuity condition, both recall and causal attribution were higher for predictive cues compared with nonpredictive cues. However, the reversal of the learned predictiveness effect in the change condition was incomplete as no difference between predictive and nonpredictive cues was observed across both measures. It is unclear why we failed to find instructed reversal of learned predictiveness yet Mitchell et al. (2012) found strong evidence for a reversal. One difference between the two protocols is the conceptual scenario employed. In the study of Mitchell et al. (2012), fictitious seeds, acting as the predictive and nonpredictive cues are cross-pollinated in order to grow various shapes of tree, the outcomes. This scenario potentially favors a more categorical inferential process. Given the nature of the causal relationship between seeds and trees, that is, that only one seed will grow a specific tree, this aspect of the design raises the possibility that a complete reversal was facilitated based on conceptual aspects of the scenario in addition to the manipulation of interest. Specifically, if each outcome is most likely attributable to only one of the compound components during Phase II and not the other, this means that at test, ratings are more likely to reflect a mutually exclusive causal structure across those same components. However, whether this conceptual component would be sufficient to strengthen the magnitude of the reversal remains an empirical question.

Subsequent experiments provided clear evidence of an automatic bias in Phase II recall according to Phase I training. Making use of an orthogonal manipulation between Phase I predictiveness and instructions issued at the item (Experiment 2), and categorical (Experiment 3) level, a robust benefit in recall for previously predictive compared to previously nonpredictive cues was observed for cues known to be allergenic. Surprisingly, this reversed for known noncauses as recall was higher for nonpredictive cues. This reversal raises the question of how Phase I training influences further learning under conditions of instructional manipulation. Although the result for known causes suggests a role for automatic bias, this does not appear to combine with inferential reasoning in an additive manner to direct learning.

One explanation for this result is that instead of directing attention in a rigid fashion, predictive utility enhances cognitive control. That is, previously predictive cues are both easier to attend and ignore, depending on task requirements. According to this explanation, learning is facilitated for predictive known causes and suppressed for predictive cues known to be noncausal.

An alternative explanation can be made on the assumption that differences in congruence between instruction and predictive history are crucial for the rate of learning about the Phase II compounds. In both Experiments 2 and 3, some Phase II compounds can be seen to possess congruence between instruction and Phase I predictive utility. For example, in Experiment 2, for compounds AY an CW, predictive cues in Phase I are instructed as causal for Phase II and nonpredictive cues in Phase I are instructed as noncauses in Phase II and thus remain irrelevant. In contrast, a switch in the utility of the cues occurs for the Phase II compounds BZ and DX in which cues that are nonpredictive in Phase I are instructed as causal in Phase II, and cues that were predictive in Phase I are known to subsequently be noncausal. The incongruence between phases of the utility of the cues in the latter set of compounds may produce a cost to learning about all stimuli present on those trials. Likewise, according to this explanation, when the predictiveness of the cues from Phase I and the instructions about causation in Phase II are congruent, learning about all presented cues is achieved relatively easily. This raises the possibility that less is learned in general about Phase II compounds in which participants have to attend to the previously nonpredictive cue and ignore the previously predictive cue.

The rate of learning in Phase II for congruent trials types appeared to be somewhat higher than for the incongruent trials in both Experiments 2 and 3. Prediction accuracy was higher for congruent trials in Experiment 2 and trended in the same direction in Experiment 3. This result is consistent with there being interference to the general rate of learning on incongruent trials. However, it is worth noting that this result is also amenable to an explanation based on a change in cognitive control. If previously predictive cues afford greater cognitive control then each would serve as a highly attended and therefore associable cue on congruent trials or an efficiently ignored cue on incongruent trials. If cognitive control of the previously nonpredictive cues was poorer, and thus did not compensate fully for the attention or suppression of the predictive cue, then net performance would remain higher on congruent trials than on incongruent trials.

Finally, a third alternative, based on the formation of within-compound associations during Phase II, might explain the observed interaction between Phase I training and instruction. During Phase II, there were compounds in which the instructed component was predictive in Phase I, such as AY, and compounds in which the instructed component was nonpredictive in Phase I, such as BZ. If Phase I predictiveness increases the strength of the association for instructed components, then the predictive, instructed Cue A would be expected to form a strong association with O3 on trials in which AY is present. Alternatively, the nonpredictive instructed component Z would be expected to form a weaker association with O4 on trials in which BZ is present. The formation of within-compound associations between A and Y therefore predicts that learning about Y is benefitted indirectly by virtue of a Y-A-O3 chain. Alternatively, the association between B and Z predicts that learning about B receives less benefit by virtue of the B-Z-O4 chain. This leads to the prediction that learning for A should be greater than learning for Z, but also that learning for B should be less than learning for Y, which is the pattern of results observed. Indeed, evidence consistent with the formation of within-compound associations during causal learning has previously been reported (Dickinson & Burke, 1996). Although retrieval-based associative models can largely explain these effects without appealing to the presence of within-compound associations (e.g., Le Pelley & McLaren, 2001), whether these are necessary in explaining the present data will depend on further defining the nature of the interaction.

Regardless of why this interaction between predictive history and instruction occurs, an interesting feature of the results is that the outcome recall measures clearly produced a different set of results to the causal ratings across both experiments. The causal ratings closely followed the instructions issued at the start of Phase II. Theorists who claim a link between associative learning and causal inference (e.g., McLaren, Green, & Mackintosh, 1994) usually assume that the strength with which a cue retrieves an associated outcome serves as one potential source of evidence on which a causal judgment might be made. This does not preclude the possibility that causal judgments might be made on the basis of alternative sources of evidence, such as the product of explicit inferences, if the individual is given sufficient motivation to do so. Direct instructions do exactly this. Importantly, our results demonstrate that despite the instructions having a profound effect on causal ratings, and a strong effect on memory-based judgments, there was still clear evidence that Phase I learning impacted on associative memory in a way that is not explained by the instruction. It should be noted that there was also some evidence for an influence of predictiveness on causal attribution in Experiment 3. This suggests that causal judgments, although reflecting explicit inference, may nonetheless be influenced by the strength of associations in some situations.

One important point to consider here, however, relates to the predictions generated by an inferential account of associative learning. In their investigation of memory and causal reasoning in the context of blocking, Mitchell et al. (2006) have argued that an inferential account predicts a dissociation between measures of associative memory and causal reasoning. In the case of blocking, a novel (blocked) cue appears alongside a cue previously established as causal. An inferential account makes the assumption that the relationship between the blocked cue and the outcome with which it is paired is indeed encoded. The memory of that relationship then serves as a means to judge that same cue as noncausal. Thus, memory ratings for blocked cues should be comparable with cues previously established as causal. Alternatively, causal ratings for blocked cues would be low, while causal ratings for previously causal cues should be high. If this line of reasoning is extended to a more traditional demonstration of learned predictiveness, a strong inferential position might predict no evidence of learned predictiveness in associative memory, rather observing the bias in causal attribution. That is, one might expect to observe no difference in associative memory between predictive and nonpredictive cues, while causal attribution would be high for predictive cues and low for nonpredictive cues.

However, this line of reasoning becomes less clear when an instruction is introduced into the procedure. Given that the instructions are issued at the start of Phase II, participants have access to the explicit inference regarding causality while they are engaged in learning, and therefore have little reason to encode information about cues known to be noncausal. For example, if participants know that A is causal and that Y is noncausal, when encountering these cues during training there is little motivation to remember the relationship between Y and the outcome with which it is paired.

Thus, whether an inferential account would predict the same dissociation when attention is purposefully biased via explicit instruction is unclear. This raises the question of whether an inferential account of associative learning is specified in enough detail in order to generate clear predictions in situations where automatic biases may be interacting with higher order inferences.

Certainly, it remains unclear as to whether these are best thought of as distinct ways in which information is processed, or whether automatic associations form the underlying structure from which inferences emerge. In either case, once an inference is present it is unlikely to exhibit the same properties as a simple association. The fact that the effects of instruction and predictive history are dissociable but seemingly interactive seems to support this. Given that measures of causal learning were seemingly more sensitive to instruction than measures of associative memory, this raises the question of whether different kinds of learning are equivalent in their sensitivity to inferential processes. It may be that learning about causal value is under greater control of inferential processing. Whether instructions would manipulate predictive learning for example, independently of causal value, remains to be seen.

Overall, the present findings are consistent with an increasing collection of results from related paradigms demonstrating that predictiveness influences stimulus selection in tasks that are thought to reflect nonstrategic processes. For example, it has been shown that previously predictive cues facilitate learning of subsequent spatial motor response sequences (Beesley & Le Pelley, 2010), capture spatial attention (Le Pelley, Vadillo, & Luque, 2013), and attenuate the attentional blink (Livesey, Harris, & Harris, 2009). Although this provides evidence for an automatic attentional bias, it should be noted that our observed reversal in recall does not conform to the proposed relationship between associability and predictive history. Accordingly, there are a number of studies suggesting that the learned predictiveness effect does not operate via the competitive associative algorithms of attentional change described by Mackintosh (1975; Le Pelley, 2004; Pearce & Mackintosh, 2010). For instance, Le Pelley et al. (2010b) found that competition between cues in compound was not necessary for learned predictiveness to occur, and Livesey et al. (2011) found no evidence that direct (i.e., within-trial) comparison between predictive and nonpredictive cues affected the magnitude of learned predictiveness. The current study demonstrates another way in which the automatic allocation of attention appears to behave differently from model predictions. Although there appears to be a relatively automatic influence of the predictive history of the cues, that influence only matches the predictions of associative learning theories for cues that are deliberately attended.

So far, our interpretation has assumed that associability influences the strength of learning, such that cues with high α following Phase I have a direct benefit in the formation of subsequent associations. However, Mackintosh (1975) did raise the possibility that associability may influence performance as well as learning. Indeed, some models describe cue competition effects largely in terms of performance effects on test rather than acquisition deficits (Stout & Miller, 2007). Accordingly, Le Pelley, Suret, and Beesley (2009) have noted that a performance-based explanation of learned predictiveness would allow for equivalent gains in associative strength for predictive and nonpredictive cues across Phase II. The observed bias is therefore a function of increased responding to cues with high α at test, that is, those that were predictive in Phase I. Indeed, there is some evidence to suggest that predictiveness impacts performance as well as learning (Le Pelley et al., 2009). Given that we found evidence of differences in acquisition in Experiments 2 and 3, it seems unlikely that our results would be amenable to explanations of learned predictiveness which rely exclusively on performance at test to explain differences in learning scores. However, it is possible that the effects reported here reflect a combination of changes in performance as well as learning.

In conclusion, it appears that automatic biases in learning persist under conditions of explicit instructional manipulation of the learned predictiveness effect. Further, these results suggest a distinct effect of associative memory and causal reasoning in producing the bias. However, these processes do not appear to influence learning in an additive manner. Rather, it seems that reasoning and associative memory combine in ways that are not directly anticipated on the basis of attentional models of associative learning (e.g., Mackintosh, 1975). Clearly the relationship between these processes is a complicated one, and the mechanisms by which they interact remain to be established.

References

Beesley, T., & Le Pelley, M. E. (2010). The effect of predictive history on the learning of sub-sequence contingencies. *The Quarterly Journal of Experimental Psychology, 63,* 108–135. doi:10.1080/17470210902831767

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. doi:10.1163/156856897X00357

Dickinson, A., & Burke, J. (1996). Within compound associations mediate the retrospective revaluation of causality judgements. *The Quarterly Journal of Experimental Psychology, 49,* 60–80. doi:10.1080/713932614

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology, 45,* 812–863. doi:10.1006/jmps.2000.1354

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology, 57,* 193–243. doi:10.1080/02724990344000141

Le Pelley, M. E. (2010). Attention and human associative learning. In C. J. Mitchell & M. E. Le Pelley (Eds.), *Attention and associative learning: From brain to behaviour* (pp. 187–215). Oxford, UK: Oxford University Press.

Le Pelley, M. E., & McLaren, I. P. L. (2001). Retrospective revaluation in humans: Learning or memory? *The Quarterly Journal of Experimental Psychology, 54,* 311–352. doi:10.1080/02724990143000072

Le Pelley, M. E., & McLaren, I. P. L. (2003). Learned associability and associative change in human causal learning. *The Quarterly Journal of Experimental Psychology, 56,* 68–79. doi:10.1080/02724990244000179

Le Pelley, M. E., Mitchell, C. J., & Johnson, A. M. (2013). Outcome value influences attentional biases in human associative learning: Dissociable effects of training and instruction. *Journal of Experimental Psychology: Animal Behavior Processes, 39,* 39–55. doi:10.1037/a0031230

Le Pelley, M. E., Reimers, S. J., Calvini, G., Spears, R., Beesley, T., & Murphy, R. A. (2010a). Stereotype formation: Biased by association. *Journal of Experimental Psychology, 139,* 138–161. doi:10.1037/a0018210

Le Pelley, M. E., Suret, M. B., & Beesley, T. (2009). Learned predictiveness effects in humans: A function of learning, performance, or both? *Journal of Experimental Psychology: Animal Behavior Processes, 35,* 312–327. doi:10.1037/a0014315

Le Pelley, M. E., Turnbull, M. N., Reimers, S. J., & Knipe, R. L. (2010b). Learned predictiveness effects following single-cue training in humans. *Learning & Behavior, 38,* 126–144. doi:10.3758/LB.38.2.126

Le Pelley, M. E., Vadillo, M., & Luque, D. (2013). Learned predictiveness influences rapid attentional capture: Evidence from the dot probe task. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 39,* 1888–1900. doi:10.1037/a0033700

Livesey, E. J., Harris, I. M., & Harris, J. A. (2009). Attentional changes during implicit learning: Signal validity protects a target stimulus from the attentional blink. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 408–422. doi:10.1037/a0014525

Livesey, E. J., Thorwart, A., De Fina, N. L., & Harris, J. A. (2011). Comparing learned predictiveness effects within and across compound discriminations. *Journal of Experimental Psychology: Animal Behavior Processes, 37,* 446–465. doi:10.1037/a0023391

Lochmann, T., & Wills, A. J. (2003). Predictive history in an allergy prediction task. In *Proceedings of EuroCogSci 03,* The European Conference of the Cognitive Science Society (pp. 217–222). Mahwah, NJ: Erlbaum.

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

McLaren, I. P. L., Green, R. E. A., & Mackintosh, N. J. (1994). Animal learning and the implicit/explicit distinction. In N. C. Ellis (Eds.), *Implicit and explicit learning of languages* (pp. 313–332). London, UK: Academic.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32,* 183–198. doi:10.1017/S0140525X09000855

Mitchell, C. J., Griffiths, O., Seetoo, J., & Lovibond, P. F. (2012). Attentional mechanisms in learned predictiveness. *Journal of Experimental Psychology: Animal Behavior Processes, 38,* 191–202. doi:10.1037/a0027385

Mitchell, C. J., Lovibond, P. F., & Gan, C. Y. (2005). A disassociation between casual judgment and outcome recall. *Psychonomic Bulletin and Review, 12,* 1–5. doi:10.3758/BF03196791

Mitchell, C. J., Lovibond, P. F., Minard, E., & Lavis, Y. (2006). Forward blocking in human learning sometimes reflects the failure to encode a cue–outcome relationship. *The Quarterly Journal of Experimental Psychology, 59,* 830–844. doi:10.1080/17470210500242847

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian conditioning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532–552. doi:10.1037/0033-295X.87.6.532

Pearce, J. M., & Mackintosh, N. J. (2010). Two theories of attention: A review and a possible integration. In C. J. Mitchell & M. E. Le Pelley (Eds.), *Attention and associative learning: From brain to behaviour* (pp. 11–39). Oxford, UK: Oxford University Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. doi:10.1163/156856897X00366

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review, 16,* 225–237. doi:10.3758/PBR.16.2.225

Shone, L. T., & Livesey, E. J. (2013). Automatic and instructed attention in learned predictiveness. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1306–1311). Austin, TX: Cognitive Science Society.

Stout, S. C., & Miller, R. R. (2007). Sometimes-competing retrieval (SOCR): A formalization of the comparator hypothesis. *Psychological Review, 114,* 759–783. doi:10.1037/0033-295X.114.3.759

Vadillo, M. A., & Matute, H. (2007). Predictions and causal estimations are not supported by the same associative structure. *The Quarterly Journal of Experimental Psychology, 60,* 433–447. doi:10.1080/17470210601002520

Vadillo, M. A., Miller, R. R., & Matute, H. (2005). Causal and predictive-value judgments, but not predictions, are based on cue-outcome contingency. *Learning & Behavior, 33,* 172–183. doi:10.3758/BF03196061

Wills, A. J., Lavric, A., Hemmings, Y., & Surrey, E. (2014). Attention, predictive learning, and the inverse base-rate effect: Evidence from event-related potentials. *NeuroImage, 87,* 61–71. doi:10.1016/j.neuroimage.2013.10.060

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5,* 495–501. doi:10.1038/nrn1411

Wright, R. D., & Ward, L. M. (2008). *Orienting of attention*. New York, NY: Oxford University Press.

Received April 10, 2014
Revision received August 20, 2014
Accepted August 25, 2014 ■