Blocking and Backward Blocking Involve Learned Inattention

John K. Kruschke and Nathaniel J. Blair
Indiana University

In Press, *Psychonomic Bulletin & Review*. Version of October 26, 1999.

Four experiments examine blocking of associative learning by human participants in a disease diagnosis procedure. The results indicate that after a cue is blocked, subsequent learning about the cue is attenuated. This attenuated learning after blocking is obtained for both standard blocking and for backward blocking. Attenuated learning after blocking cannot be accounted for by theories, such as the Rescorla-Wagner model, that rely on lack of learning about a redundant cue, nor by extensions of the Rescorla-Wagner model designed to address backward blocking, that encode absent cues with negative values. The results are predicted by the hypothesis that people learn not to attend to the blocked cue.

When a compound of two cues, A and B, is paired with an outcome, both cues will typically accrue moderate associative strength with the outcome. If, however, the pairing of the compound with the outcome is preceded by an earlier phase of training in which cue A by itself is paired with the outcome, then cue B does not accrue much associative strength with the outcome. Apparently the previous learning about cue A has *blocked*, i.e., prevented, learning about cue B (Kamin, 1969).

There are two traditional theories of blocking that have vied for thirty years. One theory, proposed by Kamin (1969) and formalized in the classic Rescorla-Wagner (RW, 1972) model, argues that blocking is caused by *lack of learning*. According to this approach, in the second phase of training, there is lack of surprise about the outcome because cue A already predicts the outcome. To the extent that there is no predictive error, there is no change in associative weight from cue B to the outcome. The blocked cue remains essentially unaffected. A primary motivation of the RW model was accounting for blocking (cf. Miller, Barnet, & Grahame, 1995; Siegel & Allan, 1996), and the model continues to be cited as the standard explanation of blocking (e.g., Domjan, 1998, pp. 107–110).

A second theory of blocking, suggested by Sutherland & Mackintosh (1971) and extended by Mackintosh (1975), argues that participants do learn something about the redundant cue B. Specifically, participants learn to suppress attention to it because it predicts no change in reinforcement. Thus, this theory asserts that blocking is not caused solely by lack of learning, but by *learned inattention* to the blocked cue.

Mackintosh & Turner (1971) presented evidence in favor of the learned-inattention theory. They examined learning *subsequent* to blocking, in order to determine whether the blocking of a cue attenuated subsequent learning about the cue. Suppose that after having a cue blocked, the learner must associate it with an outcome. If the lack-of-learning theory is correct, then the blocked cue begins this new phase in a relatively pristine state, because there was little learned about the cue. If the learned-inattention theory is correct, then the blocked cue begins this new phase at a disadvantage, because attention to it has been suppressed, and this suppression must be overcome by new learning. The results of Mackintosh and Turner's (1971) experiment were consistent with the theory of learned inattention.

Although blocking has been observed in human learning in many situations (e.g., Arcediano, Matute, & Miller, 1997; Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993; Chapman & Robbins, 1990; Dickinson, Shanks, & Evenden, 1984; Hinchy, Lovibond, & Ter-Horst, 1995; Shanks, 1985; Waldman & Holyoak, 1992; Williams, Sagness, & McPhee, 1994), attenuation of learning about a blocked cue has not been reported for humans, to our knowledge. Mackintosh and Turner's (1971) experiment used rats as subjects, with magnitude of shock as the outcome (i.e., the unconditioned stimulus), and with suppression of lever pressing (which dispensed food pellets) as the dependent measure of learning. Different training conditions were compared across different groups of subjects. The experiments reported below extended Mackintosh and Turner's (1971) results to human participants, in a multiple-outcome disease diagnosis paradigm, and in a within-subjects design. The experimental designs used here also improve upon that of Mackintosh and Turner (1971). Unlike the earlier design, the present experiments

This research was supported in part by NIMH FIRST Award 1-R29-MH51572.

For assistance administering the experiments, we thank James Cortright, Christy Doherty, Jacob Hall, Scott Harris, Erin DeMien, Angie McKillip, Heather Shelton and Elyse Weiss.

For helpful comments on earlier drafts of this article, we thank Michael Fragassi, Mark Johansen, Teresa Treat, John Wixted, Pete Wood, Michael Young, and two anonymous reviewers.

Correspondence can be addressed to John K. Kruschke, Department of Psychology, Indiana University, 1101 E. 10th St., Bloomington IN 47405–7007, or via electronic mail to kruschke@indiana.edu. The first author's world wide web page is at http://www.indiana.edu/~kruschke/

(a) include a control condition that shows robust blocking before subsequent attenuation of learning, (b) include a control condition that tests whether novelty alone can account for the attenuation of learning, and (c) test for attenuated learning by using a more sensitive design in which the learner has the option to learn about the blocked cue or another redundant cue.

Researchers have more recently discovered *backward blocking*, in which the two training phases of the blocking paradigm are reversed: In the first phase of learning, a compound of two cues, A and B, is paired with an outcome. Both cues acquire associative strength with the outcome. In the next phase of learning, cue A by itself is paired with the outcome. In subsequent testing, cue B has lost some of its associative strength with the outcome, despite the fact that it was not present in the second phase of training (e.g., Chapman, 1991; Dickinson & Burke, 1996; Larkin, Aitken, & Dickinson, 1998; Miller & Matute, 1996; Shanks, 1985; Van Hamme & Wasserman, 1994; Wasserman & Berglan, 1998; Wasserman, Kao, VanHamme, Katagiri, & Young, 1996; Williams et al., 1994).

Backward blocking cannot be accounted for by the RW model because the model assumes that absent cues are encoded with an activation of zero, and consequently they can have no influence on responding and no influence on learning. Recently, some theorists have proposed modifications of the RW model, in which absent cues are encoded with negative values (Tassoni, 1995; VanHamme & Wasserman, 1994; cf. Dickinson & Burke, 1996; Larkin, Aitken & Dickinson, 1998; Markman, 1989). When absent cues are encoded with negative values, the second phase of training in backward blocking has the absent cue B encoded negatively, and consequently its associative strength with the outcome is reduced.

Like the basic RW model, however, these extended versions predict no attenuation of learning about a previously blocked cue, whether the cue was blocked in a standard or backward paradigm. In this article we report that learning about a blocked cue is attenuated after backward blocking (Experiment 3), as well as after forward blocking (Experiment 1). Control experiments (2 and 4) shows that this attenuation of learning cannot be accounted for by novelty of cues. The attenuation is predicted, however, by the hypothesis that blocking, whether backward or forward, is caused at least in part by learned inattention to the blocked cue.

Experiment 1: Blocking attenuates subsequent learning

People learned to associate lists of symptoms with disease labels. On each trial, a symptom, or set of symptoms, was displayed, as indicated by letters A–I in Table 1, and the learner had to guess which of six diseases was the correct diagnosis, indicated by numerals in Table 1. There were three phases of learning. In the first phase, symptom A always resulted in disease 1, denoted A$\rightarrow$1. In the second phase of training, redundant symptom B was added to symptom A, always leading to the same disease as previously occurred with symptom A (i.e., AB$\rightarrow$1), so that learning about symptom B would, presumably, be blocked. The second phase also included HI$\rightarrow$6, which acted as a comparison for the blocked symptom B. In the test phase for blocking, symptoms were presented without corrective feedback. Of several cases tested, one was a combination of symptoms B and I. If symptom B was blocked, then people should prefer the disease paired with control symptom I over the disease paired with blocked symptom B.[1]

In the subsequent, third phase of training, new symptoms and diseases were introduced, wherein ABC$\rightarrow$2 and DEF$\rightarrow$4. The central motivations for this structure are the hypotheses that (a) learners will shift attention away from cues that already have been learned as indicative of different diseases (Kruschke, 1996; Kruschke & Johansen, 1999), and (b) learners will tend not to shift attention toward a cue that they have previously learned to ignore. Specifically, for case DEF$\rightarrow$4, attention will shift away from symptom D, because it is already known to indicate disease 3, leaving attention on the distinctive symptoms E and F. For case ABC$\rightarrow$2, attention will shift away from symptom A, because it is already known to indicate disease 1. If, as a consequence of blocking, people have learned to ignore the blocked symptom B, then attention should not be directed to symptom B, leaving only symptom C significantly attended to. Then there will be only a relatively weak association made between symptom B and disease 2. The strength of the association is assessed in the final test phase, when blocked symptom B and control symptom E are presented together. We predict for this case that people will prefer the disease (4) paired with the control symptom E over the disease (2) paired with the blocked symptom B.[2]

On the other hand, if there is no learned inattention to symptom B when it is blocked, then subsequent learning about it should be unaffected. Therefore, in the third training phase, the blocked symptom B should be as strongly associ-

[1] An alternative assessment of blocking could have asked subjects for *ratings* of disease probabilities, given the blocked symptom B alone or the control symptom I alone. We avoided this dependent measure in the present studies because we wanted the procedure, and hopefully the mental processing, in the test phases to match as closely as possible the procedure in the training phases. We expect, however, that ratings would have yielded analogous results.

[2] The third phase did not examine learning of blocked symptom B *in isolation* because we assumed that people would devote all of their attention to the symptom if it were the only cue available. We assumed that learned inattention to a cue can be most sensitively assessed when alternative cues, such as symptom C, are available to be attended to. We used compound ABC instead of just BC because we assumed that learned inattention to B might be context or stimulus specific, hence the generalization of learned inattention from AB to ABC might be greater than from AB to BC. These assumptions motivated the design, but have not been specifically tested in separate experiments.

Table 1
Designs of Experiments 1 and 2.

	Experiment 1			Experiment 2:		
Phase	Blocking	Control to assess attenuation	Control to assess blocking	Blocking	Control to assess attenuation	Control to assess novelty
Training I	A→1	D→3		A→1	D→3	G→5
Training II	AB→1	D→3	HI→6	AB→1	D→3	G→5 H→?
Test for Blocking		e.g., BI			(not applicable)	
Training III	A→1 ABC→2	D→3 DEF→4	G→5 GHI→6	A→1 ABC→2	D→3 DEF→4	G→5 GHI→6
Test for Attenuation		e.g., BE			e.g., BE, BH, EH	

Note. Letters A–I denote symptoms, numerals 1–6 denote diseases.

ated with the new disease 2 as the control symptom E is associated with the new disease 4. In particular, the RW model as formalized predicts that the associative strength between blocked symptom B and disease 2 is the same as the associative strength between control symptom E and disease 4, *regardless* of the associative strength between blocked symptom B and disease 1. A variety of other symptom combinations are presented in the final testing phase to further constrain the theories.

Table 1 shows that the third training phase also included cases of symptom G paired with outcome 5 and symptoms GHI paired with outcome 6. These cases were included merely to match the third phase of training with the other experiments, to facilitate comparison of results across experiments.

Method

Participants. Forty students (25 female, 15 male, mean age 19.0 years, range 18–21) volunteered for partial credit in an introductory psychology class at Indiana University.

Design. Table 1 shows the abstract design. A complete list of symptom combinations presented during the testing phases is shown in Appendix B. Training Phase I had 20 blocks of 2 trials; training Phase II had 20 blocks of 3 trials; the test for blocking had 2 blocks of 8 test cases; training Phase III had 15 blocks of 6 trials; and, the test for attenuation had 4 blocks of 14 test cases, for a total of 262 trials.

Apparatus and Stimuli. Participants were trained individually in dimly lit, sound-dampened cubicles. They sat before an a IBM-PC compatible computer at a comfortable viewing distance. They made disease diagnoses by pressing the "D," "F," "G," "H," "J" and "K" keys on the standard computer keyboard. The symptoms were "ear ache," "skin rash," "back pain," "dizziness," "nausea," "insomnia," "bad breath," "blurred vision" and "nose bleed." Symptoms were displayed horizontally centered, in a vertical list.

Procedure. Participants were instructed that they would learn to diagnose diseases of hypothetical patients who had certain symptoms. Participants were told that each patient had one and only one disease. The instructions also stated that for some cases the official diagnosis was not yet known, but that these cases should be diagnosed according to what was learned from other cases. The full text of the instructions to the participants is provided in Appendix A.

Within each block of training or testing, the cases were randomly permuted. The assignment of abstract cues A–I to concrete symptoms, and the assignment of abstract diseases 1–6 to concrete response keys, were randomly permuted for each participant. In cases for which multiple symptoms occurred, the order of symptoms in the vertically displayed list was separately randomized on each trial.

A trial began with the presentation of a list of symptoms, and a response prompt below the list. After the participant selected a diagnosis, corrective feedback was displayed. If the response was correct, the feedback said "CORRECT!" otherwise it said "WRONG!" accompanied by a tone. If there was no response within 30 seconds, a higher pitched tone sounded, along with the word "FASTER!" The feedback also indicated the correct diagnosis, with the statement "This patient has disease X," where "X" was the letter of the correct disease. If the trial was a case for which no corrective feedback would be provided, then the feedback stated, "No official diagnosis is yet available. Your diagnosis has been recorded." Beneath the feedback was stated, "After you have studied this case (up to 30 seconds), press the space bar to see the next one." If more than 30 seconds of study time elapsed, a tone sounded, with a reminder that there are only 30 sec-

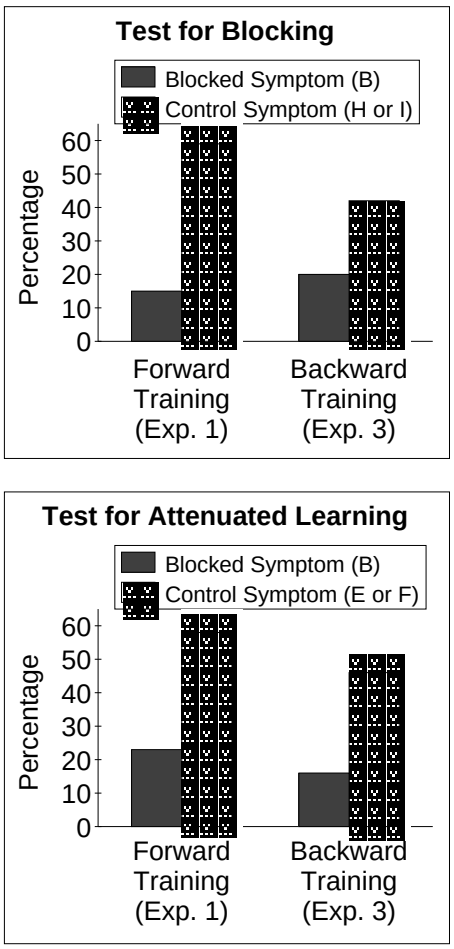

Figure 1. Essential results from Experiments 1 (forward training) and 3 (backward training). Upper panel shows evidence of blocking from the test trials in which the blocked symptom (B) was presented with a blocking control symptom (H or I). The darker bars show the percentage of choices for the disease (1) trained with the blocked symptom, and the lighter bars shown the percentage of choices for the disease (6) trained with the control symptom. The lower panel shows evidence of attenuated learning about the blocked cue, from the test trials in which the blocked symptom (B) was presented with an attenuation control symptom (E or F). The darker bars show the percentage of choices for the disease (2) trained with the blocked symptom, and the lighter bars shown the percentage of choices for the disease (4) trained with the control symptom.

onds of study time, and then the next case was displayed. There were no explicit markers between phases of training, except for the instructions prior to the final testing phase.

Results

Figure 1 displays response percentages for the two critical tests (Appendix B reports all the response percentages in detail). The first test phase evidenced robust blocking (upper panel of Figure 1): When shown the blocked symptom B along with the control symptom H (or I), people preferred the disease trained with the control symptom (i.e., disease 6) over the disease trained with the blocked symptom (i.e., disease 1), 58.8% to 15.0%, $\chi^2(\text{df=1,N=59})/2 = 10.4$, $p < .005$. The χ^2 value has been divided by 2 as the most conservative precaution against possible lack of independence between the two repetitions of the case seen by each participant (Wickens, 1989, p. 28).

Results from the final test phase evidenced notably attenuated learning about the blocked symptom B (lower panel of Figure 1): When shown blocked symptom B with control symptom E (or F), people preferred the disease trained with the control symptom (i.e., disease 4) over the disease trained with the blocked symptom (i.e., disease 2), 58.1% to 22.5%, $\chi^2(\text{df=1,N=129})/4 = 6.30$, $p < .05$. This difference cannot be attributed to lesser learning of ABC$\rightarrow$2 than of DEF$\rightarrow$4, because the test accuracies for these cases were the same (72.5%; see Appendix B).

These results are inconsistent with the hypothesis that blocking is caused entirely by lack of learning about the blocked cue. Instead, the results are consistent with the hypothesis that blocking is caused, at least in part, by learned inattention. One possible alternative explanation is that learning about the blocked symptom is attenuated in the third phase because the symptom is not novel, whereas the control symptoms are novel. This alternative explanation is tested in Experiment 2.

Experiment 2: Novelty does not explain attenuation after forward blocking

The design of Experiment 2 is shown on the right side of Table 1, where it can be seen that that the control to assess blocking, in Experiment 1, is replaced by a control to assess novelty. In this experiment, the second phase of training introduces a control symptom (H) that occurs as frequently as the blocked symptom (B). The control symptom is not given corrective feedback during the second phase; learners are merely told that no diagnosis is yet available for this case. If familiarity, or lack of novelty, accounts for the attenuated learning about the blocked symptom in Experiment 1, then there should be comparable attenuation of learning about the control symptom in this experiment.

Method

Fifty-six students (44 female, 12 male, mean age 18.9 years, range 17–22) volunteered for partial credit in an introductory psychology class at Indiana University. The design is shown in Table 1, and a complete list of symptom combinations presented in the testing phase is shown in Appendix C. Training Phase I had 20 blocks of 3 trials; Phase II had 20 blocks of 4 trials; Phase III had 15 blocks of 6 trials; and, the test for attenuation had 4 blocks of 17 trials, for a total of 298

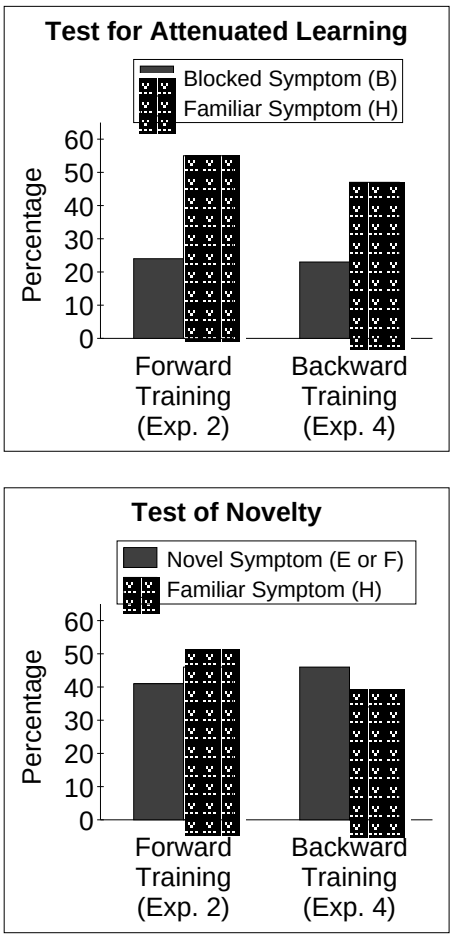

Figure 2. Essential results from Experiments 2 (forward training) and 4 (backward training). Upper panel shows evidence of attenuated learning about the blocked cue, from the test trials in which the blocked symptom (B) was presented with the equally familiar control symptom (H). The darker bars show the percentage of choices for the disease (2) trained with the blocked symptom, and the lighter bars shown the percentage of choices for the disease (6) trained with the control symptom. Lower panel shows no significant effect of novelty, from the test trials in which a novel control symptom (E or F) was presented with the familiar control symptom (H). The darker bars show the percentage of choices for the disease (4) trained with the novel symptom, and the lighter bars shown the percentage of choices for the disease (6) trained with the familiar symptom.

trials. The experiment was conducted identically to Experiment 1, with the same instructions.

Results

Figure 2 displays results from the essential test cases (Appendix C provides full details). There was again strong evidence of attenuated learning about the blocked cue (upper panel of Figure 2): When the blocked symptom B was combined with the equally familiar control symptom H, people preferred the disease trained with the control symptom (i.e., disease 6) over the disease trained with the blocked symptom (i.e., disease 2), 55.4% to 24.1%, $\chi^2(\text{df=1,N=178})/4 = 6.88$, $p < .01$. A direct conflict of the novel control symptom E (or F) and the familiar control symptom H showed no significant difference in preference (lower panel of Figure 2): People chose the disease paired with the novel control symptom (i.e., disease 4) 40.6% of the time, and the disease paired with the familiar control symptom (i.e., disease 6) 45.5% of the time, $\chi^2(\text{df=1,N=193})/4 = 0.16$, n.s.[3]

These results show that attenuated learning about a blocked cue is not due entirely to its familiarity relative to other cues, because a control cue that was equally familiar was not attenuated. Experiment 4, below, shows a small attenuating effect of familiarity, but it is much smaller than the effect of blocking.

Experiment 3: Attenuation after backward blocking

The design of Experiment 3 is identical with Experiment 1 except that the first two phases of training are reversed. With reference to Table 1, the first phase of training presents AB→1, and the second phase of training presents A→1. If symptom B is *backward* blocked, then it should be a weaker indicator of its outcome relative to the control symptoms H and I. This backward blocking is assessed in the first test phase. The third phase of training is the same as the third phases of the previous experiments. If people have learned to suppress attention to the backward blocked cue B, then, as in Experiment 1, people should not learn as strong an association between the blocked cue and its outcome (disease 2) as the association between the attenuation-control cue E (or F) and its outcome (disease 4).

The RW model does not predict backward blocking, because absent cues are assumed to be encoded with a value of zero, and consequently no changes in associative strength can occur for absent cues such as symptom B in the second phase of Experiment 3. Extensions of the RW model, that encode absent cues with negative values, can address backward blocking, but do not predict subsequent attenuation of learning about a backward blocked cue. Like the RW model, these extensions predict that the associative strength between backward blocked symptom B and disease 2 is the same as the associative strength between control symptom E and disease 4, *regardless* of the associative strength between backward blocked symptom B and disease 1.[4]

[3]All statistical tests reported in this article that yield non-significant differences remain non-significant at the .05 level when using less conservative χ^2 values that are not divided by the number of repeated tests per subject.

[4]Predictions of extended RW models, that encode absent but expected cues as negative values, are complicated by the fact that the conflicting cues presented together in test phases of our experiment might entail a number of absent but expected cues. We have con-

Method

Eighty-five students (54 female, 31 male, mean age 19.5 years, range 17–26) volunteered for partial credit in an introductory psychology class at Indiana University. The design was identical to Experiment 1 (see Table 1) except that the first two training phases were reversed. The experiment was conducted identically to Experiments 1 and 2, with the same instructions.

Results

Data from the essential test cases are displayed in Figure 1 by the bars marked as "backward training." (Complete data are presented in Appendix B.) The first test phase (upper panel of Figure 1) manifested robust backward blocking: When shown backward blocked symptom B with control symptom H (or I), people preferred the disease trained with the control symptom (i.e., disease 6) over the disease trained with the backward blocked symptom (i.e., disease 1), 42.4% to 20.0%. This difference is reliable, $\chi^2(\text{df=1,N=106})/2 = 6.8$, $p < .01$. The magnitude of forward blocking obtained in Experiment 1 appears to be greater than the magnitude of backward blocking obtained in Experiment 3, but the difference is not reliable, $\chi^2(\text{df=1,N=164})/2 = 0.84$, n.s., a result consistent with Blair & Kruschke (1999).

Results from the final test phase (lower panel of Figure 1) evidenced strongly attenuated learning of the backward blocked symptom: When shown backward blocked symptom B with attenuation-control symptom E (or F), people preferred the disease trained with the control symptom (i.e., disease 4) over the disease trained with the backward blocked symptom (i.e., disease 2), 45.6% to 16.2%, $\chi^2(\text{df=1,N=210})/4 = 11.9$, $p < .01$. (This difference cannot be attributed to lesser learning of ABC→2 than of DEF→4, because the test accuracies for these cases were not significantly different: 66.2% vs. 75.9%, $\chi^2(\text{df=1,N=483})/4 = 0.56$, n.s.) The preference of disease 4 over disease 2 was not significantly greater for forward blocking than for backward blocking, $\chi^2(\text{df=1,N=339})/4 = 0.03$, n.s. Hence, there appears to be no major difference between the magnitude of attenuation after forward blocking and after backward blocking.

These results are inconsistent with the hypothesis that backward blocking is caused entirely by diminished association of the blocked cue with its original outcome. That is, models of backward blocking that rely on negative encoding of an absent cue cannot account for these results. Instead, the results are consistent with the hypothesis that blocking is caused, at least in part, by learned inattention. Another possible alternative explanation is that learning about the backward blocked symptom is attenuated in the third phase because the symptom is not novel, whereas the control symptoms are novel. This alternative explanation is addressed in Experiment 4.

ducted computer simulations of several variations of these models to confirm our claim that the models cannot account for our results.

Experiment 4: Novelty does not explain attenuation after backward blocking

The design of Experiment 4 is the same as the design of Experiment 3 except that the control to assess blocking, in Experiment 3, is replaced by a control to assess novelty. In other words, Experiment 4 is identical to Experiment 2 (see Table 1) except that the first two phases have been reversed. The initial phase of training introduces a control symptom (H) that occurs as frequently as the backward blocked symptom (B). If familiarity, or lack of novelty, accounts for the attenuated learning about the backward blocked symptom in Experiment 3, then there should be comparable attenuation of learning about the control symptom in this experiment.

Method

Eighty-five students (42 female, 43 male, mean age 19.6 years, range 18–25) volunteered for partial credit in an introductory psychology class at Indiana University. The experiment was conducted identically to Experiments 1, 2 and 3, with the same instructions.

Results

Figure 2 displays results from the essential test cases; complete data are reported in Appendix C. When shown the backward blocked symptom B with the equally familiar control symptom H (upper panel of Figure 2), people preferred the disease trained with the control symptom (i.e., disease 6) over the disease trained with the blocked symptom (i.e., disease 2) 47.2% to 22.7%, $\chi^2(\text{df=1,N=237})/4 = 7.27$, $p < .01$. A direct conflict of the novel control symptoms, E or F, with the familiar control symptom, H (lower panel of Figure 2). showed a small difference consistent with a novelty effect, such that people chose the disease trained with the novel control symptoms 45.9% of the time and the disease trained with the familiar control symptom just 33.8% of the time. This difference was not reliable by our conservative test, however, $\chi^2(\text{df=1,N=271})/4 = 1.55$, $p > .10$. (When not divided by the number of repetitions per subject, the χ^2 value does reach significance at the .05 level.) The magnitude of attenuation after forward blocking was about the same as the magnitude of attenuation after backward blocking: For case BH (upper panel of Figure 2), the preference for the control-symptom disease (6) over the blocked-symptom disease (2) was not significantly different between Experiments 2 and 4, $\chi^2(\text{df=1,N=415})/4 = 0.22$, n.s. These results show that attenuated learning about a backward blocked cue is not due entirely to its familiarity relative to control cues; moreover, learning about a control cue that was equally familiar was only marginally attenuated.

Conclusion: Blocking involves learned inattention

Experiments 1 and 3 demonstrated that a cue that is blocked, whether forward or backward, suffers attenuated associability. Experiments 2 and 4 suggested that the effect cannot be attributed merely to the novelty of the control cues relative to which the attenuation was assessed. The attenuation of a blocked cue cannot be accounted for by the lack-of-learning theory formalized in the RW model, and cannot be accounted for by extensions of the RW model that posit negative encoding of absent cues.

The influential configural model proposed by Pearce (1994) has several advantages over the elemental approach of the RW model (see Pearce, 1994 for a review), but unfortunately the configural model does not predict attenuation of subsequent learning about a blocked cue. The model recruits a configural unit for each distinct cue combination, such that only the exactly matching stimulus fully activates the unit. Output activations are determined by the summed influence of all partially activated configural units, but only the single maximally activated configural unit modifies its associative weights. There is not space here for a detailed review of the configural model, but we will briefly describe the model's predictions for Experiment 1. The critical case to consider is the test for attenuated learning, case BE, and the resulting strengths of activation for outcomes 2 and 4. Stimulus BE activates configural units ABC, DEF, and AB. Configural unit AB has no association with either outcomes 2 or 4, however, and therefore has no *differential* influence on outcomes 2 and 4. Configural units ABC and DEF develop equal magnitude weights to outcome 2 and 4, respectively, and so the model predicts equal activation of outcomes 2 and 4, unlike the strong preference for outcome 4 exhibited by people.

As another possible alternative to attentional theory, the comparator hypothesis (Miller & Matzel, 1988) suggests that all cues acquire learned associations with their outcomes, to the extent that the cue and outcome are contiguous. Expression of a learned association depends, however, on the extent to which it is of greater magnitude than indirect associations via other cues present during training. To our knowledge, the comparator hypothesis has only been applied to situations that involve a single unconditioned stimulus, and therefore it is not clear exactly how the approach should be applied to our multiple-outcome situation. It seems, however, that the comparator hypothesis cannot account for apparent attenuation of learning of a blocked cue in Experiment 1. The critical test case is BE, which has a direct association with 2 from B and indirect associations with 2 via A and C, and which has a direct association with 4 from E and indirect associations with 4 via D and F. These associations to 2 and 4 are symmetric because of their symmetric contiguities during training, and therefore the comparator hypothesis predicts equal response strengths for both outcomes, contrary to the human preference for response 4.

Our attentional explanation assumes that each cue is gated by an attentional strength, with total attention limited in capacity (as in ADIT Kruschke, 1996). The attention allocated to a cue affects both the associability of the cue and the influence of the cue on response generation. Like Pearce's (1994) configural model, and like ALCOVE (Kruschke, 1992) and RASHNL (Kruschke & Johansen, 1999), an exemplar unit is recruited for each distinct cue combination. Beyond those models, however, each exemplar unit encodes not only the presence or absence of cues, but also the attention paid to each cue. Thus, an exemplar unit does not record the *raw* stimulus, but the stimulus *as processed.* The attentional strengths on the input cues, and the attentional strengths within memorized exemplars, (and the associative weights,) are shifted to reduce error. The remembered attention strengths feed back onto the input gates, so that attentional distributions are learned for specific stimulus configurations. For example, consider the case of forward blocking. During initial trials of AB$\rightarrow$1 training (after previous A$\rightarrow$1 training), symptom B attracts some attention by default, thereby reducing attention to symptom A, and consequently reducing the magnitude of response 1 (because, as stated above, attention affects response production as well as associability). To alleviate this error, attention is shifted away from B toward A. The exemplar unit that encodes configuration AB also encodes this reduced attention to B. When ABC is presented subsequently, it partially activates exemplar unit AB, which feeds back its learned suppression of attention to B, and hence causes reduced attention to B during learning about ABC. That is, there is attenuated learning about a previously blocked cue. Consider now the case of backward blocking. During trials of A$\rightarrow$1 training (after previous AB$\rightarrow$1 training), the AB exemplar is only partially activated, and so the correct response is only partially evoked. To alleviate this error, attention *within the AB exemplar unit* is shifted toward A, away from B. Subsequent training with ABC partially activates exemplar unit AB, which feeds back its learned suppression of attention to B, and hence causes reduced attention to B during learning about ABC. That is, there is attenuated learning about a previously backward blocked cue. Formal specification and computer simulation of this theory is presently underway in our lab. Accurate modeling of attenuated learning after forward blocking (Experiment 1) has already been completed, along with mathematical analysis showing that Mackintosh's classic (1975) model of attention learning is very nearly a special case of two different connectionist models (Kruschke, 1999).

We make no claim that attention alone can account for all the interesting phenomena in associative learning. The attentional account just outlined assumes error-driven learning of associations as in the RW model, and assumes a form of configural encoding similar to Pearce's (1994) model. A complete model of associative learning might also have to include negative encoding of absent-but-expected cues or a comparator mechanism for response generation. Our claim is that

these non-attentional mechanisms, as presently formulated in the literature, cannot account for attenuation of learning after forward or backward blocking, but an attentional mechanism can.

Learned attention shifts are an efficacious solution to the problem of having to learn associations quickly, without interfering with previously learned associations. Attention is shifted toward cues that have associations that are already evoking the correct response. Attention is shifted away from cues that have been previously associated with a different response. In this way, previous learning is protected and preserved, while new learning is accelerated. These attentional shifts satisfy the need for speed in learning, but also produce side effects that can be interpreted as irrational or nonnormative. Kruschke & Johansen (1999) describe how a wide spectrum of these effects can be explained by rapid attention shifts.

References

Arcediano, F., Matute, H., & Miller, R. R. (1997). Blocking of Pavlovian conditioning in humans. *Learning and Motivation, 28*, 188–199.

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*, 414–432.

Blair, N. J. & Kruschke, J. K. (1999). Retrospective revaluation and configural learning in human categorization. Submitted for publication. Available via WWW at http:// www.indiana.edu/ ~kruschke/ blairkruschke99.html.

Chapman, G. B. (1991). Trial order affects cue interaction in contingency judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 837–854.

Chapman, G. B. & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory and Cognition, 18*, 537–545.

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

Dickinson, A., Shanks, D. R., & Evenden, J. L. (1984). Judgement of act-outcome contingency: The role of selective attribution. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *36A*, 29–50.

Domjan, M. (1998). *The Principles of Learning and Behavior* (4th edition). Pacific Grove, CA: Brooks/Cole.

Hinchy, J., Lovibond, P. F., & Ter-Horst, K. M. (1995). Blocking in human electrodermal conditioning. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, *48B*, 2–12.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment*, pp. 279–296. New York: Appleton-Century-Crofts.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *22*, 3–26.

Kruschke, J. K. (1999). Toward a unified model of attention in associative learning. Submitted for publication. Available online http:// www.indiana.edu/ ~kruschke/ tumaal.html.

Kruschke, J. K. & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *25*, 1083–1119.

Larkin, M. J. W., Aitken, M. R. F., & Dickinson, A. (1998). Retrospective revaluation of causal judgments under positive and negative contingencies. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*(6), 1331–1352.

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

Mackintosh, N. J. & Turner, C. (1971). Blocking as a function of novelty of CS and predictability of UCS. *Quarterly Journal of Experimental Psychology*, *23*, 359–366.

Markman, A. B. (1989). LMS rules and the inverse base-rate effect: Comment on Gluck and Bower (1988). *Journal of Experimental Psychology: General*, *118*, 417–421.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin, 117*, 363–386.

Miller, R. R. & Matute, H. (1996). Biological significance in forward and backward blocking: Resolution of a discrepancy between animal conditioning and human causal judgment. *Journal of Experimental Psychology: General*, *125*, 370–386.

Miller, R. R. & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation*, Vol. 22, pp. 51–92. San Diego, CA: Academic Press.

Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, *101*, 587–607.

Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning: II. Current Research and Theory*, pp. 64–99. New York: Appleton-Century-Crofts.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *Quarterly Journal of Experimental Psychology*, *37B*, 1–21.

Siegel, S. & Allan, L. G. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin & Review*, *3*, 314–321.

Sutherland, N. S. & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press.

Tassoni, C. J. (1995). The least mean squares network with information coding: A model of cue learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 193–204.

Van Hamme, L. J. & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation, 25*, 127–151.

Waldman, M. R. & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121*, 222–236.

Wasserman, E. A. & Berglan, L. R. (1998). Backward blocking and recovery from overshadowing in human causal judgment: The role of within-compound associations. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, *51B*, 121–138.

Wasserman, E. A., Kao, S.-F., VanHamme, L. J., Katagiri, M., & Young, M. E. (1996). Causation and association. In D. R. Shanks, D. L. Medin, & K. J. Holyoak (Eds.), *The Psychology of Learning and Motivation: Causal Learning*, Vol. 34, pp. 207–264. Academic Press.

Wickens, T. D. (1989). *Multiway contingency tables analysis for the social sciences*. Hillsdale, NJ: Erlbaum.

Williams, D. A., Sagness, K. E., & McPhee, J. E. (1994). Configural and elemental strategies in predictive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 694–709.

Appendix A: Full text of instructions to participants

At the beginning of the experiment, the participant was seated in front of the computer, with the following instructions displayed on the screen:

> INSTRUCTIONS
>
> This experiment examines how people learn to make accurate medical diagnoses. You will be presented with many patients' case histories. For each case history you will be shown the symptoms the patient has, and you will be asked to choose which illness you think the patient has. After you make your diagnosis, you will be told the correct diagnosis. All you have to do is try to learn which symptoms tend to go with which illnesses so that you can make as many correct diagnoses as possible.
>
> Re-read the previous paragraph if it is unclear.
>
> Then, press the space bar to continue.
>
> There are six possible diseases that the patients have, and each patient has one and only one of the diseases. In order to keep things as straight-forward as possible, we'll simply label the diseases with letters D, F, G, H, J and K. For each case history, you indicate your diagnosis by pressing one of these six letters on the keyboard. You'll have up to 30 seconds to make your diagnosis for each case history. At first you will just be guessing, but after many cases your accuracy will improve.
>
> If any of the instructions on this screen are unclear, please re-read them now. Otherwise, press the space bar to continue.
>
> Various diseases will be gradually phased into the learning, so that you not learn all six diseases right away.
>
> Try to be as accurate as possible! Don't let your attention wane. For each case, silently repeat to yourself the symptoms and correct disease. This will help you learn.
>
> Press the space bar to continue.
>
> There will occasionally be cases for which the official diagnosis is not yet known, and so you will not be told the officially correct diagnosis. In these cases, just make your best educated guess based on the other cases you have learned about.
>
> To reiterate, your task is to learn which symptoms tend to go with which diseases, so that you can make as many correct diagnoses as possible.
>
> Press the space bar to continue.
>
> If you have any questions, please ask now.
>
> WAIT for the experimenter to close the cubicle curtain and leave the room.
>
> Press the space bar to begin the experiment.

Before the final test phase, the computer displayed the following instructions:

> IMPORTANT! READ THIS BEFORE CONTINUING!
>
> In the next part of the experiment you must again make diagnoses on the basis of case histories. There will be cases with just single symptoms, or with combinations of symptoms you might not have seen before. In these cases, make your best guess based on what you learned from the earlier part of the experiment. You will not be told the correct diagnosis, but your best educated guess is very important for our study, so DON'T just respond randomly in these cases!
>
> Because you will not be told the correct diagnosis, you should try not to alter your opinion on the basis of these cases. Base your opinion solely on what you learned in the previous part of the experiment.
>
> Press the space bar to continue.

Appendix B: Choice percentages in test trials of Experiments 1 and 3.

	Experiment 1: Forward Blocking						Experiment 3: Backward Blocking					
	Disease						Disease					
Symptoms	1	2	3	4	5	6	1	2	3	4	5	6
Test for Blocking:												
AB	81.3	3.8	3.8	3.8	5.0	2.5	77.6	6.4	2.4	4.1	5.3	4.1
D	2.5	0.0	96.3	1.3	0.0	0.0	1.2	0.0	95.9	0.6	1.8	0.6
HI	1.3	0.0	0.0	1.3	2.5	95.0	7.1	3.5	4.1	3.5	1.8	80.0
BH/BI	**15.0**	6.3	3.8	6.3	10.0	**58.8**	**20.0**	12.4	4.7	11.2	9.4	**42.4**
BD	15.0	3.8	66.3	6.3	2.5	6.3	15.9	14.7	42.4	12.4	10.6	4.1
AD	42.5	13.8	30.0	2.5	11.3	0.0	25.3	15.3	31.2	16.5	10.0	1.8
AH/AI	65.0	1.3	1.3	5.0	3.8	23.8	35.9	9.4	5.9	15.3	9.4	24.1
DH/DI	2.5	6.3	43.8	5.0	10.0	32.5	3.5	15.9	33.5	10.6	10.6	25.9
Test for Attenuation:												
A	94.4	1.3	1.3	1.3	0.6	1.3	90.9	0.9	3.2	1.5	1.8	1.8
ABC	13.8	72.5	0.6	4.7	2.5	5.0	17.4	66.2	2.6	6.8	0.9	6.2
D	3.1	0.6	93.8	1.3	0.0	1.3	0.9	1.5	92.6	2.6	1.5	1.2
DEF	1.3	5.0	12.5	72.5	2.5	6.3	2.1	7.1	7.1	75.9	2.9	5.0
G	0.6	0.0	1.9	0.6	95.6	1.3	0.9	1.5	0.6	1.8	93.2	2.1
GHI	1.3	6.3	2.5	5.6	6.9	77.5	0.6	6.8	0.9	5.6	7.9	78.2
BE/BF	2.5	**22.5**	2.5	**58.1**	3.1	11.3	19.4	**16.2**	3.8	**45.6**	2.9	11.8
CE/CF	1.3	39.4	5.0	42.5	3.1	8.8	4.7	35.9	4.1	44.4	3.2	7.6
BH/BI	1.3	21.9	0.6	6.3	3.8	66.3	20.6	17.4	3.2	5.6	2.6	50.6
CH/CI	1.9	48.8	1.9	3.1	5.0	39.4	7.4	43.8	2.9	7.9	2.4	35.6
AB	56.9	27.5	3.1	4.4	0.6	7.5	60.0	21.2	3.5	7.4	1.8	6.2
AC	34.4	56.3	1.3	1.3	3.1	3.8	32.6	50.3	3.8	6.2	1.2	5.9
DE/DF	1.3	6.3	31.3	51.9	3.1	6.3	2.1	5.9	31.8	52.1	3.2	5.0
GH/GI	0.6	4.4	3.8	3.8	34.4	53.1	3.2	6.2	1.8	6.8	28.2	53.8

Note. Letters A–I denote symptoms, numerals 1–6 denote diseases. A slash denotes structurally equivalent cases collapsed into a single row; e.g., BE/BF indicates results for cases BE and BF combined. Data in bold font are plotted in Figure 1.

Appendix C: Choice percentages in test trials of Experiments 2 and 4.

	Experiment 2 Forward Blocking						Experiment 4 Backward Blocking					
	Disease						Disease					
Symptoms	1	2	3	4	5	6	1	2	3	4	5	6
A	93.3	1.3	1.8	1.8	1.8	0.0	93.5	1.5	2.1	0.6	1.5	0.9
ABC	4.9	83.5	1.3	5.8	0.9	3.6	10.9	79.4	1.8	2.1	0.9	5.0
D	1.3	0.4	94.2	2.2	1.3	0.4	1.2	2.1	91.8	1.8	2.9	0.3
DEF	1.3	2.2	5.4	87.1	0.9	3.1	1.5	4.4	7.1	82.4	0.9	3.8
G	0.9	0.9	0.9	0.4	95.5	1.3	2.6	0.6	1.2	1.8	92.1	1.8
GHI	0.4	3.1	0.4	2.7	8.9	84.4	0.3	2.4	1.5	4.7	6.8	84.4
BE/BF	4.5	29.9	2.2	54.0	0.9	8.5	10.3	24.1	2.6	51.2	2.1	9.7
CE/CF	1.3	49.6	0.9	43.3	0.4	4.5	1.8	44.7	2.4	43.5	1.5	6.2
BH	5.4	**24.1**	0.0	13.8	1.3	**55.4**	13.6	**22.7**	3.2	11.8	1.5	**47.2**
AB	46.9	38.8	1.3	6.7	0.4	5.8	51.2	31.8	2.1	6.5	1.8	6.8
AC	20.5	70.5	0.4	5.4	0.4	2.7	19.5	67.6	1.5	3.8	2.1	5.6
DE/DF	0.4	6.7	23.2	63.8	0.9	4.9	2.1	4.4	25.0	60.3	2.9	5.3
GH	0.4	8.9	1.3	5.8	17.4	66.1	1.5	4.4	1.5	5.6	26.5	60.6
GI	0.9	5.4	1.8	5.8	21.0	65.2	2.6	6.5	1.2	7.6	25.6	56.5
EH/FH	0.4	10.3	1.3	**40.6**	1.8	**45.5**	1.8	11.2	5.3	**45.9**	2.1	**33.8**
EI/FI	0.9	7.1	2.2	46.0	0.0	43.8	2.4	6.5	1.2	45.0	4.7	40.3
CI	2.2	39.7	0.9	6.7	0.0	50.4	3.2	47.6	2.1	4.7	3.5	38.8

Note. Letters A–I denote symptoms, numerals 1–6 denote diseases. A slash denotes structurally equivalent cases collapsed into a single row; e.g., BE/BF indicates results for cases BE and BF combined. Data in bold font are plotted in Figure 2.