

## Attention and salience in associative blocking

STEPHEN E. DENTON and JOHN K. KRUSCHKE  
*Indiana University, Bloomington, Indiana*

The associative learning effect called blocking has previously been found in many cue-competition paradigms where all cues are of equal salience. Previous research by Hall, Mackintosh, Goodall, and dal Martello (1977) found that, in animals, salient cues were less likely to be blocked. Crucially, they also found that when the to-be-blocked cue was highly salient, the blocking cue would *lose* some control over responding. The present article extends these findings to humans and suggests that shifts in attention can explain the apparent loss of control by the previously learned cue. A connectionist model that implements attentional learning is shown to fit the main trends in the data. Model comparisons suggest that mere forgetting, implemented as weight decay, cannot explain the results.

The associative learning effect known as blocking occurs when a cue that has already been learned to predict an outcome is paired with a novel cue (Kamin, 1968). People and animals tend not to associate the new cue with the outcome; that is, learning about the cue has apparently been blocked. Blocking is a historically crucial finding, because it disconfirms many models of learning in which associative strength is incremented by the mere contiguity of cue and outcome, without any competitive mechanisms.

The dominant explanation of blocking was formalized in the classic model of Rescorla and Wagner (1972). The idea is that associative strength changes only to the extent that the outcome is unexpected. If a cue comes to predict the outcome, then the outcome will still be expected when the cue is later paired with the novel cue. Because the outcome is expected, there will be little or no change in associative strength. The novel cue is blocked simply because an association between it and the outcome is not learned. An alternative explanation of blocking was propounded by Mackintosh (1975), whose attentional theory suggested that something is learned about the redundant relevant cue—namely, that it should be ignored. Blocking is therefore viewed as learned inattention or learned attention shifting.

The central role of attention in learning has been emphasized repeatedly, and researchers have long studied how humans and animals learn to allocate attention across potentially informative cues (Trabasso & Bower, 1968). When cues compete, the cue that is allocated more attention will dominate learning. Therefore, shifts of attention, and the learning of these shifts, have been implicated in a variety of cue-competition phenomena.

Attention is affected by the relative validity of cues. Studies indicate that a partially reinforced cue is utilized more in isolation than when it is experienced in compounds containing more valid elements (Wagner, Logan, Haberlandt, & Price, 1968). Edgell et al. (1996) found that even the addition of irrelevant cues has a degrading effect on the utilization of relevant information; and when additional cues are relevant, the initial cues will become even less utilized. Ultimately, a cue with the greater validity has a greater probability of attracting attention, at the expense of attention to other cues. This seems accurate, however, only if all cues are of equal salience (see, e.g., Kruschke & Johansen, 1999). In the case that some cues are of higher salience, the highly salient cues also have more power to attract attention from the other cues. The conflict between attending to valid cues versus salient cues has important implications on learning theories. Attending to salient cues could be deemed irrational, because all that should matter, ultimately, is the actual validity of the cue, not the salience.

In an effort to uncover the parallels between the effects of relative salience and relative validity on the amount of control exercised by a stimulus, Hall, Mackintosh, Goodall, and dal Martello (1977) performed three conditioning experiments with animals (rats and pigeons) using cues of differing salience. Their study suggests that the effect of learning highly salient cues parallels the effect of learning highly valid cues. Just as a highly valid cue will overshadow less valid cues and achieve greater control over responding, a highly salient cue will overshadow less salient cues and become more associated with the outcome. In particular, if the initially learned cue is combined with a more salient cue in compound training, then the blocking effect will be reduced or will be absent altogether. The magnitude of this blocking effect was already known to be a function of the relative saliences of the blocking cue (i.e., the initially trained cue) and the blocked cue (Feldman, 1975; Kamin, 1969). The finding that a salient to-be-blocked stimulus is less likely to be blocked has been replicated, using rats, by Arcediano, Escobar, and Miller

---

This research was supported in part by National Science Foundation Grant BCS-9910720. We thank Patrick Hill, Blake Hulet, Angela Prather, Ruth Ann Weaver, Jimmy Burroughs, Kendra Gibbs, and Daniel Spechar for assistance administering the experiment. Correspondence can be addressed to J. K. Kruschke, Department of Psychological and Brain Sciences, 1101 E. 10th St., Indiana University, Bloomington, IN 47405-7007 (e-mail: kruschke@indiana.edu; www.indiana.edu/~kruschke).

(2004), and the finding has been extended to flavor aversion learning (Klein, Weston, McGee-Davis, & Cohen, 1984).

The surprising and theoretically challenging finding of Hall et al. (1977) involved the blocking stimulus. In animals, when a low-salience stimulus signals reinforcement by itself, it will acquire significant control over behavior. But this initially learned stimulus will *lose* some of that control if it is subsequently reinforced only in combination with a second more salient stimulus. This relative salience effect parallels the well-established relative validity effect, wherein a partially reinforced cue will show a loss of control when it is paired with a more valid cue (see Kruschke & Johansen, 1999, and references therein). In Hall et al.'s results, even though the less salient blocking cue is still a perfectly valid predictor of an outcome, it will lose some control over responding.

This finding is of great consequence, because previous models of learned associations, such as the Rescorla–Wagner (1972) model do not predict this result. In the Rescorla–Wagner model, the associative strength between a stimulus and a response can be reduced only if the stimulus is no longer a valid predictor of the response. The loss of associative strength by a stimulus critically depends on overpredicting the outcome when the cue is present. In connectionist terms, the weight from a cue to a response can decrease only if on some trial the cue is present and the outcome is overpredicted. In the Hall et al. (1977) procedure, the low-salience stimulus continues to be a valid predictor and the outcome should never be overpredicted, but the cue's ability to elicit a response is nonetheless reduced.

Attention shifting theory, on the other hand, can account for this result. Even if the learned associative weight between the blocking cue and the response remains strong, attention can still be shifted away from the cue. The salient to-be-blocked cue will draw attention away from the faint blocking cue. This shift of attention causes reduced responding to the blocking cue.

The effects of salience in blocking have not yet been studied in humans. Thus, this study's important empirical contribution is to uncover how human learning in a blocking task is aided or disrupted by salience manipulation and to discover the effects of salience on attention shifting. The control condition for the experiment will be the standard blocking task with no salient cues, and the experimental condition will involve making particular cues salient, including the normally blocked cue. Our design includes a rich set of within- and between-subjects comparisons that have not been assayed previously. The hypothesis for the experiment is that highly salient cues will disrupt normal attention shifting in the blocking task. The predicted outcome of the experiment should replicate for humans the main findings of the animal experiment conducted by Hall et al. (1977). Thus, we predict that blocking will be reduced if the normally blocked cue is salient, and that the normally blocking cue will lose control over responding when the normally blocked cue is salient.

Results of computer simulations of the EXIT model (Kruschke, 2001a, 2001b) are presented. The modeling explores the theoretical issues in more rigorous detail. There are three theoretical questions that are of importance in this study. First, is there an attentional interpretation of the expected reduction in blocking of salient stimuli? Second, is there an attentional interpretation of the expected reduction in the strength of the blocking cue? Third, is there another mechanism besides attention that could explain the expected reduction in the strength of the blocking cue? This final question required us to compare an attentional mechanism with an alternative mechanism that seemed most appropriate to show a reduction in cue strength. For the comparison, we will consider a model that includes a weight decay mechanism that implements forgetting. The conclusions drawn from our modeling efforts indicate that attentional learning, as formalized in the EXIT model, provides an account of the reduction in blocking of salient stimuli and of the reduction in the strength of the blocking cue, whereas a weight decay mechanism is found to be lacking in its ability to explain the reduction in the blocking cue's strength. Attentional learning is therefore emphasized as a crucial mechanism needed to account for our results.

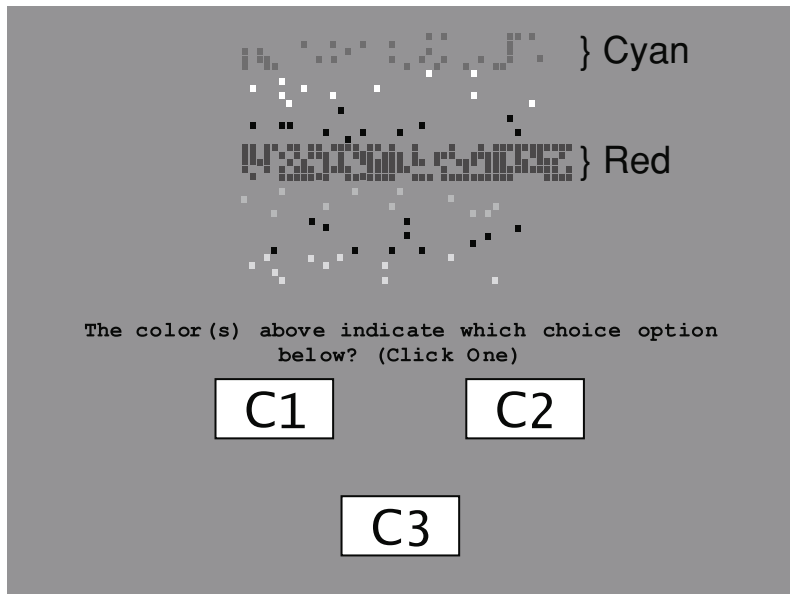
## EXPERIMENT

### Effects of Salience on Blocking

Our goal was to compare a standard blocking task, in which all cues are of equal salience, with a task in which some cues are highly salient. In our stimuli, each cue was a distinct color, displayed as an array of dots. The salience of a cue was the relative density of its color dots. An example of a stimulus display is shown in Figure 1. The stimulus is separated into seven bar fields, each of which contains one color. When a cue is present, the corresponding color has an increased dot density in its respective field. When a cue is salient, its dot density is increased further. Participants made responses by mouse clicking one of the three boxes labeled with response options.

Every participant went through the same blocking design. Details of the implemented design are shown in Table 1. Each of the cues, denoted by the letters A, B, C, and so forth in Table 1, corresponds to a particular color that appeared at a greater dot density when the cue was present.

The experimental design was split into three phases. The early training phase required participants to learn a one-to-one pairing of cues A, C, and F to responses. The late training phase contained three types of trials: (1) A.B trials—the blocking manipulation; (2) C trials—continued training on a single stimulus; and (3) D.E trials—a novel compound. Cue B was designed to be blocked by the already learned cue A. Cue E acts as a control or comparison cue to B, while cues C and F act as controls or comparison cues for A. The final testing phase included trials of all the individual cues as well as a novel cue, and many pairwise combinations that we believed would further inform and constrain theoretical interpretations.



**Figure 1.** Example stimulus display from the second training phase. This stimulus contains two cues: one salient, corresponding to the densely dotted color red in the fourth stripe of the array, and one faint (nonsalient), corresponding to the relatively sparsely dotted color cyan in the first stripe of the array. All other colors were presented at a very sparse background density. The words *red* and *cyan* did not appear in the actual stimulus display; the words appear here to help the reader interpret this grayscale reproduction.

Two experimental conditions differed only in the salience of the stimuli. In the equal-salience condition (ESC), participants saw cues of all equal salience. In the high-salience condition (HSC), participants saw two cues that had greater dot densities than did the other five cues whose densities matched those in the ESC. These cues (B and E) were always salient when presented. The higher salience is indicated by boldface in Table 1.

**Method**

**Participants**

Ninety-three students volunteered for partial credit in an introductory psychology class at Indiana University. Students participated only if they stated that they had normal color vision (no color blindness), but no test of color vision was conducted. Participants

were assigned to two groups based on their subject number (assigned sequentially as they arrived in the lab), even numbers in one group and odd numbers in the other. Forty-six participants were in the experimental group (HSC), and 47 participants were in the control group (ESC).

**Stimuli**

The stimuli were random pixilated dot patterns within a 35 × 45 cell array positioned in the top half of a computer screen. An example of the computer display is shown in Figure 1. Seven colors were used as cues. The colors were as follows (accompanied in parentheses by their corresponding red-green-blue [RGB] values on a 0–255 scale): red (200, 0, 0), orange (254, 174, 0), yellow (240, 240, 0), green (0, 230, 0), cyan (0, 180, 180), blue (0, 0, 200), and purple (128, 0, 153). The brightness, saturation, and hue of the seven colors were adjusted so that each color was roughly equal in salience, as subjectively judged by the experimenters. To judge color salience,

**Table 1**  
**Experimental Design**

Phase	Cue(s) → Correct Response × Frequency			No. Blocks
Early	A → R1 × 2	C → R2 × 2	F → R3 × 2	To criterion (8 minimum)
Late	<b>A</b> , <b>B</b> → R1 × 2	C → R2 × 2	D, E → R3 × 2	15
Test	<b>A</b> , <b>B</b> → R1 × 3	C → R2 × 3	D, E → R3 × 3	2
	A → (R1, R2, R3)	D → (R1, R2, R3)	F → (R1, R2, R3)	
	<b>B</b> → (R1, R2, R3)	<b>E</b> → (R1, R2, R3)	N → (R1, R2, R3)	
	A, C → (R1, R2)	A, E → (R1, R3)	C, D → (R2, R3)	
	<b>B</b> , C → (R1, R2)	<b>B</b> , D → (R1, R3)	C, E → (R2, R3)	
	A, D → (R1, R3)	<b>B</b> , E → (R1, R3)	C, F → (R2, R3)	
	A, F → (R1, R3)	A, N → (R1, R3)	C, N → (R2, R3)	

Note—Each cell indicates cues → correct response. If there is no correct response, allowed responses are shown in parentheses. The cues in bold were salient for the high-salience condition.

equal numbers of dots for each color were compared in an effort to make sure that none of the seven colors “stood out.” It was our desire that the colors themselves not interfere with the salience manipulation of dot density. All the color dots appeared against a grey (150, 150, 150) background.

The colors were separated so that dots of a particular color would appear only within a horizontal bar field that made up one seventh of the pixel array; see Figure 1. Within the bars, the dispersion of the color dots was randomized on each trial. The ordering of the horizontal bars from top to bottom of the stimulus display was randomized on each trial, so that the color locations differed each time a stimulus was displayed.

All seven colors appeared in every stimulus array (except in the initial part of early training as explained below). The correct response was indicated by which color or colors were more prevalent. In each stimulus array, one or two of the seven colors were presented with an increased density (i.e., there were more dots of the particular color). The cue color’s dots took up a larger proportion of the bar field that contained that particular color. In Figure 1, for example, red fills 65% of its horizontal bar field, cyan fills 15% of its field, and all other colors are present at 5% in their respective fields.

In the ESC, all the cue colors when “present” appeared at a dot density of 15% (faint). In the HSC, some cue colors appeared at a density of 65% (salient), while others remained at 15% (faint). In both conditions, the background color density for all “absent” cues was 5%.

### Procedure

Participants sat individually before a standard desktop computer, in a dimly lit, ventilated, sound-dampened booth. They responded by using the mouse to click on one of three response choices, labeled C1, C2, and C3 (see Figure 1). The assignment of response choices to color cues was randomized separately for each participant.

Written instructions were presented on the computer screen; participants read them at their own pace. The instructions indicated that the task was to match the color dot stimuli with the responses. Participants were advised that in order to progress through the experiment they needed to maintain high accuracy. The instructions indicated that participants should attend to color in the stimuli and to ignore the locations of the dots themselves, because they were random. The participants were advised that initially they would be guessing, but that feedback would be provided so that they could quickly learn matches from cues to response items. The instructions also indicated that some of the cues might be novel and would not be provided with feedback, so that the participants should make their best guess. (Complete instructions are included in Appendix A.)

Participants went through three phases in the experiment, shown in Table 1. The only difference between the two conditions was that in the HSC the colors assigned to the B and E cues appeared at a greater density (65%, as described previously) than in the ESC (15%). The experiment progressed through multiple blocks of early training until an accuracy criterion was reached and then through 15 blocks of late training; it ended with 2 blocks of testing. Each block in the early and late training phases consisted of six trials, two of each stimulus type randomly ordered. To ensure that the early training items A, C, and F were well learned by all the participants, an accuracy criterion was included in the early training phase. The accuracy criterion was set at 11 out of 12 correct across the previous two blocks. Each block of the testing phase included 27 trials, which included all of the stimulus types shown in Table 1. The entire experiment lasted approximately 20 min.

A trial consisted of the stimulus displayed on the computer screen along with the three response choices and a response prompt (see Figure 1). In the early and late training phases, after a response choice was selected, feedback was given. The feedback text was displayed in the center of the computer screen. The participant was told whether the response was correct or incorrect, and the correct response was shown. In the testing phase, only the A.B, C, and D.E

trials provided feedback. All the other trials in the testing phase provided no feedback; instead, text was shown that merely indicated that the response had been recorded.

Several time constraints were placed on participants. The color dot stimulus appeared for only 1 sec, so that participants would not have time to attempt to learn dot locations (they were uninformative anyway). Once a response choice was selected, feedback was presented for only 1 sec. If the participant responded incorrectly, an error tone sounded and there was a 1-sec delay before the participant could continue; this delay was included to motivate participants to improve their accuracy. After 1 sec, the screen cleared, and a “continue” button remained centered on the screen. Participants had to click on the “continue” button to proceed; this button also served to recenter the cursor and initiate the next trial.

Each block of the testing phase included each stimulus type shown in Table 1 once, except for the late training items (A.B, C, and D.E), each of which appeared three times. These items appeared three times more often in the testing phase so that participants were still being given feedback regularly. This feedback was intended to reduce the amount of confusion that a participant would experience in going through 21 trials and receiving feedback on only 3 of them. Instead, participants went through 27 trials in each testing block, 9 of which provided feedback.

To facilitate the learning of the early training items, two blocks of training were initially done with no background colors, so that the stimuli were made up of only one horizontal bar field. Then the background density was increased throughout the rest of the early training phase until it reached the level for the late training and testing phase (5%). The background noise was ramped up in 1% intervals, starting at 0%. If the participant maintained an accuracy above 90% correct over two consecutive blocks, then the stimulus would step up to the next background density level until it reached 5%. Participants were required to do two blocks at the 5% level with high accuracy (above 90%) before proceeding to the next training phase (late training). The early training phase could be done in as few as 8 blocks, if the participant maintained an accuracy above 90%. The maximum number of blocks of early training allowed was set at 30, but none of the participants reached this cutoff.

In the final testing phase, some choice responses were not allowed. If a participant chose the disallowed response option, he or she would get a message saying that the response was invalid, and that he/she should choose the best alternative. The allowed responses are summarized in Table 1. The restriction of response choices was intended to reduce participants’ use of nonassociative strategies (discussed later) when they were presented with novel cue combinations. In general, response choices that were never paired with either cue in the stimulus display were not allowed. For example, test cues A.C had response R3 disallowed because R3 never occurred with A or C during training. The novel cue N was included to test whether people treated F as novel in the final test phase. Therefore, cue combination A.N had the same allowed responses as did A.F, and C.N had the same allowed responses as did C.F.

### Results

Because we were specifically interested in what learners learned, we needed to make sure that participants showed good accuracy on all training items. The first training phase included an accuracy criterion. In the second training phase, the D.E cue combination was the only completely novel stimulus, so we judged second-phase learning on the basis of the participant’s ability to respond correctly to this item.

The D.E stimulus appeared six times in the final test phase. For our data analysis, we included only participants who were correct 5 or 6 times on this item in the test phase. The probability of getting 5 or 6 correct out of 6

total would be less than 5% if the participant was guessing (if it is assumed that guessing would be correct one-third of the time). This criterion eliminated 20 participants in the ESC, leaving 27, and eliminated 14 in the HSC, leaving 32. The number of excluded participants in the ESC did not differ significantly from the number excluded in the HSC [ $\chi^2(1, N = 93) = 1.472, p = .23$ ]. The excluded participants did not respond extremely differently from those who were included, and there were no reversals of trends in the data when the excluded participants were included.

All statistical tests were two-tailed, with significance (i.e., reliability) taken to be  $p < .05$ . All chi-squared tests assumed independence of responses across trials within subjects.

The primary results are shown in Table 2. The full results appear in Appendix B, Table B1. Response choices that were not allowed are indicated by dashes in these tables; these responses were considered invalid. Participants' invalid responses were recorded and are summarized in Appendix B, Table B2.

### Late Phase Maintenance

Performance in the test phase on late training items was good, as can be seen in Table 2. The test phase had six A.B trials per participant, with the ESC having an average performance of 96% correct and the HSC having an average performance of 92% correct. Performance on the six C trials was 95% correct for the ESC and 95% correct for the HSC. After the elimination of nonlearners, performance on the six D.E trials was necessarily good, with the ESC having an average accuracy of 95% and the HSC having an average accuracy of 96%.

### Is Cue B Blocked?

The strength of the blocked cue can be gauged by the B-alone test trials, of which each participant saw two. Table 2 shows 22% R1 responses in the ESC, but 70% R1 responses in the HSC. This between-subjects comparison shows that B was better able to control responding in the HSC than in the ESC. To test the reliability of this difference, the number of R1 responses was counted for each

participant. The average number of R1 responses for the ESC was 0.44 out of 2 possible (participants responded R1 22% of the time) and mean R1 responses for the HSC was 1.41 (70% R1 responses). The mean difference of 0.96 (48%) was significant [ $t(57) = 4.61, p < .001, d = 1.21$ ] (95% CI = 27%, 69%).<sup>1</sup> This suggests that blocking was stronger in the ESC than in the HSC.

A within-subjects test of blocking involves putting the blocked cue B in conflict with the control cue E. In the ESC, we looked at any trial that combined the blocked cue B with control cues D or E, because D and E were equivalent in this condition. In the HSC, blocking was judged only in comparison with the E cue, because B and E had high salience and all other cues were less salient.

In the ESC, when participants were presented with a B.D or B.E test trial, they chose the response corresponding with the control cue (R3) 78% of the time, and the response corresponding with the blocked cue (R1) only 22% of the time. This preference for R3 indicated strong blocking. To test the reliability of this preference, we considered each participant's four test trials that combined the blocked cue B with the control cues D and E. The participant's number of R3 responses (the response consistent with D and E) was counted. If participants were at chance levels, then the mean response would be 2. The mean response across the group's 27 participants was 3.11, which was significantly different from chance levels [ $t(26) = 5.15, p < .001, d = .99$ ]. Thus, there was a robust blocking effect in the ESC.

In the HSC, when participants were presented with a B.E test trial, they chose the response corresponding with E (R3) 66% of the time and the response corresponding with B (R1) 34% of the time. There were only two test trials per subject that combined B and E, so if participants were responding at chance levels we would expect their mean response to have been 1. The mean response was 1.31, which is significantly different from chance levels [ $t(31) = 2.27, p = .031, d = .40$ ]. Thus, there was also significant blocking in the HSC.

Like the B-alone results, the B.E results indicated weaker blocking in the HSC than in the ESC. That is, the

**Table 2**  
Response Choice Percentages in the Test Phase

Stimulus Type	Equal-Salience Group				High-Salience Group			
	Cue Type	Response Type			Cue Type	Response Type		
		R1	R2	R3		R1	R2	R3
Late phase maintenance	A.B	<b>96</b>	1	4	<b>A.B</b>	<b>92</b>	2	6
	C	1	<b>95</b>	4	C	2	<b>95</b>	3
	D.E	1	4	<b>95</b>	D.E	2	3	<b>96</b>
Early phase retention	A	<b>96</b>	2	2	A	<b>66</b>	30	5
	F	6	7	<b>87</b>	F	11	27	<b>63</b>
Important novel test items	B	22	17	61	<b>B</b>	<b>70</b>	6	23
	D	15	4	<b>82</b>	D	17	45	<b>38</b>
	E	7	6	<b>87</b>	<b>E</b>	5	5	<b>91</b>
	B.D	24	–	76	<b>B.D</b>	70	–	30
	B.E	20	–	80	<b>B.E</b>	34	–	66

Note—Boldface cues were salient. The bold italicized percentages correspond to the normative response. Percentages within rows do not all sum to 100%, owing to rounding error. Dashes indicate disallowed responses.

difference of responding R3 versus R1 was 78% versus 22% in the ESC when a B.D or B.E item was presented, which was greater than the difference of 66% versus 34% in the HSC when a B.E item was presented. Unfortunately the between-group difference did not reach statistical significance [ $t(57) = 1.35, p = .182$ , two-tailed]. This lack of significance is somewhat curious, given that B-alone was learned much better in the HSC than in the ESC. The muted difference cannot be explained by differential responding to E, because learning of E alone was very similar across conditions. In the ESC, E alone elicited an R3 response 87% of the time, and in the HSC, E alone elicited an R3 response 91% of the time. The fact that the difference between groups was not significant may have been due merely to lack of statistical power, because this difference had a moderate effect size ( $d = .35$ ) and hence our power was only about 0.26.

### Does the Blocking Cue A Lose Strength in the HSC?

**Primary evidence.** The A-alone test item was the primary assay of cue A's strength. In the ESC, the A cue commanded 96% of R1 responses, whereas in the HSC, participants responded R1 only 66% of the time. The mean percentage difference of 31% was significant [ $t(57) = 3.31, p = .002, d = .86$ ] (95% CI = 12%, 49%). From this comparison, it appears that the A cue did indeed lose strength in the HSC.

**Ancillary evidence.** Further evidence for the reduction of control by A comes from several cue combinations tested. The A cue was tested in combination with C, D, and N (a novel cue). The A.C item (see Appendix B, Table B1) showed a reduction in R1 responding in the HSC. The A.C item elicited 46% R1 responding in the ESC and only 31% R1 responding in the HSC, but this trend was not significant [ $t(57) = 1.21, p = .23, d = .32$ ]. The A.N item elicited 87% R1 responding in the ESC, and that percentage dropped to 78% in the HSC, but this trend was also nonsignificant [ $t(57) = 1.25, p = .22, d = .33$ ]. These trends may have been nonsignificant merely because of a lack of statistical power, because our power for both the A.C and A.N comparisons individually was only about .23. Support for this assertion was provided by the fact that these trends became significant when A.C and A.N were pooled together with the A.D cue combination. When the R1 responses for A.C, A.D, and A.N were pooled for each participant, we found a significant difference in R1 responding between the two conditions ( $M_D = 12%$ ) [ $t(57) = 2.09, p = .04, d = .29$ ] (95% CI = 1%, 24%).

The A.D comparison was informative on its own. When shown an A.D pair, participants in the ESC responded R1 94% of the time. In the HSC, participants responded R1 77% of the time. There was a significant difference in R1 responses on the A.D item ( $M_D = 18%$ ) [ $t(57) = 2.28, p = .027, d = .59$ ] (95% CI = 2%, 34%). The A cue controlled responding less in the HSC, and this could not have been because D was learned better in this condition. The D cue was learned only very weakly in the HSC, as

will be discussed in a subsequent section. Therefore, this comparison provides more evidence that the A cue was weakened in the HSC.

Additional evidence for a strength reduction of the A cue was attained by an analysis of a control stimulus for the A cue—namely, the F cue. The F-alone cue appeared in initial training, just like A-alone, but then did not appear again until the test phase. The strength of the F cue was therefore a lower bound for the strength of the A cue. In the ESC, the F cue test item elicited an R3 response 87% of the time, and in the HSC, the percentage dropped significantly to 63% ( $M_D = 25%$ ) [ $t(57) = 2.62, p = .011, d = .69$ ] (95% CI = 6%, 43%). The F cue did not appear in late training and therefore lost some of its ability to control responding when tested at the end of the experiment, but more in the HSC than in the ESC. Thus there appeared to be a retention loss or interference effect for the F control cue that was more detrimental in the HSC. The drop in strength of F appeared comparable to the drop in A, because the difference between conditions on the A.F trials was not significant [ $t(57) = 0.48, p = .962$ ] (see Appendix B). Overall, the similarity between the treatment of the A cues and of the F cues provides additional evidence that the A cue lost strength in the HSC. If the A cue had not lost strength in the HSC, then when participants were presented with an A.F stimulus, R1 responding would be higher in the HSC than in the ESC, but it was not. This result does raise some further questions, however, about why the F cue was treated differently between conditions. This issue will be discussed later in the article.

### Is D Overshadowed?

**Primary evidence.** During late training, the D cue was always paired with the E cue. In the HSC, the salient E cue should have strongly overshadowed the faint D cue. However, in the ESC, training of the D.E cue combination should have led to weaker mutual overshadowing of D and E. The data do indicate that D was overshadowed more, and not learned as well, in the HSC. In the ESC, the D cue by itself elicited an R3 response 82% of the time, whereas in the HSC, participants only responded R3 38% of the time (essentially at chance levels). The mean difference of 44% was significant [ $t(57) = 4.22, p < .001, d = 1.10$ ] (95% CI = 23%, 65%). This gives evidence that the D cue was not learned in the HSC, or at least not learned nearly as well as in the ESC.

**Ancillary evidence.** The data also indicate that the D cue had only weak control over responding in either group. When paired with the control C cue (C.D cue; see Appendix B, Table B1), participants responded R3 only 9.3% of the time in the ESC and 6% of the time in the HSC. This provides some evidence that D was overshadowed in both conditions, because the D cue was unable to control responding when paired with a strong comparison cue.

The D cue also did poorly in controlling responding when paired with the A cue. When shown an A.D pair (see Appendix B, Table B1), participants in the ESC responded R3 only 6% of the time. In the HSC, participants

responded R3 23% of the time. The significance of this difference and its importance were discussed in the previous section as evidence for the reduction in strength of the A cue.

Learning of D in the HSC was so greatly reduced that the distribution of responses to a D cue was not differentiable from the responding to a novel cue N (see Appendix B, Table B1). The N cue appeared only in the final testing phase of the experiment, so the participants should have had no learned association between it and any outcome. In the HSC, the pattern of responses for the D cue was not significantly different from that for the N cue [ $\chi^2(2, N = 128) = 0.555, p = .758$ ]. Thus, the D cue appears to have been largely overshadowed by the salient E cue that it was paired with during training.

### What Else Does the N Cue Tell Us?

Because the N cue was not associated with any particular outcome, it would be expected that the pattern of responses to N would be near uniform for all response options, but this was not the case. The pattern of responses was significantly different from uniform in both the ESC [ $\chi^2(2, N = 54) = 30.333, p < .001$ ] and in the HSC [ $\chi^2(2, N = 64) = 9.500, p = .009$ ].

As can be seen from Table B1 (Appendix B), the pattern of responses to the N cue was different between the ESC and the HSC. This difference proved to be significant [ $\chi^2(2, N = 118) = 17.186, p < .001$ ]. In the ESC, there was a tendency toward the response consistent with a D, E, or F cue (R3), whereas in the HSC, there was a tendency toward the response consistent with a C cue (R2). There was apparently a bias toward certain response items, which differed between conditions. In the discussion, this bias will be explored further.

### Discussion of Results

These results for humans replicate and extend the findings for animals of Hall et al. (1977). We found reduced blocking and a reduction in control by the blocking cue in the HSC. Our results also extended the results of Hall et al. because our design allowed for many more ways to assess blocking and the relative strength of the blocking cue. The animals in the Hall et al. experiments saw at most two cues and testing consisted only of the two cues separately and combined. Participants in our experiment saw a total of seven cues and were tested on the seven cues as well as fourteen different cue combinations. These tests allowed us to assess the reduction of the A cue in a variety of ways, including the A-alone, A.C, A.N, A.D, and A.F comparisons. The Hall et al. experiments only tested the reduction in the previously learned cue by testing it alone after compound training.

Our design also allowed us to have within-subject controls as well as between-subject controls. The experimental design of Hall et al. (1977) only allowed them to have between-subject controls. Thus we were able to assess blocking by comparing the blocked cue (B) with the control cue (E) within both conditions as well as comparing the blocked cue across conditions. Our data simulta-

neously support the findings of Hall et al. and provide a richer data set, which will prove useful for constraining models.

### MODELING

Given that the present findings matched our predictions, the next step is to see whether the data can be accounted for by a formal model. This is necessary in order to give the empirical results a more rigorous grounding in attentional learning theory. A model with attentional mechanisms is proposed, because extant models with only associative learning mechanisms, such as the Rescorla–Wagner (1972) model, have been unable to account for data from similar experimental designs, whereas models with attentional mechanisms have fared better (Kruschke, 2001b; Le Pelley, 2004).

As previously mentioned, the Rescorla–Wagner (1972) model cannot show a reduction in associative strength (a cue cannot be unlearned) unless a present cue overpredicts an outcome. Simple modifications can be made to the Rescorla–Wagner model such that learned weights can be reduced. This has been done in multiple ways. Markman (1989) showed that coding absent cues as negative values in a connectionist network using the Rescorla–Wagner learning rule can decrease their weights. The reasoning behind this coding was that if a cue is expected but does not occur, then its absence should be explicitly encoded. Van Hamme and Wasserman (1994) used a similar modification of the Rescorla–Wagner model to show a reduction in the associative weight for absent cues. Their modification was to use a negative learning rate for absent cues (cf. Wasserman & Berglan, 1998).

These modifications would not be of much use for the present results, because the reduced strength of the blocking cue A occurred in spite of the fact that it was still present during every training trial in which the outcome is response R1. Although it is possible that participants in our experiment treated cue A as an absent cue in late training of the HSC—that is, it was completely overshadowed by the salient blocked cue B—for modeling purposes, it would not make sense to encode cue A as absent on these trials. The reduction in the strength of the blocking cue A in the HSC may have been due to its being ignored, but this would need to be accounted for by an attentional mechanism in the model.

The Rescorla–Wagner (1972) model was intended only for present/absent stimuli. For the present experiment, all cue colors were present on every trial and should be included in the stimulus representation. It seems reasonable to encode the stimulus representation as partial cue activations corresponding to color densities in the stimulus display. Therefore, all input activations are positive but have different magnitudes. Alternate encoding strategies, with negative components, are disallowed, because no cue color is ever absent.

Our stimulus representation bears some resemblance to a recent model by Ghirlanda (2005). In that model, resting activity in the absence of stimulation (or in the presence of

low-intensity background stimuli) is reflected by a small positive activation across an array of input elements at all times, even during intertrial intervals. Whereas this is similar to our “absent” cues’ having a small background presence, in our model a truly absent cue has zero activation. Also, in the model proposed by Ghirlanda, inputs are represented by graded activation on a continuum, whereas our input units represent nominal cues.

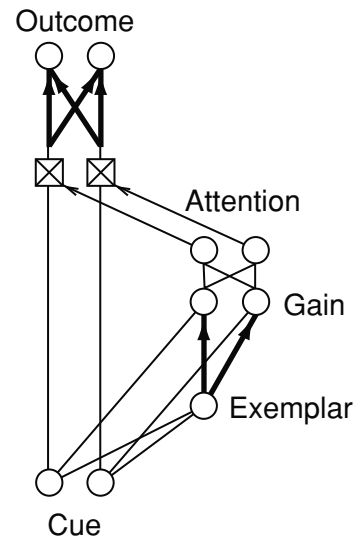
Because it is unlikely that associative learning alone will account for our data, our goal in modeling is to uncover what additional mechanisms are possibly at work. Two mechanisms that could explain a reduction in cue strength are attention shifting and forgetting. An attention shifting mechanism would shift attention away from a cue in such a way that responding to the cue is reduced. A forgetting mechanism would actually reduce the strength of the learned association by having it decay over time.

The experimental data were evaluated with models that use attention shifting and associative weight decay. A modified form of the EXIT model (Kruschke, 2001a, 2001b) was used as the attention shifting model, because the EXIT model is one formal implementation of attentional learning theory. Weight decay was added to the model to create what we call here the full model. The full model has both attention shifting and decay. To assess the effects of both attention and forgetting, nested model comparisons were done with four models. The four models are referred to as the *full* model (both attention and weight decay), the *attention* model (without decay), the *decay* model (without attention), and the *restricted* model (without attention or decay). Because both decay and attention are parameterized in the models, the three nested models can be obtained by setting certain parameters in the full model to zero. A description of the full model follows. Subsequently, the other models will be described as special instances of the full model.

### The Full Model

The full model is a modified version of the EXIT model, which has been thoroughly detailed elsewhere (Kruschke, 2001a, 2001b). Here we will describe the model only briefly and elaborate on the modifications that were made to accommodate this particular experiment. The EXIT model has previously been compared with other classic models of attention, such as Mackintosh’s (1975) model of attention in animal learning (Kruschke, 2001b). EXIT has been shown to fit results from various experiments and to make novel predictions regarding relative learning rates after blocking, and correlations across individuals in the magnitude of blocking and other effects (e.g., Kruschke, 2001b, 2005; Kruschke, Kappenman, & Hetrick, 2005).

The EXIT model (shown in Figure 2) is a connectionist architecture that represents each cue as an input node. In previous versions of the EXIT model, cues were either present or absent. If the cue was absent, its corresponding input node had a zero activation, and if the cue was present, the activation was positive. In the present experiment, all the cues are present at all times; all the cue colors were



**Figure 2.** The general architecture of the EXIT model. This diagram illustrates a case with two cues, one exemplar, and two outcomes. Thick arrows denote connections with learned weights; thin lines denote fixed-weight connections. Cue node activation is described in Equation 1 in the text; exemplar node activation in Equation 4; gain node activation in Equation 5; attention activation in Equation 6; and outcome activation in Equation 2. Learning of outcome weights is described in Equation 10; learning of gain weights, in Equation 12.

in each stimulus display. The cues were distinguished only by dot densities. Therefore, we set the activations of the input nodes all positive and differing only in magnitude. The input activation for the  $i$ th node is set as

$$a_i^{\text{in}} = \begin{cases} b & \text{if background density} \\ f & \text{if faint density} \\ s & \text{if salient density} \end{cases} \quad (1)$$

with  $b < f < s$ . For all the model fits presented later in this article, the faint density was arbitrarily set to one ( $f = 1.0$ ). The background and salient densities were free parameters that were estimated separately for each of the model fits.

On any particular trial, the stimulus is represented by a distribution of activations across the input or cue nodes. Cue nodes correspond to colors, and their activation magnitudes correspond to dot densities. When a stimulus is presented, the input nodes are activated, and the activation spreads to the output nodes via weighted connections. The output nodes correspond to response categories. Input node  $i$  is connected to output node  $k$  by a link with an associative weight denoted  $w_{ki}$ . The input-to-category associative weights are initialized at zero, but change with learning, as will be described later.

Each input node activation is multiplied by a nonnegative attention strength. By default, all the cue nodes get some attention proportional to their activation. The attentional strengths modulate the influence of the input activations, in such a way that the output activation is de-



terminated by a weighted sum across the attentionally gated input activations. Formally, the activation of output node  $k$  is determined as follows:

$$a_k^{\text{out}} = \sum_i w_{ki} \alpha_i a_i^{\text{in}}, \quad (2)$$

where  $\alpha_i$  is the attention strength on input node  $i$ . The source of these attention values is described below.

The output activations are mapped to choice probabilities in such a way that highly activated output nodes are given high response probabilities. Specifically, the probability of choosing response  $r$  is given by

$$p(r) = \frac{\exp(\phi a_r^{\text{out}})}{\sum_k \exp(\phi a_k^{\text{out}})}, \quad (3)$$

where  $\phi$  is a *decisiveness* parameter. The equation is simply a version of Luce's (1959) decision rule. When the parameter  $\phi$  is large, the network is highly decisive; small activation advantages are translated into large choice preferences. When  $\phi$  is small, the network is indecisive; large activation differences are translated into less extreme choice probabilities. This parameter influences the magnitude of the choice probabilities that the network generates, but has no influence on learning, attention shifting, or any of the internal workings of the network.

To account for possible response choice biases, an additional input node was added to the model similar to the context node in the original EXIT model. This bias node is fully activated on every trial ( $a_b^{\text{in}} = 1.0$ ), and attention on the bias cue is constant ( $\alpha_b = 1.0$ ). The bias-to-category associative weights are adjusted separately from the input-to-category weights in the manner described below. The intent is to have the bias-to-category weights learn response biases.

The attentional system propagates activation as does the original EXIT.<sup>2</sup> The attentional system allows the network to learn what attentional shift to make across the cues and also to reproduce this shifted pattern of attention in the future. The learning of attentional distributions is accomplished by connections between the cues and attentional gates or gain nodes. The propagation of the activation to the attention gates is mediated by exemplar nodes that encode cue configurations. The exemplar nodes are activated to the extent that the current stimulus is similar to the stimulus that is represented by the exemplar. The exemplar activation values are given by

$$a_x^{\text{ex}} = \exp\left(-c \sum_i |\psi_{xi} - a_i^{\text{in}}|\right), \quad (4)$$

where the superscript "ex" indicates that this is an exemplar node, and where  $\psi_{xi}$  is the magnitude of cue  $i$  in exemplar  $x$ .<sup>3</sup> The width of each exemplar node's receptive field is determined by the *specificity* parameter  $c$ . The larger the specificity, the less the learned attentional distribution generalizes from one exemplar to another. The exemplar-mediated mapping from cues to attention gates

allows the network to learn exemplar-specific distributions of attention.

The attentional gates or gain nodes are activated by a combination of one-to-one connections from the input nodes and weighted connections from the exemplar nodes. The activation of the gain node  $i$  is given by

$$g_i = a_i^{\text{in}} \exp\left(\sum_x w_{ix} a_x^{\text{ex}}\right), \quad (5)$$

where  $w_{ix}$  is the associative weight from exemplar node  $x$  to gain node  $i$ . The associative weights are initialized at zero and adjusted in the manner described below. Notice that a cue gets attentional gain proportional to its own activation and modulated by learned associations from exemplars.

The attention on the cues is hypothesized to have a limited capacity, and therefore the cues compete for attention. The capacity constraint is formalized by making the attention to each cue be the normalized gain of that cue:

$$\alpha_i = \frac{g_i}{\left(\sum_j g_j^P\right)^{1/P}}. \quad (6)$$

The normalization uses a Minkowski metric, and this introduces a power parameter. This power parameter ( $P$ ) is the *attentional capacity* and determines the total attentional weighting that can be allocated across cues. When  $P$  is large, there is more attentional multiplication overall, and overall learning is faster. When  $P$  is small, close to zero, there is severe competition between cues.

To summarize, EXIT generates a response as follows. First, a stimulus is presented, activating the corresponding cue nodes proportional to their saliences. Attention is distributed across the cues, with attentional allocation determined by a combination of cue salience and previously learned attentional shifts. The attentionally gated cue activations are spread to the output nodes.

After a response, corrective feedback is supplied, and the network determines the discrepancy between the correct response and the output activations that it generated. The goal of the network is to reduce this error, where error is measured as the sum square deviation between the teacher values  $t_k$  and the generated activation values across the output nodes,

$$E = .5 \sum_k (t_k - a_k^{\text{out}})^2. \quad (7)$$

Because the responses are categorical, not metric, teacher values were coded as the "humble teachers" of Kruschke (1992), whereby

$$t_k = \begin{cases} \max(+1, a_k^{\text{out}}) & \text{if stimulus is in category } k \\ \min(0, a_k^{\text{out}}) & \text{if stimulus is not in category } k \end{cases} \quad (8)$$

These teacher values are defined so that output activations greater than +1 are not counted as errors if the stimulus should be associated with response choice  $k$ , and output

activations less than zero are not counted as errors if the correct response is not  $k$ .

The model uses gradient descent on error to drive all aspects of learning. Gradient descent on error computes the change that would most quickly reduce the discrepancy between the correct response and the predicted outcome. Error reduction in the model goes through two stages. First, attention is shifted away from cues that cause error, and toward cues that reduce error. Second, the network adjusts its associative weights by simple error reduction learning using an attention-weighted version of the delta rule. Three sets of weights are adjusted. The first set connects the (attentionally gated) cues to the outputs. The second set connects the bias node to the outputs. The third connects the exemplar nodes to the attention nodes.

**Attention shifting.** The first response to corrective feedback is the rapid shift of attention to reduce error. This is accomplished with gradient descent on error with respect to the gains:

$$\Delta g_i = \frac{\lambda_g \sum_k (t_k - a_k^{\text{out}}) (w_{ki} a_i^{\text{in}} - \alpha_i^{P-1} a_k^{\text{out}})}{\left( \sum_j g_j^P \right)^{1/P}}, \quad (9)$$

where  $\lambda_g$  is a rate parameter called the *shift rate for attention*, which simply determines how large a shift is made in response to error. Although one large shift is desired, the shift is done in 10 relatively small steps on each trial, because attention is a highly nonlinear function of gain. After each small step, activation is repropagated to the category nodes to generate a new error and attention is adjusted. The result of these 10 small steps constitutes the single large shift.

**Associative weight learning with decay.** After the attention is shifted, the three sets of associative weights are adjusted, also by gradient descent on error. The first set of weights connecting the cues to output is adjusted by the following learning rule:

$$\Delta w_{ki} = \lambda_w (t_k - a_k^{\text{out}}) \alpha_i a_i^{\text{in}} - \delta_w w_{ki}, \quad (10)$$

where  $\lambda_w$  is the *learning rate for output weights* and  $\delta_w$  is the *output weight decay rate*.

Weight decay, which was not in the original EXIT, simply allows associations between cues and outcomes to be “forgotten.” The introduction of weight decay causes connection strengths to continually decay back toward zero by an amount proportional to their magnitude. One consequence of this is that the network suppresses irrelevant components of the weight vector where associations are no longer being strengthened (Hinton, 1989; Krogh & Hertz, 1992). So decay represents a mechanism through which stable relationships are given preferred status over more transient ones (McLaren & Mackintosh, 2000). For the purposes of the present experiment, it is important that weight decay serves to undo the effects of training, unless that training is reinforced by further presentations of the stimuli. Because decay is cumulative, it should serve to

create a recency effect in which early learned associations will show the greatest decay if they are not reinforced.

The second set of associative weights connecting the bias node to the output nodes are adjusted in a similar fashion as the output weights:

$$\Delta w_{kb} = \lambda_b (t_k - a_k^{\text{out}}), \quad (11)$$

where  $\lambda_b$  is a constant of proportionality called the *bias learning rate*. Once again, the bias node acts as another input node that is always active ( $a_b^{\text{in}} = 1.0$ ) and maintains a constant level of attention ( $\alpha_b = 1.0$ ). The only important distinctions between Equations 10 and 11 are that the bias node has its own learning rate and the weights do not decay. Bias weight decay was not implemented, because the bias is intended to be cumulative. The decay rate on the bias can be thought of as having been arbitrarily fixed at zero.

The third set of weights connecting the exemplar nodes to the gain nodes is also adjusted by gradient descent on error, with error defined here as the sum of squared differences between the shifted gain value and the initial, preshift value. Gradient descent yields

$$\Delta w_{ix}^g = \lambda_x (g_i^{\text{shift}} - g_i^{\text{init}}) g_i^{\text{init}} a_x^{\text{ex}}, \quad (12)$$

where  $\lambda_x$  is the learning rate for the associate weights from exemplar to gain nodes and is called the *attentional weight learning rate*. This allows the network to learn the shifted attentional distributions, because the postshift values act as the teachers for the gain node activations.

In summary, learning proceeds as follows. Feedback is encoded as teacher values at the outcome nodes. The error between the teacher values and the predicted outcomes is computed. Attention is shifted to reduce the error. Then the weights to output nodes and to gain nodes are adjusted to reduce error. As a consequence of learning, the weights better reproduce the desired attentional allocation and outcome on subsequent trials.

The full model has a total of 10 free parameters. This high-dimensional parameter space has many local minima, and therefore the best-fitting parameters (for all models) provided in Appendix C, Table C1, may not be global minima; but they are the result of extensive searches made using a wide range of initial parameter values.

### The Attention Model

The attention model, or no-decay model, is a special case of the full model. This model is obtained by setting the output weight decay rate parameter ( $\delta_w$ ) to zero. This model therefore has nine free parameters. The attention model is the most similar to the original EXIT model (Kruschke, 2001a, 2001b). This model uses attention shifting, but excludes the mechanism for forgetting.

### The Decay Model

The decay model, or no-attention model, is obtained by removing the attention component of the full model. This is done by setting the attentional shift rate ( $\lambda_g$ ) and the gain

weight learning rate ( $\lambda_x$ ) to zero. Exemplar specificity ( $c$ ) becomes extraneous once attention is removed, and therefore this parameter is also set to zero.<sup>4</sup> The decay model has a total of seven free parameters.

### The Restricted Model

At the bottom of the nested model hierarchy is the restricted model. This model is a special case of the full model in which both attention and decay have been removed. The restricted model is obtained by setting the attentional parameters of the full model to zero, as was done in the decay model, and by setting the decay rate parameter to zero, as was done in the attention model. The restricted model is analogous to the Rescorla–Wagner (1972) model, with two supplements. The restricted model uses the response-mapping choice rule and the capacity-limiting mechanism that were not present in the original Rescorla–Wagner model.<sup>5</sup>

### Fit of the Models

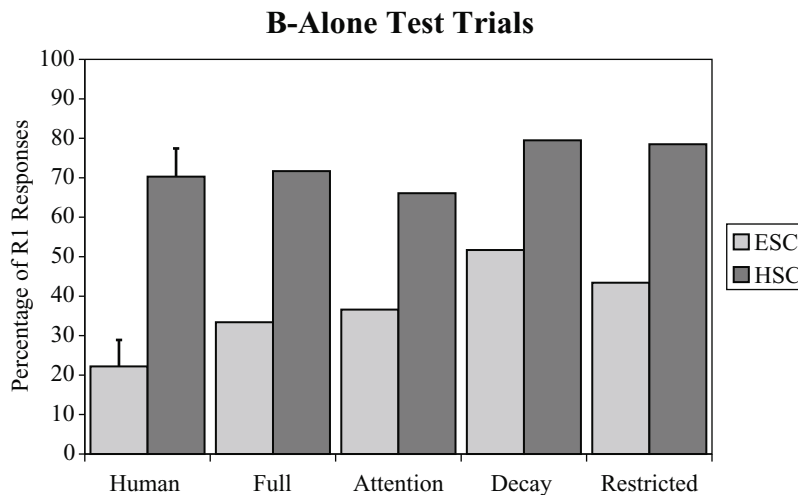
The models were trained on 12 blocks of early training and 15 blocks of late training. The number of blocks in early training was chosen to approximate the mean blocks to criterion observed in human participants who learned the task (11.75 blocks). The model then made predictions on 2 testing blocks. Stable model predictions were achieved by running 50 simulated subjects, each with a different random ordering of training and test trials within blocks. The models were fit using the generalized likelihood ( $G^2$ ) to the mean percentages of both testing blocks and the last 2 blocks of the early training phase (i.e., all the data shown in Table B1). Thus, the models were fit to 120 choice percentages, containing 72 independent data values. The last 2 blocks of the early training phase were included so that the models would have to show accuracy on the training items comparable to human performance. For each model,

a single set of parameter values was used for all simulated subjects in both equal- and high-salience conditions. The best-fitting parameters and model predictions for all four models are shown in Appendix C, Table C1.

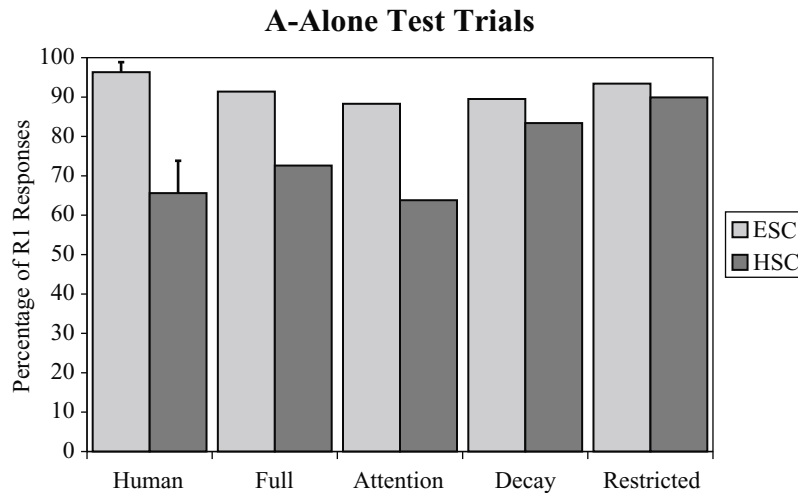
**Model comparison.** Of the four models, the best fit to the human data was, of course, provided by the full model, with  $G^2 = 115.7$ . This was followed by the attention model, with  $G^2 = 193.6$ ; the decay model, with  $G^2 = 259.6$ ; and the restricted model, with  $G^2 = 379.7$ . The largest expected difference for nested model comparison would be the difference between the full and the restricted model, because the restricted model has four fewer parameters than the full model. Given that the critical  $\chi^2$  ( $\alpha = .05$ ,  $df = 4$ ) was only 9.49, it is easy to see that the fits of the models are all significantly different. This means that both attention and decay mechanisms improve the fit of the model significantly. Therefore, we must take a closer look at the model predictions to assess the effects and importance of attention versus decay.

**Is cue B blocked?** In the ESC, all the models show blocking of the B cue and they all show a reduction in blocking in the HSC relative to the ESC. As can be seen in Figure 3, the more complex models (the full and attention models) appear more similar to the human data for the B-alone item, but all the models show the correct trend of reduced blocking in the HSC. Although the models make somewhat different predictions on blocking test trials, overall they cannot be qualitatively differentiated on the basis of their abilities to block the B cue. The more important items for model comparison involve the A (i.e., blocking) cue.

**Does the blocking cue A lose strength in the HSC?** The models do show important differences in their abilities to show the reduction in responding to the blocking cue (A) in the HSC. Figure 4 shows that the models with attention mechanisms (the full and attention models) are



**Figure 3.** Percentage of R1 responses to the blocked cue B. Error bars indicate one standard error. Human, human data. Full, full model. Attention, attention model. Decay, decay model. Restricted, restricted model. ESC, equal-salience condition. HSC, high-salience condition.



**Figure 4.** Percentage of R1 responses to the blocking cue A. Error bars indicate one standard error. Human, human data. Full, full model. Attention, attention model. Decay, decay model. Restricted, restricted model. ESC, equal-salience condition. HSC, high-salience condition.

able to show a reduction in the strength of responding to cue A similar to that of the human data. The decay model shows some weak reduction in responding to the A cue, but the decay model does not match the magnitude of reduction seen in human data.

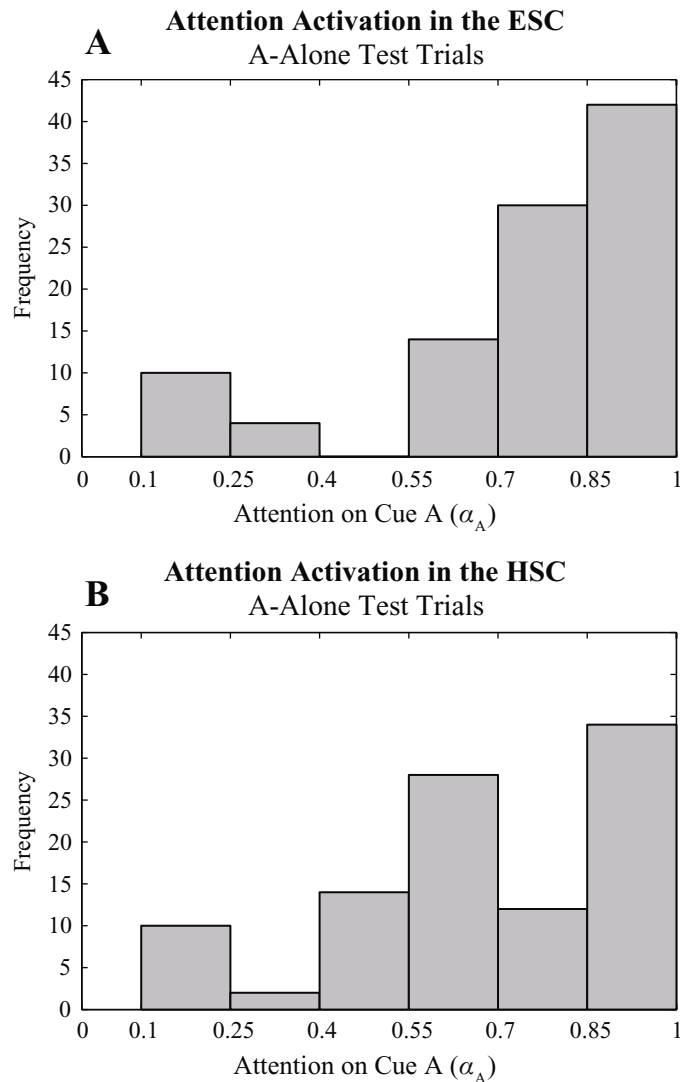
In the attention model, the reduction of correct responding to the A cue is not caused by a reduction in associative strength from the A cue. After the initial training phase, which is identical in the ESC and the HSC, the average associative weight between cue A and response R1 is 3.09 ( $SD = 1.06$  across 50 simulated subjects). After the second training phase in the HSC, this average weight *increases* slightly to 3.10 ( $SD = 1.06$ ). The weight cannot decrease in the second phase, because there is no weight decay and there are no trials with A present while responses are overpredicted.

In the attention model, the reduction of correct responding to the A cue in the HSC is achieved via shifting of attention. During the later training phase in the HSC, attention is shifted away from the A cue, for many of the simulated subjects. Figure 5 shows the allocation of attention to cue A on A-alone test trials. (The attention values are the  $\alpha$  values computed in Equation 6.) Each simulated subject received two A-alone trials during testing. The histograms show the number of trials across all simulated subjects in which the attention value on cue A fell within the depicted range. The upper panel of Figure 5 shows results for the ESC, where it can be seen that most simulated subjects maintained high attention on cue A. The majority of trials show attention values close to 1.0 (full attention on cue A). Results for the HSC are shown in the lower panel of Figure 5, where it can be seen that attention values on cue A are more distributed. In the HSC, the number of simulated subjects that maintain full attention on cue A is smaller than in the ESC, and thus the distribution is shifted toward lower values of attention on cue A.

Why is attention shifted away from A on A-alone test items? The shift is caused by generalization from what was learned on A.B training items. On A.B training items in the HSC, the salient cue B grabs attention away from cue A (for many simulated subjects). This shift of attention away from A is learned by the exemplar-to-gain associations. Subsequently, when A appears by itself in the test phase, the A.B exemplar is partially activated and attention is shifted away from A. The reduced attention to A then causes a weaker activation of R1—that is, a reduced R1 response activation.

**The insufficiency of forgetting.** The fact that the decay model does not show a robust reduction in responding to the A cue in the HSC lends support to the claim that more is taking place than mere forgetting. Further evidence for the insufficiency of forgetting is given by the test trials that used the F cue. The F cue is a control cue learned in the initial learning phase but then not seen again until the testing phase. As can be seen in Figure 6, accuracy on the F cue is lower in the HSC than in the ESC. The decay model cannot account for a reduction in correct responding to F in the HSC, because decay treats F the same in both the ESC and the HSC: The weights from F to the response nodes decay constantly throughout late training while F is not present. Thus, the decay model shows a reduction in correct responding to F when compared with the restricted model, but is unable to capture the differential effect of salience. (In Figure 6, correct responding to F is slightly higher in the ESC than in the HSC because of a generalized response bias toward R3 learned in the ESC.)

One approach to salvaging an explanation based on forgetting is to hypothesize that forgetting was accelerated in the HSC by interference from the salient cues. This acceleration would probably have to be *cue specific*, because the decay rate on faint cues would need to be greater than



**Figure 5. Histograms of attention allocation in the attention-only model for the A-alone test cue. The upper panel shows simulations of the equal-salience condition (ESC), and the lower panel shows simulations of the high-salience condition (HSC). The count indicates single trials.**

the decay rate on salient cues. Such cue-specific decay rates are tantamount to a form of selective attention, and so it is unclear that such an approach would be meaningfully different from an attentional account.

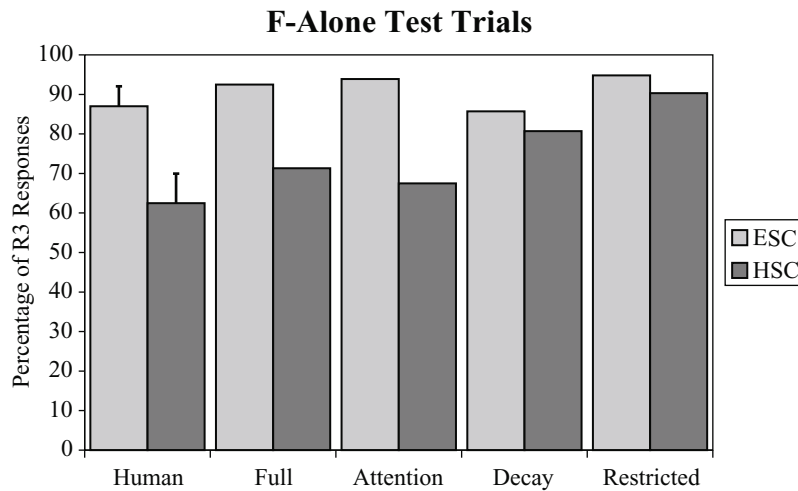
**The importance of attention.** We have seen that the attention model does a better job than the decay model in showing the reduction in correct responding to the A cue in the HSC. Attention shifting is cue specific, unlike weight decay, and shifting therefore allows for cues to differentially control responding between the ESC and the HSC. The shift of attention away from A that occurs on A.B trials when B is salient explains both the decreased blocking of B and the reduction in strength of cue A.

The attention model is also able to account for the reduction in responding to cue F, which presents problems for the decay model (Figure 6). Attention is distributed

differently if salient cues are present than when they are not. In the HSC, the salient B and E cues draw attention away from other cues. This shifting of attention is generalized in the test trials. In particular, when F is presented alone, attention is partially diverted, and correct responding is reduced.

## SUMMARY AND DISCUSSION

Previous research on animals has suggested that more salient cues are harder to block, and that the strength of a blocking cue will be reduced when it is paired with a more salient cue (Hall et al., 1977). In the present study, these findings were extended to humans. It was found that a normally blocked cue was much harder to block when it was salient. It was also found that a previously learned



**Figure 6. Percentage of R3 responses to cue F. Error bars indicate one standard error. Human, human data. Full, full model. Attention, attention model. Decay, decay model. Restricted, restricted model. ESC, equal-salience condition. HSC, high-salience condition.**

blocking cue did lose some of its ability to control responding when paired with a highly salient to-be-blocked cue. Our experiment's design included several tests that extended the results of Hall et al. and provided a rich data set for modeling.

We contend that the results can be explained in part by attentional mechanisms. The reduction in blocking of B and the reduction in correct responding to A in the HSC indicates that the highly salient stimuli drew attention away from previously learned but less salient stimuli. The loss of control by the blocking cue can be explained by a shift of attention away from it. The loss does not necessitate a reduction in the strength of the learned association.

Only the models with attention mechanisms showed an appropriate reduction in the strength of the blocking cue in the HSC relative to the ESC. The decay-only model failed to show appropriate losses for the blocking cue and the first-phase control cue.

A modified form of the EXIT model was a viable candidate to account for the human data, but ultimately even the full model left some aspects of the data unaccounted for. This remaining discrepancy may have been due to the large number of data points and hence noise inherent in the data, but it may also indicate that the model is incomplete. A comparison of the human data and the model predictions suggests that the data require mechanisms that the model does not have.

Consider responses to the novel cue (N) in the test phase. This cue should not have been associated with any particular outcome, but there was a significant tendency for participants to respond R3 in the ESC and R2 in the HSC. In the ESC, the overall bias toward response R3 can be explained as a learned heuristic, as follows. Although there is a strong association from A to R1 and from C to R2, several cues indicate R3—namely, cues D, E, and F. Response R3 therefore seems to become an “everything else” bin. This

would explain why response R3 is more likely to be chosen when a novel cue is presented (because N is neither A nor C), and why there is a strong R3 response when the blocked cue B is presented (because B is neither A nor C, and an association between B and R1 was never learned). The model has no way to learn an “everything else” heuristic.

In the HSC, the overall bias toward response R2 can be explained in terms of participants' creating a categorical distinction between faint and salient cues. In the second training phase, participants learn that the salient cue B is associated with R1 and that the salient cue E is associated with R3. The only response that remains strongly associated with a faint cue is R2, which should be associated only with cue C. From this association, learners might induce a heuristic: If the cue is faint (like C), respond R2. This explains why the faint novel cue is more likely to elicit an R2 response and why the same is true of the faint D cue (which is overshadowed by the salient E cue during training and treated much like a novel cue). Bolstering this suggestion is the fact that invalid response percentages (Table B2) in the HSC were the highest when the cues were both faint and the disallowed response option was R2. The model has no way to induce a heuristic based on a categorical faint–salient distinction.

Even though the between-condition differences seen with the F cue can be explained by attentional redistribution (as described previously in this article), it is also possible that this effect was partially a product of these heuristics. The “everything else” strategy would boost correct responses to F in the ESC, and a categorical distinction between faint and salient cues would reduce correct responding to the faint F cue in the HSC.

These alternative strategies do not explain all the discrepancies between the model and the data. Learners probably invented yet other strategies that are not implemented in the present model.

In the future, it may be fruitful to proceed in any of three directions with this research. One would be to create a new model that includes additional mechanisms such as those outlined above. A second direction would be to design new empirical methods that prohibit nonassociative strategies, or at least limit the alternative strategies that participants could bring to bear on the experimental task. A third direction would be to extend training to subsequent phases that further assess learned attention. An experimental design analogous to those used by Kruschke and Blair (2000) or Kruschke (2005) could be adapted to reveal the difficulty with which subsequent associations to A are learned. If attention to A has been reduced, then subsequent learning about it should be retarded.

### REFERENCES

- ARCEDIANO, F., ESCOBAR, M., & MILLER, R. R. (2004). Is stimulus competition an acquisition deficit or a performance deficit? *Psychonomic Bulletin & Review*, **11**, 1105-1110.
- EDGE, S. E., CASTELLAN, N. J., JR., ROE, R. M., BARNES, J. M., NG, P. C., BRIGHT, R. D., ET AL. (1996). Irrelevant information in probabilistic categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1463-1481.
- FELDMAN, J. M. (1975). Blocking as a function of added cue intensity. *Animal Learning & Behavior*, **3**, 98-102.
- GHIRLANDA, S. (2005). Retrospective reevaluation as simple associative learning. *Journal of Experimental Psychology: Animal Behavior Processes*, **31**, 107-111.
- HALL, G., MACKINTOSH, N. J., GOODALL, G., & DAL MARTELLO, M. (1977). Loss of control by a less valid or by a less salient stimulus compounded with a better predictor of reinforcement. *Learning & Motivation*, **8**, 145-158.
- HINTON, G. E. (1989). Connectionist learning procedures. *Artificial Intelligence*, **40**, 185-234.
- KAMIN, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami Symposium on the Prediction of Behavior: Aversive stimulation* (pp. 9-33). Coral Gables, FL: University of Miami Press.
- KAMIN, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279-296). New York: Appleton-Century-Crofts.
- KLEIN, S. B., WESTON, D., MCGEE-DAVIS, T., & COHEN, L. (1984). The relative contributions of predictiveness and salience in flavor aversion learning. *Learning & Motivation*, **15**, 188-202.
- KROGH, A., & HERTZ, J. A. (1992). A simple weight decay can improve generalization. In J. E. Moody, S. J. Hanson, & R. P. Lippmann (Eds.), *Advances in neural information processing systems* (Vol. 4, pp. 950-957). San Mateo, CA: Morgan Kaufmann.
- KRUSCHKE, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.
- KRUSCHKE, J. K. (2001a). The inverse base-rate effect is not explained by eliminative inference. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1385-1400.
- KRUSCHKE, J. K. (2001b). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, **45**, 812-863.
- KRUSCHKE, J. K. (2005). Learning involves attention. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (pp. 113-140). Hove, U.K.: Psychology Press.
- KRUSCHKE, J. K., & BLAIR, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, **7**, 636-645.
- KRUSCHKE, J. K., & JOHANSEN, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1083-1119.
- KRUSCHKE, J. K., KAPPENMAN, E. S., & HETRICK, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 830-845.
- LE PELLEY, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology*, **57B**, 193-243.
- LUCE, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- MACKINTOSH, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, **82**, 276-298.
- 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.
- MCLAREN, I. P. L., & MACKINTOSH, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior*, **28**, 211-246.
- RESCORLA, R. A., & WAGNER, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.
- TRABASSO, T., & BOWER, G. H. (1968). *Attention in learning: Theory and research*. New York: Wiley.
- VAN HAMME, L. J., & WASSERMAN, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning & Motivation*, **25**, 127-151.
- WAGNER, A. R., LOGAN, F. A., HABERLANDT, K., & PRICE, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology*, **76**, 171-180.
- WASSERMAN, E. A., & BERGLAN, L. R. (1998). Backward blocking and recovery from overshadowing in human causal judgement: The role of within-compound associations. *Quarterly Journal of Experimental Psychology*, **51B**, 121-138.

### NOTES

1. Effect size ( $d$ ) was computed by dividing the mean difference by the pooled standard deviation.
2. In the present models, attention is only shifted across input nodes. The bias node is excluded from the attentional system because the bias represents tonic context cues instead of phasic trial cues.
3. The exemplar activation equation in the original EXIT model includes a salience parameter ( $\sigma$ ) for each cue. This parameter was excluded (fixed at one) in the present models because it is not useful in the present application. Cue salience is implicit in the cue's input activation value and does not need to be represented as a separate multiplier.
4. In the decay model, the exemplar specificity is extraneous, because exemplar activations have no influence. This is due to the exemplar-to-gain weights remaining fixed at zero because there is no learning on these weights ( $\lambda_x = 0$ ).
5. Limited attention capacity is due to the  $P$  parameter, which in the restricted model acts only to normalize the input (there is no gain). At the limit ( $P \rightarrow \infty$ ), the restricted model would reflect the unlimited capacity of the Rescorla-Wagner (1972) model.

**APPENDIX A**  
**Full Text of Instructions to Participants**

---

Welcome! Thanks for participating.

Please turn off cell phones! Also please stow personal CD players, radios, and all papers and pens, etc. You can place your personal belongings beneath the desk.

Please place your right hand on the mouse, and your left hand at the space bar of the keyboard.

Press the SPACE BAR to continue.

You will see a multi-colored dot pattern on the top part of the computer screen. Your task will be to match the colors in the dot pattern to the correct response choice option at the bottom of the screen. There will be three choice options and you will select one by clicking on it with the mouse.

After you make a choice, the computer will tell you the correct answer. At first you will just be guessing, but after a few repetitions your accuracy will improve.

Press the SPACE BAR to continue.

[Example stimulus is displayed on the upper portion of the screen.]

The color dot patterns appear as shown above. The dot and color locations are randomized, so you will never see the exact same arrangement of colored dots twice. The COLORS themselves will indicate the correct response choice.

Press the SPACE BAR to continue.

All you need to do is learn which choice option on the lower part of the screen is indicated by the color dot pattern on the upper part of the screen. Remember that the colors in the pattern indicate the correct response and the dot positions themselves are random. Be sure to pay close attention because the colored dot patterns only appear for a brief time.

You will continue to be drilled on training cases until you achieve very high accuracy in each phase. Each phase moves on to the next phase only after you have learned well. If you respond accurately, there will be fewer cases in each phase, and the experiment will end sooner.

Press the SPACE BAR to continue.

Try to learn as well as you can, because throughout the experiment you'll be asked to guess which choice option is best indicated by novel colored dot distributions. Another reason to learn well is that errors are accompanied by a buzzing sound and a delay period.

THE EXPERIMENT DOES NOT PROGRESS TO SUBSEQUENT PHASES UNTIL YOU HAVE VIRTUALLY PERFECT ACCURACY. THEREFORE, TO SHORTEN THE LENGTH OF THE TRAINING, RESPOND AS ACCURATELY AS YOU CAN. ERRORS TAKE EXTRA TIME!!!

Press the SPACE BAR to continue.

There will be a lot of trials to work through, but it really only takes about 15 minutes, as long as you maintain a high accuracy in responding.

Please do not adjust the speaker volume. Please do not touch any part of the computer other than the keyboard and the mouse. We appreciate your cooperation.

If you have any questions, please ask now. Otherwise, PLEASE WAIT FOR THE EXPERIMENTER TO CLOSE THE CURTAIN BEHIND YOU, and then press the SPACE BAR to begin.

---



**APPENDIX B**  
**Complete Participant Results**

The complete set of human data is shown in Table B1. The table shows mean response percentages for each cue type in the test phase and the last two blocks of the early training phase. All items in this table are used in the model fits and therefore constrain the parameters of the models.

Table B2 shows the response choice percentages for disallowed responses. These percentages are the number of times that a disallowed response was chosen, divided by the total number of times that the stimulus appeared in the test phase. Each participant was presented each of these items twice, so it is possible that participants learned that certain response options were not allowed, but this is unlikely, given the length of the testing phase.

**Table B1**  
**Response Choice Percentages**

Phase	Equal-Saliency Group				High-Saliency Group			
	Cue Type	Response Type			Cue Type	Response Type		
		R1	R2	R3		R1	R2	R3
Early	A	98.1	0.0	1.9	A	98.4	1.6	0.0
	C	0.0	100.0	0.0	C	2.3	95.3	2.3
	F	0.9	0.9	98.1	F	0.0	0.8	99.2
Test	A	96.3	1.9	1.9	A	65.6	29.7	4.7
	<b>A.B</b>	95.7	0.6	3.7	<b>A.B</b>	91.7	2.1	6.3
	A.C	46.3	53.7	—	A.C	35.9	64.1	—
	A.D	94.4	—	5.6	A.D	76.6	—	23.4
	A.E	88.9	—	11.1	<b>A.E</b>	31.3	—	68.8
	A.F	63.0	—	37.0	A.F	61.9	—	38.1
	A.N	87.0	—	13.0	A.N	78.1	—	21.9
	<b>B</b>	22.2	16.7	61.1	<b>B</b>	70.3	6.3	23.4
	B.C	11.1	88.9	—	<b>B.C</b>	70.3	29.7	—
	B.D	24.1	—	75.9	<b>B.D</b>	70.3	—	29.7
	B.E	20.4	—	79.6	<b>B.E</b>	34.4	—	65.6
	C	1.2	95.1	3.7	C	2.1	94.8	3.1
	C.D	—	90.7	9.3	C.D	—	93.6	6.4
	C.E	—	83.3	16.7	<b>C.E</b>	—	31.3	68.8
	C.F	—	66.7	33.3	C.F	—	78.1	21.9
	C.N	—	90.7	9.3	C.N	—	93.8	6.3
	D	14.8	3.7	81.5	D	17.2	45.3	37.5
	D.E	1.2	3.7	95.1	<b>D.E</b>	1.6	2.6	95.8
E	7.4	5.6	87.0	<b>E</b>	4.7	4.7	90.6	
F	5.6	7.4	87.0	F	10.9	26.6	62.5	
N	13.0	18.5	68.5	N	18.8	50.0	31.3	

Note—Boldface letters indicate salient cues. Dashes indicate disallowed responses.

**Table B2**  
**Response Choice Percentages for Disallowed Responses**

Phase	Equal-Saliency Group				High-Saliency Group			
	Cue Type	Response Type			Cue Type	Response Type		
		R1	R2	R3		R1	R2	R3
Test	A.C	—	—	9.3	A.C	—	—	3.1
	A.D	—	1.9	—	A.D	—	29.7	—
	A.E	—	0	—	<b>A.E</b>	—	3.1	—
	A.F	—	1.9	—	A.F	—	26.6	—
	A.N	—	3.7	—	A.N	—	35.9	—
	B.C	—	—	1.9	<b>B.C</b>	—	—	6.3
	B.D	—	7.4	—	<b>B.D</b>	—	7.8	—
	B.E	—	9.3	—	<b>B.E</b>	—	18.8	—
	C.D	0	—	—	C.D	6.3	—	—
	C.E	1.9	—	—	<b>C.E</b>	1.6	—	—
	C.F	0	—	—	C.F	6.3	—	—
	C.N	1.9	—	—	C.N	1.6	—	—

Note—Boldface letters indicate salient cues. Dashes indicate allowed responses.

**APPENDIX C**  
**Model Predictions**

Best-fitting parameter values for all the models are shown in Table C1. Model predictions are shown in Tables C2–C5. These tables show the response percentages predicted by the model for each cue type. The early training results are from the last two blocks of early training. The models were fit simultaneously to the training and testing data.

**Table C1**  
**Best-Fitting Parameter Values**

Parameter	Model			
	Full	Attention	Decay	Restricted
$\delta_w$	0.0069	–	0.0051	–
$\lambda_g$	0.122	0.970	–	–
$\lambda_x$	0.369	0.0317	–	–
$c$	1.82	1.86	–	–
$P$	3.05	2.44	6.91	1.86
$\phi$	3.41	2.46	3.67	3.46
$\lambda_w$	2.35	3.04	0.220	0.248
$\lambda_b$	0.328	0.815	0.543	0.124
$b$	0.837	0.864	0.0001	0.429
$f$	1.0 (fixed)	1.0 (fixed)	1.0 (fixed)	1.0 (fixed)
$s$	1.56	1.59	1.66	2.15

**Table C2**  
**Best-Fitting Predictions of the Full Model**

Phase	Cue Type	Equal-Salience Group			High-Salience Group			
		Response Type			Response Type			
		R1	R2	R3	Cue Type	R1	R2	R3
Early	A	95.9	2.0	2.1	A	95.9	2.0	2.1
	C	1.7	96.0	2.3	C	1.7	96.0	2.3
	F	1.7	1.1	97.1	F	1.7	1.1	97.1
Test	A	91.4	3.1	5.5	A	72.6	15.7	11.6
	A.B	94.4	2.1	3.5	<b>A.B</b>	95.8	1.6	2.6
	A.C	47.6	52.4	–	A.C	28.6	71.4	–
	A.D	83.9	–	16.1	A.D	82.1	–	17.9
	A.E	84.1	–	15.9	<b>A.E</b>	20.6	–	79.4
	A.F	53.6	–	46.4	A.F	54.4	–	45.6
	A.N	92.6	–	7.4	A.N	85.1	–	14.9
	B	33.4	22.0	44.6	<b>B</b>	71.7	10.0	18.3
	B.C	5.2	94.8	–	<b>B.C</b>	55.6	44.4	–
	B.D	21.0	–	79.0	<b>B.D</b>	69.6	–	30.4
	B.E	21.2	–	78.8	<b>B.E</b>	36.6	–	63.4
	C	2.1	94.3	3.6	C	1.8	96.2	2.0
	C.D	–	86.2	13.8	C.D	–	95.4	4.6
	C.E	–	86.0	14.0	<b>C.E</b>	–	23.3	76.7
	C.F	–	56.6	43.4	C.F	–	78.5	21.5
	C.N	–	92.2	7.8	C.N	–	95.8	4.2
	D	9.5	11.0	79.5	D	22.9	41.5	35.6
	D.E	3.1	3.2	93.7	<b>D.E</b>	2.3	2.1	95.6
E	9.3	10.5	80.2	<b>E</b>	3.8	3.8	92.4	
F	3.3	4.2	92.5	F	10.6	18.1	71.3	
N	20.7	23.1	56.1	N	24.3	44.9	30.8	

Note—Boldface letters indicate salient cues. Dashes indicate disallowed responses.

APPENDIX C (Continued)

**Table C3**  
**Best-Fitting Predictions of the Attention Model**

Phase	Equal-Salience Group				High-Salience Group			
	Cue Type	Response Type			Cue Type	Response Type		
		R1	R2	R3		R1	R2	R3
Early	A	91.8	3.4	4.9	A	91.8	3.4	4.9
	C	3.5	91.1	5.4	C	3.5	91.1	5.4
	F	4.8	2.6	92.6	F	4.8	2.6	92.6
Test	A	88.3	2.9	8.7	A	63.8	21.3	14.9
	A.B	95.6	1.8	2.6	<b>A.B</b>	95.3	1.6	3.2
	A.C	39.3	60.7	–	A.C	32.3	67.7	–
	A.D	80.1	–	19.9	A.D	84.7	–	15.3
	A.E	79.8	–	20.2	<b>A.E</b>	25.7	–	74.3
	A.F	50.7	–	49.3	A.F	47.5	–	52.5
	A.N	94.7	–	5.3	A.N	90.2	–	9.8
	B	36.6	20.7	42.7	<b>B</b>	66.1	10.5	23.4
	B.C	2.3	97.7	–	<b>B.C</b>	62.6	37.4	–
	B.D	15.2	–	84.8	<b>B.D</b>	65.8	–	34.2
	B.E	15.3	–	84.7	<b>B.E</b>	47.0	–	53.0
	C	1.0	96.2	2.9	C	1.4	96.8	1.8
	C.D	–	84.2	15.8	C.D	–	94.5	5.5
	C.E	–	84.5	15.5	<b>C.E</b>	–	25.8	74.2
	C.F	–	64.6	35.4	C.F	–	72.6	27.4
	C.N	–	93.4	6.6	C.N	–	95.8	4.2
	D	4.7	6.4	88.9	D	19.5	43.3	37.2
	D.E	1.9	1.1	97.0	<b>D.E</b>	2.7	1.6	95.6
E	4.6	6.5	88.9	<b>E</b>	5.6	4.4	90.0	
F	3.2	2.9	93.9	F	12.5	20.0	67.5	
N	14.7	24.7	60.7	N	21.1	49.1	29.8	

Note—Boldface letters indicate salient cues. Dashes indicate disallowed responses.

**Table C4**  
**Best-Fitting Predictions of the Decay Model**

Phase	Equal-Salience Group				High-Salience Group			
	Cue Type	Response Type			Cue Type	Response Type		
		R1	R2	R3		R1	R2	R3
Early	A	93.9	3.1	3.0	A	93.9	3.1	3.0
	C	3.1	93.9	3.1	C	3.1	93.9	3.1
	F	3.1	3.1	93.8	F	3.1	3.1	93.8
Test	A	89.5	4.9	5.6	A	83.4	8.8	7.8
	A.B	94.4	2.8	2.8	<b>A.B</b>	94.8	2.7	2.6
	A.C	40.8	59.2	–	A.C	29.8	70.2	–
	A.D	72.0	–	28.0	A.D	82.1	–	17.9
	A.E	72.0	–	28.0	<b>A.E</b>	16.0	–	84.0
	A.F	64.3	–	35.7	A.F	55.8	–	44.2
	A.N	92.2	–	7.8	A.N	89.3	–	10.7
	B	51.7	22.7	25.6	<b>B</b>	79.5	10.5	10.0
	B.C	9.3	90.7	–	<b>B.C</b>	52.7	47.3	–
	B.D	28.2	–	71.8	<b>B.D</b>	84.2	–	15.8
	B.E	28.1	–	71.9	<b>B.E</b>	29.7	–	70.3
	C	2.9	94.2	2.9	C	2.8	94.4	2.8
	C.D	–	82.3	17.7	C.D	–	92.8	7.2
	C.E	–	82.0	18.0	<b>C.E</b>	–	24.9	75.1
	C.F	–	77.4	22.6	C.F	–	78.7	21.3
	C.N	–	95.7	4.3	C.N	–	96.1	3.9
	D	9.2	10.1	80.7	D	21.0	28.1	50.9
	D.E	2.8	2.8	94.4	<b>D.E</b>	2.6	2.7	94.8
E	9.1	10.0	80.9	<b>E</b>	3.6	3.9	92.5	
F	6.5	7.8	85.7	F	8.0	11.3	80.7	
N	26.5	31.0	42.5	N	27.3	38.6	34.1	

Note—Boldface letters indicate salient cues. Dashes indicate disallowed responses.

(Continued on next page)

## APPENDIX C (Continued)

**Table C5**  
**Best-Fitting Predictions of the Restricted Model**

Phase	Equal-Salience Group				High-Salience Group			
	Cue Type	Response Type			Cue Type	Response Type		
		R1	R2	R3		R1	R2	R3
Early	A	88.7	5.6	5.6	A	88.7	5.6	5.6
	C	5.7	88.6	5.7	C	5.7	88.6	5.7
	F	5.7	5.6	88.6	F	5.7	5.6	88.6
Test	A	93.4	2.8	3.8	A	89.9	5.3	4.8
	A.B	93.9	3.0	3.0	<b>A.B</b>	94.7	3.0	2.3
	A.C	49.9	50.1	–	A.C	39.4	60.6	–
	A.D	78.0	–	22.0	A.D	91.0	–	9.0
	A.E	77.9	–	22.1	<b>A.E</b>	15.9	–	84.1
	A.F	53.9	–	46.1	A.F	50.0	–	50.0
	A.N	93.3	–	6.7	A.N	92.3	–	7.7
	B	43.4	23.9	32.7	<b>B</b>	78.5	12.0	9.6
	B.C	7.7	92.3	–	<b>B.C</b>	45.6	54.4	–
	B.D	22.8	–	77.2	<b>B.D</b>	86.4	–	13.6
	B.E	22.8	–	77.2	<b>B.E</b>	24.7	–	75.3
	C	3.0	93.8	3.2	C	2.9	94.1	3.0
	C.D	–	81.0	19.0	C.D	–	94.5	5.5
	C.E	–	81.0	19.0	<b>C.E</b>	–	19.9	80.1
	C.F	–	58.0	42.0	C.F	–	63.0	37.0
	C.N	–	94.3	5.7	C.N	–	95.3	4.7
	D	8.6	8.3	83.1	D	28.1	33.8	38.1
	D.E	3.2	3.2	93.6	<b>D.E</b>	2.8	3.0	94.3
E	8.6	8.3	83.1	<b>E</b>	2.4	2.5	95.1	
F	2.7	2.5	94.8	F	4.4	5.3	90.3	
N	25.9	24.4	49.7	N	29.9	36.6	33.5	

Note—Boldface letters indicate salient cues. Dashes indicate disallowed responses.

Manuscript received July 19, 2005;  
 revision accepted for publication March 6, 2006.)