Journal of Experimental Psychology:
Learning, Memory, and Cognition
2005, Vol. 31, No. 5, 830–845

0278-7393/05/$12.00 DOI: 10.1037/0278-7393.31.5.830

Eye Gaze and Individual Differences Consistent With Learned Attention in Associative Blocking and Highlighting

John K. Kruschke, Emily S. Kappenman, and William P. Hetrick
Indiana University Bloomington

The associative learning effects called *blocking* and *highlighting* have previously been explained by covert learned attention, but evidence for learned attention has been indirect, via models of response choice. The present research reports results from eye tracking consistent with the attentional hypothesis: Gaze duration is diminished for blocked cues and augmented for highlighted cues. If degree of attentional learning varies across individuals but is relatively stable within individuals, then the magnitude of blocking and highlighting should covary across individuals. This predicted correlation is obtained for both choice and eye gaze. A connectionist model that implements attentional learning is shown to fit the data and account for individual differences by variation in its attentional parameters.

Keywords: blocking, highlighting, learned attention, eye tracking, individual differences

The phenomenon of *blocking* has become a touchstone for theories of learning. Reported initially by Kamin (1968, 1969) and found thereafter in numerous species and procedures, the phenomenon forced a revolution in theories of learning. In the blocking procedure, a person initially learns to predict an outcome from a single cue. Subsequently, the cue is always accompanied by a second cue, still leading to the same outcome. People tend not to learn a strong association between the second cue and the outcome; that is, previous learning about the first cue has blocked learning about the second cue. This blocking challenges some theories of learning because the second cue has co-occurred with the outcome quite often and their association should be learned.

A complementary phenomenon called *highlighting* suggests augmented learning about a cue, in contrast to the diminished learning in blocking (Kruschke, 2003a). This phenomenon is extremely challenging to learning theories, even those that were created to account for blocking (such as the Rescorla-Wagner model, which will be explained below). In a highlighting procedure, participants initially learn that a pair of cues leads to an outcome. Later, participants learn that one of those cues, paired with a different cue, leads to a different outcome. The result is that the association from the distinctive cue to the later-learned outcome is apparently very strong; that is, learning about the distinctive cue has been highlighted.

To our knowledge, only attentional theory (Kruschke, 2003a) explains both highlighting and blocking. According to this approach, highlighting occurs because people have learned to attend to the distinctive cue for the new outcome. Blocking occurs, at least in part but not fully, because people have learned to ignore the redundant cue. The attentional theory will be explained in more detail later; for now we wish to emphasize that the attention posited in this theory is a theoretical construct referring to a covert aspect of cognition. The formal theory simply states that more strongly attended cues are multiplied by a larger factor in responding and learning. The magnitude of the covert attentional factor is inferred from observed choice data via a mathematical model.

The present research is based on the additional premise that overt eye gaze is correlated with the covert attention hypothesized by the learning theory. If this is true, then blocked cues should be gazed at less than control cues, and highlighted cues should be gazed at more than control cues. In this article we report strong confirmation of these predictions. Although the formal attentional theory makes no commitment to the eye gaze premise, and although the exact mechanisms by which covert attention is manifested in eye gaze are unspecified, the new eye gaze results are highly suggestive of attentional processes in associative learning.

In mathematical models of attentional learning (e.g., Kruschke, 2001b), the magnitude of attentional shifting and learning is governed by corresponding parameter values. These parameter values might be determined in part by situational influences and random noise. We make the additional assumption that the attentional shifting and learning magnitudes are, in part, relatively stable and enduring individual traits. This assumption implies that individuals with low attentional shifting rates will show small amounts of blocking and highlighting, whereas individuals with high attentional shifting rates will show large amounts of blocking and

John K. Kruschke, Emily S. Kappenman, and William P. Hetrick, Department of Psychology, Indiana University Bloomington.

This research was supported in part by Grant BCS-9910720 from the National Science Foundation to John K. Kruschke, a Young Investigator Award from the National Alliance for Research on Schizophrenia and Grant R03-MH066149-01A1 from the National Institute of Mental Health to William P. Hetrick, and the Clinical Science Research Training Grant T32-MH017146-20 from the National Institute of Mental Health to Indiana University, Richard McFall, Principle Investigator.

We thank Patrick Hill, Blake Hulet, Angela Prather, and Ruth Ann Weaver for their assistance in administering the experiment. Thanks also to Derek Koehler, Douglas Medin, and Michael Young for constructive comments on a draft. John K. Kruschke's World Wide Web page is at http://www.indiana.edu/~kruschke/

Correspondence concerning this article should be addressed to John K. Kruschke, Department of Psychology, Indiana University Bloomington, 1101 East 10th Street, Bloomington, IN 47405-7007. E-mail: kruschke@indiana.edu

highlighting. In other words, the magnitudes of blocking and highlighting should covary across individuals.

This article reports the first experiment in which blocking and highlighting have been measured in the same individuals. Results confirm the predicted correlation of blocking and highlighting, in both choice responses and in eye gaze. Although it is conceivable that other individual differences could account for the observed covariation of blocking and highlighting, the new results are highly suggestive of attentional processes in associative learning.

To bolster the argument that attention can account for covariation of blocking and highlighting, we present results of computer simulations of the EXIT model (Kruschke, 2001a, 2001b). The model was originally proposed to address choice proportions only; in this article we introduce a mapping from attention in the model to eye gaze in human data. The simulations demonstrate that blocking, highlighting, and eye gaze covary robustly with attentional parameters. Changing nonattentional parameters, such as associative learning rate or choice decisiveness, do not yield such correlations.

In subsequent sections of this introduction, we describe the designs of typical blocking and highlighting experiments, and we supply a few more details and background for attentional learning theory. We also provide a little more justification for the hypothesis bridging covert attention to eye gaze and provide some background regarding individual differences in attention and learning. After the introduction, we present the new experiment and its results. Subsequently, we describe model simulations demonstrating how the model accounts for covariation of blocking and highlighting. We conclude with a review of the main results and a brief discussion of the multifaceted aspects of attention in learning.

Blocking and Highlighting Designs

Table 1 shows details of a typical design for a blocking experiment. In the early training phase, a trial with cue A always has outcome X, denoted A→X in Table 1. During this early phase, intermixed trials of F→Y occur so that the learner must discriminate A and X from other potential cues and outcomes. After A→X has been well learned, the later training phase includes trials of cue A always accompanied by a second cue B, still leading to the same outcome, and denoted A.B→X in Table 1. To test the degree of learning about cue B, the later training phase includes an equal number of trials of C.D→Y. If all that matters for learning is the frequency of co-occurrence, then the strength of association from B to X should be the same as the strength of association from D to Y. To test this prediction, the final test phase includes trials that display cues B and D together (denoted B.D→?), asking participants to provide their best guess on the basis of what they have learned. Results for probe B.D→? are that people have a strong preference for outcome Y, which suggests that the association from D to Y is stronger than the association from B to X.

Table 1
Essence of Blocking and Highlighting Designs

	Design			
Phase	Blocking		Highlighting	
Early	A→X	F→Y	I.PE→E	
Late	A.B→X	C.D→Y	I.PE→E	I.PL→L
Test	B.D→? (Y) A.C→? (X)		PE.PL→? (L) I→? (E)	

Note. Each cell indicates cues→correct response. In the test phase, typical response tendencies are shown in parentheses.

Table 1 also shows details of a typical design for a highlighting experiment. In the early training phase, people see trials with cues PE and I indicating the outcome E, denoted I.PE→E in Table 1. E denotes the early-learned outcome, PE denotes a perfect predictor of outcome E, and I denotes an imperfect predictor. In the later training phase, participants learn about a new outcome *L* indicated by cues *I* and *PL,* denoted I.PL→L. Notice that the early- and later-trained outcomes have symmetric cue structures: Each outcome has one perfect predictor and one shared imperfect predictor. If people learn the symmetry (e.g., as normative statistical models require), then cue I should not be differentially predictive of the outcomes, and cues PE and PL should be equally predictive of their respective outcomes. This prediction is examined in the final test phase. When presented with cue I by itself, there is a robust tendency for people to respond with the early-trained outcome E. But this tendency is not just a general preference for the higher base-rate outcome in the face of ambiguous cues, because when presented with cues PE and PL, there is a robust tendency to respond with the later-trained outcome L. The results suggest that people's knowledge is asymmetric: Cue I is more strongly associated with outcome E than with L, but cue PL is more strongly associated with outcome L than cue PE is associated with outcome E.

Explanations of Blocking and Highlighting

The dominant explanation of blocking states that associations are built only when the outcome is surprising or mispredicted. This idea of error-driven learning was formalized in the classic model of Rescorla and Wagner (1972). When a learner sees a case of A.B→X during the later training phase, the outcome X is fully anticipated by the cue A. Therefore, because there is little surprise in the outcome, there is little learning about cue B. Other theorists, notably Mackintosh (1975), suggested that there is not merely lack of learning about the redundant cue B, there is learned suppression of attention to it. Some researchers take for granted the theoretical stance that blocking involves learned (in-)attention (e.g., Crookes & Moran, 2003), but evidence that attention is involved in blocking is relatively rare. One type of evidence comes from studies of learning about a blocked cue subsequent to its being blocked. If people have learned to ignore a blocked cue, then subsequent learning about that cue should be retarded. This prediction has been confirmed (in both humans and rats; Kruschke & Blair, 2000; Mackintosh & Turner, 1971). Although the Rescorla–Wagner model can explain blocking, per se, it cannot explain subsequent retarded learning about a blocked cue. Notice that the two mechanisms are not mutually exclusive: There can be lack of learned association and learned inattention simultaneously. Indeed, the connectionist model described later incorporates both processes governed by the same goal: error reduction.

The Rescorla–Wagner model also cannot explain highlighting, because the model predicts that symmetric associations should ultimately be learned. A variation of the model was proposed by Markman (1989), whereby absent-but-expected cues are encoded with negative activations, but no method was described whereby cue expectancies could be learned, and this approach to highlighting has not been further pursued in the literature. Juslin, Wennerholm, and Winman (2001) proposed that highlighting was not so much a learning effect as a decision effect. According to this theory, when a participant is presented in the test phase with a cue combination that she or he does not know, such as PE.PL, the participant infers that the correct outcome must be one that she or he does not know. Therefore the participant eliminates the well-known early-learned outcome E, and selects the poorly known later-learned outcome *L.* Decision strategies such as eliminative inference are, no doubt, being used by participants in these experiments, but the particular approach proposed by Juslin et al. (2001) was shown to have serious shortcomings by Kruschke (2001a). For example, when human participants are presented with cue I by itself, they tend to respond with outcome E, but when presented with the triplet of cues I.PE.PL, the preference for outcome E is greatly reduced or even reversed. This reduction is not addressed by the eliminative inference model.

Highlighting seems currently to be best explained by attention shifting. When learning the later-trained cases of I.PL→L, attention shifts away from cue I because cue I predicts the previously learned outcome E. Instead, attention shifts to the distinctive cue PL, and a strong association is learned between it and outcome L. This type of attentional theory was described by Medin and Edelson (1988) to explain the inverse base-rate effect, from which the highlighting effect was derived.

The attentional theory was formalized in the attention to distinctive input (ADIT) model (Kruschke, 1996) and its successor, the EXIT model (Kruschke, 2001a, 2001b). The EXIT model has been shown to account for both highlighting and retarded learning after blocking. Later in this article we describe the EXIT model more thoroughly and introduce predictions of eye gaze from attentional strengths in the model. We also show that if we increase the attentional parameters, then blocking and highlighting and gaze preferences are simultaneously increased. Thus, if the attentional parameter values are considered to reflect individual differences, the model accounts for covariation of blocking and highlighting and gaze preference across individuals.

Attention and Eye Gaze

The attentional theory of blocking and highlighting posits covert attentional mechanisms. An attended cue is supposed to have a stronger impact on responding, or a larger learning rate, or both. These covert attentional processes influence overt choice preferences. These models fit choice data fairly well in a variety of experiments. The good fits countenance the covert attentional constructs. It would be even more compelling to have independent evidence that selective attention to cues is involved.

One potential measure of attention to cues is eye gaze. Intuitively, a person should gaze longer at cues she or he has learned to attend to than at cues she or he has learned to ignore. There is, however, no necessary link between the covert attention posited by learning theories and overt eye gaze. Indeed, there are reports of failures to find cognitive processes reflected in eye gaze, when there plausibly could have been (e.g., Anderson, Bothell, & Douglass, 2004; Lohmeier, 1996). But there are many precedents suggesting that eye gaze is an overt indicator of cognitive attention or relevance. For example, Kaakinen, Hyönä, and Keenan (2002) had people read expository text with a certain topical perspective in mind, so that some topics would be more relevant to the reader than others, and presumably the reader would pay more attention to the relevant text than the irrelevant passages. The researchers found that eye fixations on topic-relevant text were indeed of longer duration than fixations on irrelevant text.

Most directly related to the present investigation, Rehder and Hoffman (2005, in press) reported eye tracking results from category learning tasks. Different cues were differently predictive of the category label. People had to learn the correct category label for each set of training cues. After learning the categories, people looked longer (or even exclusively) at relevant cues than at irrelevant cues. The current investigation goes beyond Rehder and Hoffman's in several ways. First, the category structures are interestingly different. Whereas Rehder and Hoffman noted correspondences of eye gaze with dimensional relevance—meaning simple monotonic correspondence of stimulus relevance with gaze—we are looking for *reduced* gaze to a *predictive* cue (i.e., the blocked cue B), *differential* looking to *equally* predictive cues (i.e., cues PE and PL in highlighting), and *context specific* gaze at a single cue (i.e., cue I in highlighting). Second, we measure gaze at present and absent cues in random locations, whereas Rehder and Hoffman considered always-present dimensions in fixed locations. Thus, in Rehder and Hoffman's experiments, spatial location was learned as an indicator of relevance, but in our experiments, viewers looked at all cues and therefore differential gaze could be much more subtle in magnitude. Third, we are specifically interested in assessing individual differences and correlations across tasks. Fourth, we do explicit model fitting of choice and gaze simultaneously, with emphasis on explaining correlated individual differences via attentional parameters.

In summary, we have reasonable precedents for the hypothesis that eye gaze could correlate with covert learned attention. The hypothesis is bolstered by the new results reported in this article.

Individual Differences and Correlations

Our premise is that the degree of attentional shifting or learning is, at least in part, an individual characteristic that is stable over short time scales (but may change with context or over developmental time ranges). In a formal model of attentional learning that we detail later, individual levels of attentional processing are reflected by parameter values that govern attention in the model. The model predicts that as attentional shifting and learning increase, so should the magnitudes of blocking and highlighting, as measured by both choice and gaze. Therefore the magnitudes of highlighting and blocking should covary across individuals.

There are numerous precedents for treating attentional ability as a stable individual trait. Here we point out only a few.

Individual differences in blocking per se have been previously reported and used as indices of attentional processing. For example, Crookes and Moran (2003) found age and gender differences in blocking. In their blocking task (called the *mouse in house* task), people learned to associate color patches at the top of a computer screen with displayed goal boxes to which they had to move the cursor using a joystick. Learning (or lack thereof) was measured as the latency until the correct goal box was entered. The authors found greater blocking for females than for males, and a larger proportion of participants who showed blocking as age increased through childhood to early adolescence. Crookes and Moran (2003, p. 472) stated that "[blocking] can be interpreted as the habitual filtering of unnecessary information" (i.e., as learned inattention to the blocked cue). They were especially concerned with blocking as a diagnostic measure of attentional abilities in the context of schizophrenia, which tends to show symptoms in adolescence and in males more than females. For our purposes, we take this research as a precedent that there are meaningful individual differences in blocking, and that these differences can be interpreted in terms of attentional abilities.

Gibbons, Rammsayer, and Lubow (2001, Experiment 2) reported correlated individual differences in two attentional learning tasks. One task involved visual search, the other task involved learning the relationship of a cue and an outcome. In the visual search task, people had to determine whether a display consisted of 20 identical squiggles or, instead, had a single different squiggle among 19 identical distractors. The design included an initial phase with multiple trials that used the same distractor squiggles. This phase was intended to encourage people to learn to ignore that specific distractor. A subsequent phase included some trials in which the initial distractor became the unique target among novel distractors. Results showed relatively slow detection of the target on these trials (i.e., apparently, people had indeed learned to ignore the initial distractor). In the learning task, a trial display consisted of letter trigrams surrounded by a polygonal shape, both of which could change across trials. Participants had to learn what aspect of the display indicated that a penalty counter would increment. Participants could press a button when they thought the counter was about to increment, thereby reducing the penalty if they were correct. The indicative cue was the shape of the surrounding polygon. The key result was that people were slower to learn about the relevance of the polygon if they had been preexposed to a situation in which the polygon was irrelevant. In the preexposure task, people simply had to count the number of repetitions of the trigrams while the surrounding polygon was held constant. Again, the key findings for our purposes was that both effects were interpreted in terms of learned attention and that there were individual differences correlated across tasks.

Engle (2002) and colleagues have argued that individual ability in executive attention, meaning the capacity to avoid distraction from misleading cues, is crucial for explaining individual ability in working memory. To provide evidence for the putative correlation of attention and working memory, Kane, Bleckley, Conway, and Engle (2001) used an eye tracking procedure to assess attention. People were required to respond to a target that could appear on either side of the display. The target was preceded by a prompt on the opposite side of the display. People who were better able to suppress visual saccades to the prompt tended to have higher scores on working memory tasks. These results indicate individual differences in attention, as measured by eye gaze, that are correlated with short-term learning.

Beyond the empirical work cited above, there are also precedents of formal modeling that specifically address individual differences in learned attention. For example, Webb and Lee (2004) examined individual differences in a category learning task reported by Kruschke (1993). In that task, people learned to classify simple rectangles that varied (across trials) in height and in the position of an internal vertical line segment. In a filtration structure, people could learn to ignore one aspect (e.g., segment position) and correctly classify by attending to the other aspect (e.g., height). Webb and Lee (2004) found individual differences in patterns of learning (e.g., some participants achieved high accuracy very rapidly whereas others learned gradually). Webb and Lee characterized those differences, in part, by different attentional learning rates in a formal attentional learning covering map (ALCOVE) model (Kruschke, 1992).

In summary, we described a few precedents for our hypotheses regarding individual differences in attentional abilities. We established that there are, in fact, individual differences in attentional abilities, as manifested in associative blocking, antisaccade tasks, visual search, learned irrelevance, and so on. Performance in these attentional tasks covaries across individuals, such that people with relatively high performance in one task tend to have relatively high performance in another. Individual differences have been addressed in a formal model as differences in the value of an attentional learning parameter. In the present research we apply these ideas to eye gaze in blocking and highlighting. To the extent that blocking and highlighting rely on attentional shifting and learning, then magnitudes of blocking and highlighting should covary across individuals.

Having given some background to motivate an expected correlation between blocking and highlighting, it is worth mentioning that such a correlation is by no means a foregone conclusion. For example, the classic Rescorla-Wagner model of blocking predicts no highlighting. The eliminative inference model of highlighting (Juslin et al., 2001) predicts no blocking (see the subsection Other Models near the end of this article). The rule-plus-exception (RULEX) model (Nosofsky, Palmeri, & McKinley, 1994) might address both blocking and highlighting, but it is unlikely that it would predict any correlation of their magnitudes (again, see the subsection Other Models below). In other words, there is no necessary theoretical connection between magnitude of blocking and magnitude of highlighting.

It would also be wrong to infer that blocking and highlighting should be correlated just because the stimuli and learning task are so similar. In other empirical work (Kappenman, Kruschke, & Hetrick, 2005), we observed no hint of a correlation between highlighting and illusory correlation, which is another learning phenomenon related to base rates. Thus, it is not the case that any two learning effects will correlate merely by virtue of generic task similarity.

Never before has there been an investigation of blocking and highlighting in the same individuals, and never before have eye movements been measured during these tasks. Finding that blocking and highlighting covary would suggest that attention is involved in both effects.

Experiment: Eye Tracking During Blocking and Highlighting

Our goal was to measure eye gaze at the various cues. We therefore required cues that must be fixated to be perceived and that are spatially separated. Many previous experiments in blocking and highlighting have used written words as cues, and we continued this procedure. Participants made responses by clicking with the cursor one of four boxes also marked by words. An example of a stimulus display, with a superimposed eye gaze trajectory, is shown in Figure 1. One cue is a word displayed in the top left (orange) rectangle, and another cue is a different word displayed in the top right (purple) rectangle. The left–right position of the cues was counterbalanced across trials.

Every participant went through a blocking design and a highlighting design in counterbalanced order across participants. Details of the implemented designs are shown in Tables 2 and 3. These designs duplicate the abstract designs of Table 1. For example, where the abstract design has A.B→X, the implemented design has two copies: A1.B1→X1 and A2.B2→X2. This duplication makes more than one correct response possible in the early phase of highlighting, so that participants cannot merely learn to click a single response regardless of the cues. The duplication also makes the task more challenging. Moreover, the left–right locations of the cues were counterbalanced; for example, both A1.B1→X1 and B1.A1→X1 were displayed on different trials.

Each cue type appeared in a unique-color rectangle. For example, for a given experiment run, the blocking cues A1 and A2 may have always appeared on an orange rectangle, whereas the blocked cues B1 and B2 may have always appeared on a purple rectangle. The colors, therefore, did not indicate the correct response, because multiple correct responses occurred for any color. People might therefore learn to ignore color. On the other hand, people might learn to use the colors to guide attention (e.g., attending more to the color of the blocking cues and less to the color of the blocked cues). Results reported below indicate that color had little if any effect on gaze preference, and so this aspect of the procedure is not emphasized here.

Method

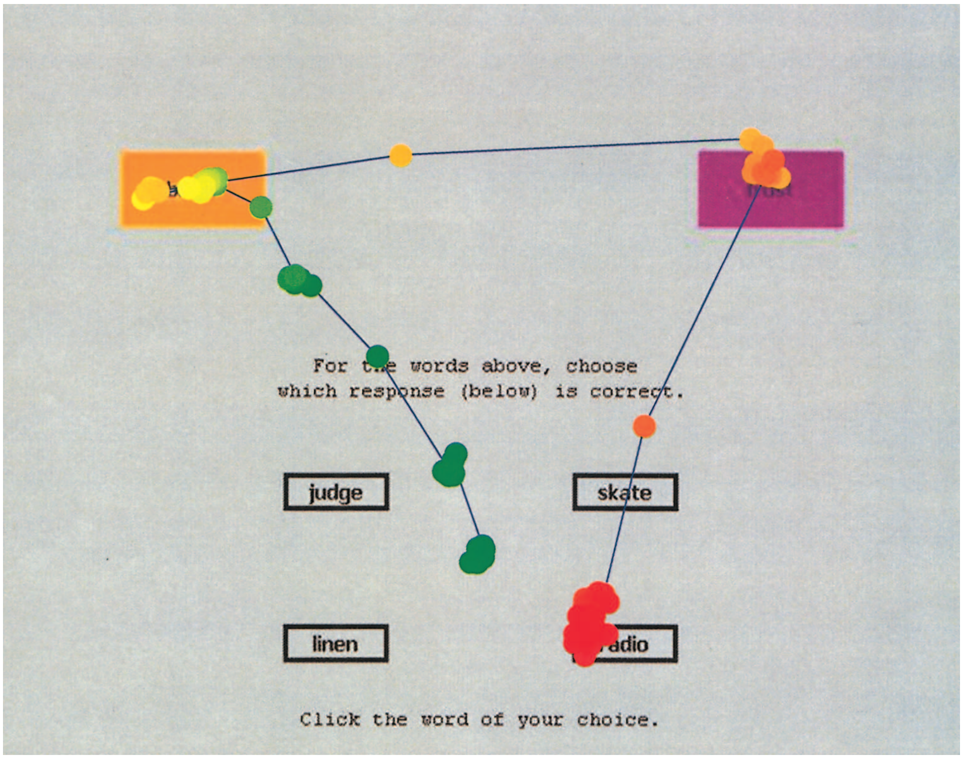
Participants

A total of 65 students from introductory psychology courses at Indiana University participated for partial course credit. Students were asked to participate only if they had normal (or corrected to normal) acuity and color vision. Participants were also told that no eye makeup could be worn at the time of the experiment because it could confuse the eye tracker.

Despite heroic attempts to coddle, cajole, and coerce the eye tracker, it could not be successfully calibrated on some participants, leaving 42 participants in the blocking experiment and 37 participants in the highlighting experiment. The calibration failures appeared to be caused by long or dark eye lashes or low-riding upper eye lids, which partially obscured the pupil. Presumably, such superficial characteristics are not systematically correlated with learning and attention. Of the 42 participants included, the mean age was 19.60 years (*SD* = 1.15), 25 were male and 17 were female, and 38 were right-handed and 4 were left-handed.

Apparatus

Participants were seated in front of a desktop computer with a 15 in. monitor and a standard keyboard and mouse. The participant straddled a rod that was attached to the seat and extended to the participant's chin. The rod height and angle were adjusted so that the chin could be comfortably



For the words above, choose which response (below) is correct.

Click the word of your choice.

Figure 1. Example of a stimulus display with an eye gaze trajectory superimposed. Each colored dot represents the gaze direction for 1/60th s. The trajectory begins with the green dots near the center, proceeds through the yellow dots at the upper left cue and the orange dots at the upper right cue, and concludes at the red dots at the lower right response box. (Plot generated in Matlab with a program created by John K. Kruschke.)

Table 2
Details of Blocking Design

Phase	Trial items				Duration
Early	A1.␣→X1	␣.A1→X1	F1.␣→Y1	␣.F1→Y1	Until 15/16 correct in 2 consec. blocks (3 min.)
	A2.␣→X2	␣.A2→X2	F2.␣→Y2	␣.F2→Y2	
Late	A1.B1→X1	B1.A1→X1	C1.D1→Y1	D1.C1→Y1	Until 15/16 correct in 2 consec. blocks (3 min.)
	A2.B2→X2	B2.A2→X2	C2.D2→Y2	D2.C2→Y2	
Test	*A.B: (each shown twice)*		*C.D: (each shown twice)*		48 trials
	A1.B1→X1	B1.A1→X1	C1.D1→Y1	D1.C1→Y1	
	A2.B2→X2	B2.A2→X2	C2.D2→Y2	D2.C2→Y2	
		D.B:			
	D1.B1→?	B1.D1→?	D2.B2→?	B2.D2→?	
	D2.B1→?	B1.D2→?	D1.B2→?	B2.D1→?	
	C1.B1→?	B1.C1→?	C2.B2→?	B2.C2→?	
	C2.B1→?	B1.C2→?	C1.B2→?	B2.C1→?	
		A.C:			
	A1.C1→?	C1.A1→?	A2.C2→?	C2.A2→?	
	A1.C2→?	C2.A1→?	A2.C1→?	C1.A2→?	
	A1.D1→?	D1.A1→?	A2.D2→?	D2.A2→?	
	A1.D2→?	D2.A1→?	A2.D1→?	D1.A2→?	

Note. Trial items indicate left-cue.right-cue→correct-response. A "␣" symbol indicates a cue position that was unoccupied. Each cue type (A, B, C, D, and F) appeared in a different color. Consec. = consecutive; min. = minimum.

rested. Participants were free to move their heads if necessary. At the lower right (from the participant's perspective) of the monitor was a small electronic camera lens, aimed at the participant's right eye. The camera was wired to an eye tracking computer in an adjacent room.

We used a SensoMotoric Instruments (Needham, MA) system with iView X software. This dark pupil system uses a remote eye tracking camera without any apparatus mounted directly on the participant's head. The face and eye are illuminated by an infrared (IR) source that rides atop the camera. The pupil appears dark to the IR sensors in the camera and the pupil position is calculated every 1/60 s (16.7 ms) by pattern recognition software. The camera is mounted on motors that automatically compensate for small head motions, detected by the corneal reflex.

Stimuli

An example of the stimulus display is shown in Figure 1. The computer monitor was approximately 70 cm from the viewer's eyes. The cue words were separated by 18.4 cm, that is, about 14.7° visual angle. Each cue word was approximately 1.6 cm wide, therefore spanning approximately 1.3° visual angle.

Stimulus and response words were limited to 5-letter nouns with familiarity, imagability, and concreteness ratings of 500 or higher, as recorded in the MRC Psycholinguistic Database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm). From this pool of candidate words, 20 were chosen that had different initial letters and relatively few obvious semantic

Table 3
Details of Highlighting Design

Phase	Trial items				Duration
Early	I1.PE1→E1	PE1.I1→E1			Until 11/12 correct in 3 consec. blocks (5 min.)
	I2.PE2→E2	PE2.I2→E2			
Late, 3-to-1 base rates	*I.PE, 3 times each:*		*I.PL, 1 time each:*		16 trials
	I1.PE1→E1	PE1.I1→E1	I1.PL1→L1	PL1.I1→L1	
	I2.PE2→E2	PE2.I2→E2	I2.PL2→L2	PL2.I2→L2	
Late, 1-to-3 base rates	*I.PE, 1 time each:*		*I.PL, 3 times each:*		16 trials
	I1.PE1→E1	PE1.I1→E1	I1.PL1→L1	PL1.I1→L1	
	I2.PE2→E2	PE2.I2→E2	I2.PL2→L2	PL2.I2→L2	
Test	*I.PE, 1 time each:*		*I.PL, 1 time each:*		24 trials
	I1.PE1→E1	PE1.I1→E1	I1.PL1→L1	PL1.I1→L1	
	I2.PE2→E2	PE2.I2→E2	I2.PL2→L2	PL2.I2→L2	
	I, 2 times each:		*PE.PL, 2 times each:*		
	I1.␣→?	␣.I1→?	PE1.PL1→?	PL1.PE1→?	
	I2.␣→?	␣.I2→?	PE2.PL2→?	PL2.PE2→?	

Note. Trial items indicate left-cue.right-cue→correct-response. A "␣" symbol indicates a cue position that was unoccupied. Each cue type (I, PE and PL) appeared in a different color. Consec. = consecutive; min. = minimum.

relationships (e.g., *shark* was excluded because of its strong semantic relation to *ocean*). The assignment of words to cues and responses was randomized anew for every participant and experiment design. The 20 words were as follows: *apple, brain, cigar, daisy, elbow, frost, glass, house, ivory, judge, knife, linen, movie, ocean, phone, queen, radio, skate, tiger, world.*

The cue words appeared on colored rectangles. The colors were chosen to be highly distinctive and of roughly equal salience. The assignment of colors to abstract cue types was randomized anew for each participant and experiment design. The six colors were as follows (accompanied in parentheses by their corresponding red–green–blue [RGB] values on a 0–255 scale): *green* (0, 210, 0), *blue* (0, 100, 255), *yellow* (225, 225, 0), *red* (210, 50, 50), *orange* (255, 165, 0), and *purple* (211, 0, 252). The background was always grey (190, 190, 190).

As indicated explicitly in Tables 2 and 3, the left–right positions of the cues were counterbalanced across trials. If each cue had instead occupied a unique position, then the mere presence or absence of each cue could have been ascertained from the participant's peripheral vision alone, and so eye gaze would have been a useless measure. Previous experiments involving highlighting and blocking have also used cues in varying positions; it is not known whether fixed-location cues would alter these effects (cf. Young & Wasserman, 2002).

The four response positions were fixed. In blocking, X1 was the upper left and X2 was the lower right, and Y1 was the upper right and Y2 was the lower left. In highlighting, E1 was the upper left and E2 was the lower right, and L1 was upper right and L2 was lower left.

Procedure

Every participant learned the highlighting and blocking designs in an order that was counterbalanced across participants. Before each design, the eye tracker was calibrated using an automated nine-point procedure supplied with the eye tracker. Within each experiment, the participant first read instructions that included an example of the stimulus display.

Immediately after the instructions and before the beginning of training, an additional calibration display was shown. This display allowed fine-tuning of the approximate screen coordinates delivered by the eye tracker. The pretraining calibration consisted of four blinking fixation spots, two centered at the cue positions and two centered at lower response positions. When analyzing the eye tracking data, a Matlab program written by the first author determined the tracker-reported location during each fixation spot[1] and then determined the best linear mapping to the true fixation point positions. This correction was applied to all subsequent tracker positions for that experiment session.

Within each trial of training, color patches appeared with the response prompt and response boxes for 750 ms. Then the cue words appeared in the color patches. Participants made a response by moving the mouse (cursor) and clicking a response box. The response prompt was then replaced by corrective feedback, with the cues remaining visible. For correct responses, the feedback said, "Yes! The correct answer is [word]." For incorrect responses, the feedback said, "Wrong! The correct answer is [word]" and was accompanied by a brief buzzing sound. On test trials for which no correct answer was supplied, the feedback said, "Your response has been recorded." Participants could study the cues and feedback as long as they liked. In all cases, the feedback was accompanied by a clickable box labeled *Next* that appeared centered among the four response words, which the participant had to click to see the next trial. The action of clicking this target also centered the cursor among the four response options before each trial. An intertrial blank screen of 500 ms occurred after the click.

The duration of each phase of training and testing is summarized in Tables 2 and 3. In the blocking design, participants continued in each training phase until a learning criterion was reached. Each training block had eight trials, and participants had to achieve 15/16 correct across two consecutive blocks (with a minimum of three blocks trained) before moving to the next phase. In the highlighting design, only the first phase had a learning criterion, requiring 11/12 correct across three blocks of four trials (with a minimum of five blocks trained).

Dependent Variable: Gaze Duration Difference

Figure 1 shows a particularly clear example of an eye gaze trajectory. Each dot shows the gaze location at a sample 1/60th s. The total duration of the trajectory is the time from cue onset (i.e., the blank colored rectangles) to response click. The color of the dot indicates the relative time at which the gaze was sampled, with green at the beginning, yellow in the middle, and red at the end of the trajectory. Many trials were not this clean, instead peppered with tracker noise, presumably uncorrelated with cue identity. A rectangular region of interest (ROI) around each cue was liberally defined to ensure that no gazes near the cues were omitted. The colored rectangle behind each cue was 15% screen width by 10% screen height, and the corresponding ROI was 27% screen width and 26% screen height. On each trial, the gaze duration at each cue was defined simply as the number of 1/60th sample locations within the corresponding ROI, from cue onset until response. This gaze duration does not include gaze during feedback or intertrial intervals.

Relative gaze duration was measured as follows. Consider a trial with cue A displayed on the left and cue B displayed on the right. Denote the time gazing at A on the left by A_L, and the time gazing at B on the right by B_R. The difference, $A_L - B_R$, indicates how much more time was spent gazing at A than at B. Unfortunately, the cause of that difference could be either the identity of the cues or the position of the cues. It turns out that the left cue tended to be gazed at first and longer than the right cue, presumably because of the strong tendency for English readers to scan text left to right. Fortunately, the positions of the cues were counterbalanced across trials, and the position effect can be removed, on average, by including the difference of gaze times when the positions of the cues were reversed. We define $AmBg = [(A_L - B_R) + (A_R - B_L)]/2$, where the notation *AmBg* is supposed to suggest A minus B for gaze durations. If the difference in gaze times to A and B is caused only by position, then *AmBg* will be zero. On the other hand, if A tends to be gazed at longer than B, then *AmBg* will be positive. Analogous average differences are used throughout the analyses below.

Results

We first report results regarding choice data, because there is little motivation for examining the eye gaze data if blocking and highlighting are not manifested in choice. Next we report the results regarding eye gaze for both blocking and highlighting. Finally, we consider individual differences and covariation across individuals.

Whenever outliers are culled in the following analyses, outliers are defined as any score below the 25th percentile minus 1.5 times the interquartile range, or any score above the 75th percentile plus

[1] The tracker-reported location of a fixation spot was determined as the median location of gazes that were both in the proximity of the fixation spot and stable across consecutive samples. A gaze was considered to be in the proximity of the spot if it was within 12% of the screen width, left or right of the spot, and within 14% of the screen height, above or below the spot. A sample gaze was considered to be stable if it was within the proximity of the two sample gazes immediately preceding it and the two sample cases immediately succeeding it. A minimum of 150 proximal and stable samples were used to determine the median location; otherwise the calibration was considered unstable and the participants' data were not included in further analysis.

1.5 times the interquartile range (and where the interquartile range is the 75th percentile minus the 25th percentile). This is a conventional definition (e.g., it is the default definition of outliers for boxplots in SPSS statistical software).

We only included data from trials in which people actually looked at the cues. That is, there were a few trials in which people responded but the eye tracker recorded no gaze directed at the cues. Those few trials were excluded from the data analysis. All statistical tests are two-tailed, with significance (i.e., reliability) taken to be $p < .05$.

In subsequent sections, all of the results are collapsed across experiment order. Results were also examined separately for the blocking-first and highlighting-first groups. The major trends were evident within both orders, with no notable differences between orders. There was little statistical power to detect differences, however, because of the small sample size within each order.

All of the results were also examined for female or male participants only. Both groups showed the same trends as the collapsed data reported below, with no obvious differences. Again, there was little statistical power to detect differences between groups, so we could not detect sex differences analogous to those reported by Crookes and Moran (2003).

We also examined gaze durations during the beginning of cue presentation when the colored rectangles were blank. No statistically significant preferences were found, although there were some weak trends consistent with blocking and highlighting as reported below. We therefore report no further details of the blank rectangle period. All of the gaze results reported below refer to the period from blank rectangle onset to response click.

Response Choice

Blocking. We are interested only in data from participants who retained memory of the training items in the test phase. Therefore we first culled outlying individuals on the basis of total percentage correct on the A.B and C.D training items in the test phase. There was just one outlier, leaving $N = 41$ for subsequent analyses of the blocking data.

Performance in the test phase on the training items was fairly good. The test phase had eight A.B trials (see Table 2), with average performance across the 41 participants of 92.3% correct. Performance on the eight C.D trials was not quite as high, with average accuracy of 87.7%. The somewhat low accuracy on C.D trials is actually useful, however, insofar as it makes blocking more challenging to demonstrate statistically, because the blocked cue B must generate an even weaker response than the somewhat weak control cues C and D.

On the test trials that combined the blocked cue B with control cues C or D, participants chose the response corresponding with the control cue 55.8% of the time and the response corresponding with the blocked cue only 21.2% of the time. (The remaining 23.0% of responses were for the two remaining choices.) For each participant we considered the 16 test trials that combined the blocked cue B with control cues C or D (see test items labeled *D.B* in Table 2) and counted the number of times the participant gave the corresponding Y outcome, minus the number of times the participant gave the corresponding X outcome, divided by the number of *D.B* trials. This difference in choice proportions is denoted *DmBc,* which is meant to suggest D minus B for choices. The mean difference, *DmBc,* across the 41 participants was 34.5%, which is reliably greater than zero, $t(40) = 5.84$, $p < .001$. Thus, there was a robust blocking effect.

As a check that there was not merely a bias to choose the control-cue response when conflicting cues appeared, we also examined choice preferences for trials that combined the blocking cue A with control cues C or D. Specifically, for each participant we considered the 16 test trials that combined the blocking cue A with control cues C or D (see test items labeled *A.C* in Table 2) and counted the number of times the participant gave the corresponding X outcome, minus the number of times the participant gave the corresponding Y outcome, divided by the number of A.C trials. This difference in choice proportions is denoted *AmCc,* which is meant to suggest A minus C for choices. Overall, participants chose the response corresponding with the blocking cue *A* 70.3% of the time, and the response corresponding with the control cue just 20.3% of the time. The mean difference of $AmCc = 50.0\%$ was reliably greater than zero, $t(40) = 8.30$, $p < .001$.

In summary, both the D.B and A.C test items showed robust choice preferences, indicating strong blocking. In subsequent correlational analyses, the *DmBc* and *AmCc* choice preferences will be summed to form an overall measure of a participant's degree of blocking as indicated by choice.

Highlighting. The attentional hypothesis relies on the assumption that the later-trained I.PL→L items are actually learned. If they are not learned well, then there is no basis for suggesting that an attentional shift has been learned. Outlying individuals were culled on the basis of total accuracy on the I.PE and I.PL training items in the test phase. Only one outlier was identified, leaving 36 participants contributing data to subsequent analyses.

In the test phase, there were four trials each of I.PL and I.PE (see Table 3). Accuracy on these items was fairly high: For I.PE, mean accuracy across the 36 participants was 93.8% correct, and for I.PL, the mean accuracy was 91.0%.

For a given participant, there were eight *I* test trials (one half of which had *I* on the left and the other half of which had I on the right; see Table 3). For each participant we counted the number of choices for the corresponding E outcome and subtracted the number of choices for the corresponding L outcome, and divided the difference by the number of I trials. We denote the resulting measure as *Ic,* where the suffixed *c* indicates choice preference. The mean choice for the corresponding early-trained outcome, E, was 69.1%, and for the late-trained outcome, L, was 16.7%. (The remaining 24.2% of responses were for the two remaining choices.) The mean difference, $Ic = 52.4\%$, was reliably greater than zero, $t(35) = 6.07$, $p < .001$.

For a given participant, there were eight PE.PL test trials (one half of which had PE on the left and the other half of which had PL on the left; see Table 3). On these PE.PL trials, outcome L was chosen 66.0% of the time, and outcome *E* was chosen 26.4% of the time. The difference is computed for each participant and denoted *PLmPEc*. The mean difference, $PLmPEc = 39.6\%$, was reliably greater than zero, $t(35) = 4.92$, $p < .001$.

In summary, both the I and PE.PL test items showed robust choice preferences, indicating strong highlighting. In subsequent correlational analyses, the *Ic* and *PLmPEc* values will be summed to form an overall measure of a participant's degree of highlighting as indicated by choice.

Eye Gaze

Having established that the choice data show substantial blocking and highlighting, we proceeded to examine the eye gaze data.

Blocking. The attention hypothesis suggests that attention to the blocked cue B should decrease, whereas attention to the blocking cue A should increase. The control cues, C and D, should have intermediate attention.

To test whether the blocked cue B gets less attention than the control cues, we would like to know if gaze time to the control cue D (or C) is greater than the blocked cue B. The corresponding average difference is $DmBg = [(D_L - B_R) + (D_R - B_L)]/2$. This value, *DmBg,* will be greater than zero if D is gazed at longer than B. For each participant, the mean difference $D_L - B_R$ was determined from eight trials: D1.B1, D1.B2, D2.B1, D2.B2, C1.B1, C1.B2, C2.B1, and C2.B2. The mean difference $B_L - D_R$ was computed from the corresponding eight trials with the left–right positions reversed. The average of those differences was then recorded for each participant as his or her value of *DmBg*. There was one outlier, leaving 40 participants included. The gaze time to D was significantly longer than the gaze time to B, with the mean $DmBg = 46.2$ ms, reliably greater than zero, $t(39) = 3.01$, $p = .005$.

To test whether cue B gets less looking time than A, we computed the average difference, $AmBg = [(A_L - B_R) + (A_R - B_L)]/2$, in a manner analogous to *DmBg*. Because A.B was a training case, the analysis was restricted to trials that had a correct response (although the conclusion remains the same even if all responses are included). There were three outliers. The gaze time to A was longer than the gaze time to B, with the mean $AmBg = 57.2$ ms, reliably greater than zero, $t(37) = 2.11$, $p = .041$.

It was also found that looking time to A exceeded looking time to C (or D). We defined $AmCg = [(A_L - C_R) + (A_R - C_L)]/2$ analogous to the previous scores. There were four outliers. The gaze time to A was longer than the gaze time to C, with the mean $AmCg = 50.5$ ms, reliably greater than zero, $t(36) = 3.37$, $p = .002$.

In summary, the eye gaze data are consistent with the prediction that one mechanism in associative blocking is learned attention: People learn to attend to the blocking cue A and to ignore the blocked cue B. In subsequent correlational analyses, the two gaze differences that involve blocked cue B, namely *DmBg* and *AmBg* will be summed to form an overall measure of a participant's degree of blocking as indicated by gaze.

Highlighting. For highlighting, the attention hypothesis suggests that attention to PL should be stronger than to PE or I. If people tend to look longer at what they are cognitively attending to, then they should tend to look longer at PL than at PE.

The most straightforward way to test this prediction is by considering looking times on PL.PE test trials. We defined $PLmPEg = [(PL_L - PE_R) + (PL_R - PE_L)]/2$ analogous to the scores defined in the blocking analysis. Of 36 participants, there were 2 outliers. The gaze time to PL was longer than the gaze time to PE, with the mean $PLmPEg = 58.4$ ms, reliably greater than zero, $t(33) = 2.48$, $p = .019$.

Gaze times to PL and PE can also be compared across PL.I and PE.I trials. The prediction is that for PL.I and PE.I trials, that is, when the perfect predictors are on the left, there should be longer looking at PL than at PE. That should also be true when the perfect predictors are on the right, that is, for I.PL and I.PE trials. We combine those cases into an average difference, $PL.ImPE.Ig = ([(PL_L - I_R) - (PE_L - I_R)] + [(PL_R - I_L) - (PE_R - I_L)])/2$. Because these are cases of training items, only correct responses were included in the analysis. There were no outliers but one participant happened to have no correct responses for one of the four summands, and was therefore excluded. The gaze time to PL was longer than the gaze time to PE, with the mean $PL.ImPE.Ig = 90.2$ ms, reliably greater than zero, $t(34) = 2.08$, $p = .045$.

In summary, the eye gaze data support the claim that one mechanism in highlighting is learned attention: People learn to attend to the later-learned distinctive cue, PL. In subsequent correlational analyses, the two gaze differences, *PLmPEg* and *PL.ImPE.Ig,* will be summed to form an overall measure of a participant's degree of highlighting as indicated by gaze.

Individual Differences and Covariation

There is variation across individuals in the degree of choice or gaze preference. One possibility is that this variation might be entirely noise, with all individuals having an equal underlying preference. On the other hand, the variation might be caused, to some extent, by individual differences in the underlying preference. In particular, some people might have more rapid or extensive attentional shifting and learning than other people. This hypothesis of individual differences in attentional shifting and learning leads to two predictions: First, within a design (i.e., blocking or highlighting) people's choice preferences and gaze differences should covary. Second, across the highlighting and blocking designs, people should covary in the magnitude of blocking and highlighting.

We pursue these predicted correlations in three steps. First we show that within each type of experiment (blocking or highlighting), choice preferences for different test cues did indeed correlate with each other, and gaze preferences for different test cues also correlated with each other. Second, within each type of experiment, choice preferences were correlated with gaze preferences. Third, across the experiments, degree of blocking was correlated with degree of highlighting, for both choice and gaze measures.

Consider the blocking experiment. For the choice data, the degree of control-cue preference on DB trials (*DmBc*) strongly correlated with the degree of blocking-cue preference on AC trials (*AmCc*), $r = .522$, $t(39) = 3.82$, $p < .001$. For the gaze data, the correlation of *AmBg* with *DmBg* was $r = .423$, $t(35) = 2.76$, $p = .009$. The high correlation of *AmBg* with *DmBg* across individuals suggests that people who ignored B in one context also tended to ignore B in the other context.

As an overarching measure of an individual's blocking manifested in choice preference, we computed the sum of DB and AC preferences ($DmBc + AmCc$). As a summary measure of an individual's blocking manifested in gaze preference, we computed the sum of *AmBg* and *DmBg*. These measures of blocking in choice and blocking in gaze were strongly correlated, $r = .481$, $t(35) = 3.24$, $p = .003$. Thus, to the extent that gaze preference indicates differential attention, and gaze preference is correlated with blocking as measured by choice, we have evidence that blocking involves differential attention. Individuals who showed stronger gaze preferences tended to show stronger blocking.

Now consider the highlighting experiment. For the choice data, the degree of L-response preference on PE.PL trials (*PLmPEc*) strongly correlated with the degree of E-response preference on I trials (*Ic*), $r = .437$, $t(34) = 2.84$, $p = .008$. This correlation is predicted by attentional theory, because the stronger the association is from I to E, the stronger should be the attentional shift away from I during learning of I.PL. Conversely, the stronger the attentional shift away from I during learning of I.PL, the better preserved is the association from I to E. For the gaze data, the gaze differences *PLmPEg* and *PL.ImPE.Ig* also had a reliable positive correlation, $r = .369$, $t(31) = 2.21$, $p = .035$.

As a summary measure of an individual's highlighting manifested in choice, we summed the PE.PL and I choice preference magnitudes (*PLmPEc* + *Ic*). As a summary measure of an individual's highlighting manifested in gaze, we summed the *PLmPEg* and *PL.ImPE.Ig* gaze differences. These two measures were positively correlated, as predicted, but the trend was only marginally significant statistically, $r = .314$, $t(31) = 1.84$, $p = .075$ (two-tailed). Thus, individuals who showed stronger highlighting in their choice data tended to show stronger differential looking in their eye gaze data. Again, to the extent that eye gaze indicates attention, we have evidence that highlighting involves attention.

Finally, consider correlations across blocking and highlighting. For the choice data, degree of blocking (i.e., *DmBc* + *AmCc*) correlated with degree of highlighting (i.e., *PLmPEc* + *Ic*), $r = .382$, $t(32) = 2.34$, $p = .026$. For the gaze data, degree of blocking (i.e., *AmBg* + *DmBg*) correlated with degree of highlighting (i.e., *PLmPEg* + *PL.ImPE.Ig*), $r = .385$, $t(27) = 2.17$, $p = .039$. In other words, people who showed stronger blocking also tended to show stronger highlighting, as measured either by choice or by eye gaze.

Modeling

The empirical results reported above are naturally predicted by attentional learning theory, but until this point in this article the theory has been only vaguely stated in informal language. A rigorous pursuit of theoretical issues behooves us to ask two further questions. First, can an attentional learning theory account for the data when it is thoroughly specified in formal detail? Second, are there other models that could account for the results? The goal of this section is to answer those questions: Yes, a formal attentional learning theory can account for the data. No, some other likely models do not so readily account for the results.

A central theme of the modeling is that individual differences can be captured by parameter value differences. That is, all individuals are assumed to be describable by the same underlying representations and processing in the model; what varies between individuals are the specific values of parameters that control their specific quantitative behaviors. Thus, a successful model should do three things: (a) generate the blocking and highlighting effects in choice responses, (b) generate the blocking and highlighting effects in eye gaze, and (c) generate the correlations of blocking and highlighting, in choice and eye gaze, across individuals.

The EXIT Model

The EXIT model is one formal implementation of attentional learning theory. It has been thoroughly detailed elsewhere (Kruschke, 2001a, 2001b) and therefore will be described only briefly here. EXIT is a connectionist model that represents each cue as an input node that has zero activation when the cue is absent and positive activation when the cue is present. Each cue node is multiplied by a nonnegative attention strength. By default, any present cue gets some attention. The attention on the cues is hypothesized to have limited capacity, and therefore the cues compete for attention. The attentionally gated cue activations are propagated across weighted connections to the output nodes, each of which represents a possible response. The output activations are mapped to choice probabilities such that the responses corresponding to the more highly activated output nodes are given a higher probability. In summary, when a stimulus is presented to the network, the corresponding cue nodes are activated, attention is distributed across the cues, the attentionally gated cue activations are spread to the output nodes, and responses are made probabilistically, corresponding to the relative activations of the output nodes.

When corrective feedback is supplied for a stimulus (just as feedback is supplied to people in the learning experiments), 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. The first action it takes to reduce the error is a shift of attention across the input nodes. Attention is shifted away from cues that cause error and toward cues that either reduce error or at least do not increase error.

After attention has been shifted, the network then adjusts its associative weights. One set of weights to be adjusted connects the (attentionally gated) cues to the outputs. These weights are adjusted by simple error-reduction learning, as in standard back propagation and the Rescorla–Wagner model. Weights from attended-to cues are adjusted more than ignored cues, by virtue of the attentional gating.

The network does not merely learn what overt response to make for the cues. It also learns what covert attentional shift to make across the cues. Recall that the network's first reaction to error is to shift attention away from error-causing cues. The network should learn to reproduce this shifted pattern of attention in the future, so that it does not generate the error again. In EXIT, this learning of attentional distributions is accomplished by connections between the cues and the attentional gates. In fact, there is a set of exemplar nodes that encode configurations of co-occurring cues, and the exemplar nodes are connected to the attentional gates. The exemplar-mediated mapping from cues to attention gates allows the network to learn exemplar-specific distributions of attention. For example, in the context of the highlighting experiment, the network can learn to retain some attention on cue I for exemplar I.PE, but the network can learn to shift attention away from cue I for exemplar I.PL. Thus, after attention has shifted, the associative weights from the exemplar nodes to the attention nodes are adjusted, such that the shifted attentional distribution is better evoked by that exemplar in the future.

Parameters related to attention. There are a number of parameters that govern the specific quantitative behavior of EXIT. Three of these parameters are primarily concerned with attention. One such attentional parameter is the attentional shifting rate (λ_g in Kruschke, 2001a, Equation 7, p. 1400), which determines how large a shift is made in response to error. A second attentional parameter is the attentional learning rate (λ_x in Kruschke, 2001a, Equation 9, p. 1400). This parameter determines how large an adjustment is made to the weights from exemplars to attention nodes. A third attentional parameter is the exemplar node specificity (*c* in Kruschke, 2001a, Equation 3, p. 1399). This parameter governs how much the learned attentional distribution generalizes from one exemplar to another. The larger the specificity, the less the learned attention generalizes. For example, when the network learns for exemplar I.PL to shift attention toward PL, the specificity parameter determines how much that learning will generalize to the test probe PE.PL. In the simulation results reported below, the three attentional parameters will be yoked into a single factor because their effects on the model's behavior are tightly linked. Enhancement of attentional influence is produced by increasing the attentional shift rate, increasing the attentional learning rate or decreasing the specificity of the learned attention, or both.

Parameters not directly related to attention. There are two parameters that primarily govern the overall rate of learning associations from cues to outcomes. One of these parameters is, of course, the learning rate for the weights connecting cues to outputs (λ_w in Kruschke, 2001a, Equation 8, p. 1400). The second parameter is the capacity (*P* in Kruschke, 2001a, Equation 5, p. 1400). This parameter determines the total attentional weighting that can be allocated across cues. Essentially, when the capacity is high, there is more attentional multiplication overall, and overall learning is faster. The attentional capacity also has an influence on the degree of competition between cues, but, at least in the present experimental designs, the parameter appears to have an influence very similar to the learning rate on the output weights.

Another parameter in EXIT determines the decisiveness of the mapping from output activations to choice probabilities (ϕ in Kruschke, 2001a, Equation 2, p. 1399). Suppose that one output node has activation of .7 and another output node has activation of .3. A highly decisive network would assign a high probability to the first response and a low probability to the second response. A weakly decisive network, on the other hand, would assign less extreme probabilities to the two responses. It is important to notice that this decisiveness parameter can greatly influence the magnitude of choice probabilities generated by the network, but it has no influence whatsoever on learning or attention shifting. The decisiveness parameter merely governs the back end of the network that maps network behavior to human choice data. The decisiveness parameter does not directly influence the internal workings of the network.

A final parameter is the context node salience (σ in Kruschke, 2001a, Equations 3 and 4, p. 1399). In every stimulus, it is assumed that there is a shared context cue (e.g., the response prompt that occurs with every stimulus in the experiment). This cue can be useful for learning differential base rates of outcomes, but in the present experimental designs it is essentially inconsequential, and so the best fitting salience turns out to be close to zero.

Fit of EXIT

The EXIT model was formulated to generate choice probabilities. It was never explicitly meant to predict eye gaze. Nevertheless, we will cautiously make the assumption that attention allocated to a cue generates eye gaze to that cue. Thus, to get a rough-and-ready prediction of eye gaze from the model, we will assume that the relative attention across cues predicts the relative amount of eye gaze across cues.

In the Results section, we reported the absolute difference in gaze durations. For example, in the blocking paradigm we defined the difference between duration of gaze at the control cue, D, and duration of gaze at the blocked cue, B, as $DmBg = [(D_L - B_R) + (D_R - B_L)]/2$. We now simply convert that difference to a relative measure, $DmBg^* = [(D_L - B_R)/(D_L + B_R) + (D_R - B_L)/(D_R + B_L)]/2$. The value of $DmBg^*$ can range from -1 to $+1$. If a person gazes exclusively at *D* and never at *B,* then $DmBg^* = 1.0$. If a person gazes at *D* 55% of the time at at *B* 45% of the time, then $DmBg^* = 0.10$.

Across participants, the means of the relative gaze durations were $DmBg^* = 0.0224$, $AmBg^* = 0.0808$, $PLmPEg^* = 0.0489$, and $PL.ImPE.Ig^* = 0.0511$.[2] We shall attempt to fit these empirically measured values with the corresponding attentional differences in the model. For example, on a D.B test trial in the blocking experiment, the model's attention to cue D is D_α and the attention to cue B is B_α, and the relative attention is $DmBg^* = (D_\alpha - B_\alpha)/(D_\alpha + B_\alpha)$. That attention ratio in the model will be directly fit to the corresponding empirical gaze ratio, $DmBg^*$. Notice that there are no additional parameters introduced to the model in making this direct mapping from model attention ratios to gaze duration ratios. This mapping is, no doubt, incomplete and quantitatively lacking, but it turns out to be good enough for reasonable fits.

Other mappings from model attention to gaze ratios are possible. For example, when the stimuli appear, there might be an initial duration in which both cues are looked at and encoded (because their positions are random) regardless of which cue will subsequently in the trial get more sustained gaze because of learned relevance. In this case, the gaze ratios might instead correspond to $(D_\alpha - B_\alpha)/(2E + D_\alpha + B_\alpha)$, where *E* reflects the time for initially encoding a cue. Other measures of relative gaze might also be used, such as the overall priority of looking at one cue or the other within the sequence of fixations. Such a measure might be especially appropriate if there were a process model describing eye movements as a function of learned relevance of cues. After we completed our modeling, we noted that Rehder and Hoffman (in press) reported a linear correspondence between model attention values and gaze preferences. Merely for simplicity, however, we will identify model attention ratio with empirical gaze ratio.

We will fit the EXIT model to overall mean data from all of the participants and then explore how the model behavior changes when parameter values deviate from the best fitting values. The idea is that the parameter values that fit the overall mean might describe an average participant, and variations from those central parameter values would yield predictions of variations across individuals. In particular, we are specifically interested in knowing whether variations in attentional parameters generate correlated variations in blocking and highlighting, as predicted by the informal theory that motivated the experiments and as suggested by the data analyses reported above.

[2] Using the proportional measure of gaze time, all of the results remained statistically reliable except a subset of those involving *PL.ImPE.Ig*,* which had a relatively large variance. All of the trends remained the same as those reported for the difference measure.

For the blocking design (see Table 2), the model was trained on four blocks of Phase 1 and four blocks of Phase 2, because these amounts of training approximate the mean blocks to criterion observed in human participants (3.78 and 4.02 blocks, respectively). The model was trained on six blocks of Phase 1 in highlighting (see Table 3), reflecting the fact that human participants took more than five blocks on average to reach criterion (mean of 5.24 blocks). Stable model predictions were achieved by running 20 simulated participants, each with a different random ordering of training trials within blocks.

A single set of parameter values was used for all simulated participants and simultaneously for both blocking and highlighting designs. (The model can fit the data even better when the blocking and highlighting experiments are fit separately.) Predictions of choice and gaze were computed for each simulated participant, and the mean model prediction was compared against the mean human prediction using root mean squared deviation (RMSD) as a measure of discrepancy. In more detail, the model attempted to fit 12 means from the test phases of the two experiments. These data were the accuracies on the training items A.B, C.D, I.PE, and I.PL, the choice differences DmBc, AmCc, Ic, and PLmPEc, and the proportional gaze differences DmBg*, AmBg*, I.PLmI.PEg*, and PLmPEg*. The parameter space was searched until a best fit was found. We used a simplex hill-climbing method, starting from several different initial parameter values. Unfortunately, the parameter space seemed to be crenulated with many local minima, so the fit we report might not be the best possible.

The predicted means from the best fitting parameter values are shown in Figure 2, along with model predictions when the attentional parameters are varied above and below the best fitting values (BFV). The *x*-axis of Figure 2 indicates the magnitude of the attentional parameters, as a proportion of the BFV. When this proportion is 1.0, the parameters are set at the BFV; thus, the model predictions plotted over the *x*-axis value of 1 are the best fitting predictions. Additional discussion of the *x*-axis appears below.

The upper panels of Figure 2 plot predicted response probability or difference of response probability. For example, the curve labeled *A.B* in the upper left panel indicates the predicted accuracy on item A.B in the test phase. The curve labeled *AmCc* indicates the predicted value of the choice *difference* AmCc (defined in the *Results* section). The lower panels plot the attention ratios in the model, which are identified with gaze ratios in the data. For example, the curve labeled *AmBg** in the lower left panel indicates the model's predicted attention ratio when item A.B is presented in the test phase.

The EXIT model attained a good quantitative fit. (The best fitting RMSD was 0.032, for the following parameter values: attentional shift rate 0.278, gain weight learning rate 0.0418, exemplar specificity 0.300, output weight learning rate 1.00, attention capacity 1.03, choice decisiveness 3.57, and context node salience 0.000.) The model shows high accuracy on the training items (curves A.B and C.D in the upper left panel of Figure 2, and curves I.PE and I.PL in the upper right panel of Figure 2), strong blocking in choice (curves DmBc and AmCc well above zero in the upper left panel of Figure 2), sizable highlighting in choice (curves PLmPEc and Ic well above zero in the upper right panel of Figure 2), notable gaze differences in blocking (curves DmBg* and AmBg* above zero in the lower left panel of Figure 2), and significant gaze differences in highlighting (curves PLmPEg* and I.PLmI.PEg* above zero in the lower right panel of Figure 2).

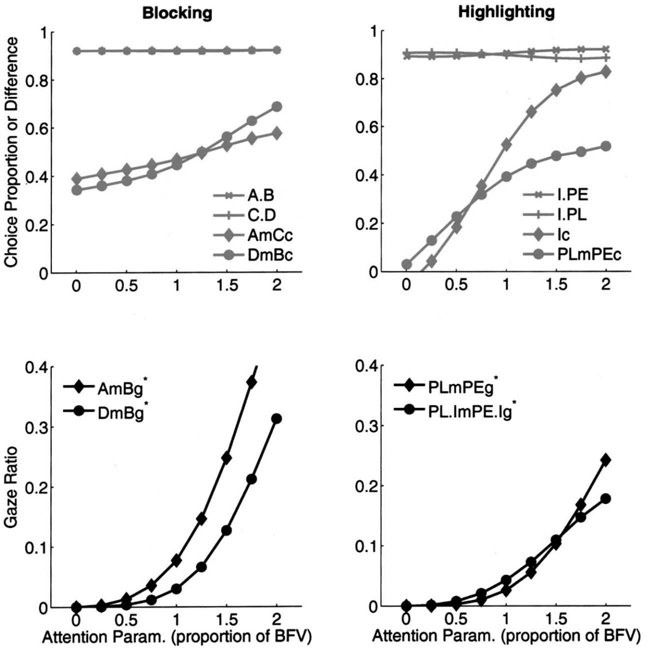


Figure 2. Predictions of the EXIT model for the blocking experiment (left panels) and highlighting experiment (right panels), when fit simultaneously to data from both. The values on the *x*-axes denote the values of the three attentionally related parameters in EXIT as a proportion of the best fitting values (BFV). The upper panels plot choice proportions (for A.B, C.D, I.PE, and I.PL) or differences of choice proportions (for AmCc, DmBc, Ic, and PLmPEc). The lower panels plot attention ratios in the model, which were fit to human gaze ratios. Param. = parameter.

Figure 2 also displays how the predictions of EXIT change when the attentional parameters are changed away from their BFV. The motivation for this exploration is the hypothesis that different individuals have different attentional characteristics. Some individuals have attentional shifting, learning, and generalization that is higher than average, whereas others have attentional shifting, learning, and generalization that is lower than average. Attentional theory suggests that people with higher attention shifting, learning, and generalization should exhibit larger blocking and highlighting in choices and in gaze. The main point of Figure 2 is to illustrate that EXIT produces exactly these trends as a function of attention.

The *x*-axis of Figure 2 indicates the values of the three attentional parameters as proportions of the BFV. For example, when the proportion is $x = 2.0$, then the attentional shifting and learning rates are set to twice their BFV, and when the proportion is $x = 0.0$, then the attentional shifting and learning rates are set to zero. The direction of the specificity is reversed on the *x*-axis, because increasing x represents increasing attentional generalization (i.e., decreasing attentional specificity). Thus, for the specificity, the proportional multiplier on the specificity is actually $2 - x$. For example, when $x = 2.0$, the specificity is set to zero, and when $x = 0.0$, the specificity is set to twice its BFV.

The mutually ascending curves in the four panels of Figure 2 show clearly that as the attentional parameters increase, the magnitudes of blocking and highlighting increase, and the gaze differences also increase. Although not displayed here, these trends also occur when each of the three attentional parameters is adjusted separately, with all other parameters held constant. The three parameters are adjusted together here because they all express the strength of the attentional system. The three parameters do have identifiable effects on behavior, however. Attentional shifting, not necessarily accompanied by attentional learning (or generalization of that learning), is critical to generate highlighting per se. Learning the shift is helpful for accelerating performance on the I.PL training trials (Kruschke, 2003b). Generalization of the learned attentional shifts is important for producing attentional (i.e., gaze) differences for novel test stimuli, such as PL.PE in highlighting and B.D in blocking.

The trends shown in Figure 2 naturally imply the correlations across participants that were observed in the human data. Imagine simulated participants that vary individually in the strength of their attentional systems. A simulated individual with attentional parameters set at $x = 0.8$ will have smaller-than-average blocking, highlighting, and gaze differences. A simulated individual with attentional parameters set at $x = 1.2$ will have larger-than-average blocking, highlighting, and gaze differences. Therefore, across simulated participants who vary in attentional strength, the magnitudes of blocking, highlighting, and gaze differences will be positively correlated. Notice that this is true even when the other parameters (the decisiveness, the associative learning rate, the capacity, and the context salience) are held constant.

Variations in single other parameters do not give rise to the correlations observed in the data. When the decisiveness is varied, there are large changes in blocking and highlighting choice proportions, but zero changes in eye gaze because the decisiveness parameter has no influence on attention or learning. The context salience is virtually zero in the best fits to the present data, and so even doubling it causes essentially no changes in the predictions. Changing the output weight learning rate (within ±0.5 of the BFV) produces *negative* correlations of blocking (DmBc and AmCc) with gaze (DmBg* and AmBg*). The same is true of changing the capacity parameter. It is conceivable that some sort of correlated variation among the other parameters could generate the types of correlation seen in human subjects, but if it exists it is not obvious to us.

Figure 2 also graphically illustrates that attentional shifting is critical in EXIT for producing highlighting, but not for producing blocking. When there is no attentional shifting (i.e., when $x = 0.0$), then the gaze ratio drops to zero because attention is never shifted from its initial uniform distribution. Highlighting also drops to zero (i.e., the Ic and PLmPEc curves are essentially at zero).[3] The Rescorla–Wagner model is a special case of EXIT in which there is zero attention shifting (and very large attention capacity), and thus it cannot exhibit highlighting.

Blocking persists, however, when $x = 0.0$ (i.e., the DmBc and AmCc curves are well above zero). Blocking occurs when $x = 0.0$ because the associative weights are adjusted by error, just as in the classic Rescorla–Wagner model. Attentional shifting in EXIT merely enhances blocking, but attention shifting is not the sole cause of blocking. We speculate that this property of attention in EXIT might mimic behavior across species or developmentally: Blocking might occur without highlighting in individuals who have small or no attentional shifting, such as in human youngsters or in some nonhuman species (Fagot, Kruschke, Depy, & Vauclair, 1998). Although attentional learning is not needed to account for blocking per se, attentional learning is needed to account for retarded subsequent learning about a previously blocked cue (Kruschke & Blair, 2000; Kruschke, 2001b, 2005).

Other Models

It is not our purpose here to claim that no other model can account for our data. We examine a few alternative models that are candidates because of their application to blocking, highlighting, or gaze by previous authors. These models in their published forms cannot account for our results, but modified versions might.

The Rescorla–Wagner model. The Rescorla–Wagner model is the classic model of blocking. It is a special case of EXIT when EXIT's attentional parameters are set to zero (and its capacity parameter is set to a very large value). As pointed out above in conjunction with Figure 2, the Rescorla–Wagner model cannot account for highlighting. Perhaps some modified version that encodes absent-but-expected cues as negative values (e.g., Markman, 1989; Tassoni, 1995; Van Hamme & Wasserman, 1994) could account for highlighting. Such an approach would have to specify how the model learns to expect cues and would need to specify how associative weights are mapped to eye gaze.

The eliminative inference model (ELMO). An alternative model that has been claimed to account for highlighting is ELMO, proposed by Juslin et al. (2001). As described in the beginning of this article, ELMO fails to account for some effects observed in highlighting experiments (Kruschke, 2001a). For the moment, we will overlook those shortcomings and investigate whether ELMO might address the results of the present article. Unfortunately, we find that ELMO cannot generate a blocking effect. In the blocking procedure (see Table 1), cases of A.B→X and C.D→Y occur equally often, and therefore would be equally well learned by ELMO. In the test phase, when probe B.D is presented, the A.B→X and C.D→Y rules would be equally evoked, and so the X and Y responses would have equal probability.

The RULEX model. The RULEX model (Nosofsky et al., 1994) was considered by Rehder and Hoffman (2005) in their study of gaze and category learning. The RULEX model is therefore also a candidate for the present research on gaze and associative learning. To our knowledge, RULEX has never been applied to blocking or highlighting. It is far beyond the scope of this article to embark on a lengthy computer simulation study of RULEX; we limit ourselves here to intuitive simulation and tentative conclusions.

RULEX posits that people try to explain category labels using rules that are as simple as possible. Rules that involve single

[3] The Ic and PLmPEc curves are not exactly at zero when $x = 0$ because the output-weight learning rate is not high enough in these simulations for the later I.PL training to completely balance the earlier I.PE training. Additional simulations confirm that when the output weight learning rate is increased, then the Ic and PLmPEc curves go to exactly zero when $x = 0$. Notice that the output-weight learning rate is not limited at a value of 1.0. It is easy to prove that in a linear associator, the error can be reduced to zero in one trial if the learning rate is set at $1/\Sigma_i\ \alpha_i^2 a_i^2$, where α_i is the attention on cue i and a_i is the activation of cue i. For the other parameter values set at their BFV, this implies that the output learning rate has a maximum well over 2.0 for I.PE and I.PL stimuli.

features or dimensions are simplest and therefore tested first. Next simplest, usually, are single-feature rules with a small number of exceptions specified by a small number of additional features. Conjunctive rules involving as few features as possible are usually considered next. Among the free parameters of RULEX is a *branching* parameter that specifies the probability of testing imperfect single-feature rules before testing conjunctive rules. RULEX retains a rule as long as it makes enough correct predictions (according to a criterion that is a parameter of the model), otherwise the rule is jettisoned and a different rule is tested. In this way RULEX is error-driven, as is the Rescorla–Wagner model and the EXIT model. Because RULEX encodes only a small number of features whenever possible, it is also attentionally selective, like EXIT.

When applied to blocking (see Table 1), RULEX would learn in Phase 1 the rule A→X. In Phase 2, that rule would correctly predict the outcome of cases A.B→X, and therefore would be retained, and no new rule involving B would be learned. When learning cases of C.D→Y, RULEX would learn one of the single-feature rules, C→Y or D→Y, but presumably not both. Finally, when confronted with test case B.D, RULEX would do one of two things. If it had learned D→Y in Phase 2, then this rule would be matched and RULEX would respond with Y. If it had learned C→Y in Phase 2, then RULEX would have no known rules involving B or D, and it would simply choose randomly from the response options. Analogous reasoning applies to test case A.C. Thus, on average, RULEX would exhibit blocking in its choice predictions.

Rehder and Hoffman (2005) hypothesized that if RULEX had eyes, it would look at the features in its learned rules more than at features that were not in its learned rules, because only the features in its learned rules are informative for making responses. From this hypothesis mapping covert rules to overt gaze, it follows that RULEX would also show blocking in gaze (i.e., look more on average at control cues C and D than at blocked cue B, and look more at the blocking cue A than at the control cues).

When applied to highlighting (see Table 1), the predictions of RULEX are more complicated. In the early training phase, RULEX would learn from training case I.PE→E either the rule I→E or the rule PE→E. Suppose RULEX learned PE→E. In the later learning phase, RULEX might test the rule I→L but find it imperfect, and therefore settle on PL→L. With these rules (PE→E and PL→L), RULEX would not show highlighting. Suppose, however, that RULEX learned I→E in the early training phase. In the later training phase, RULEX could jettison that now-imperfect early rule and learn the rules PE→E and PL→L. Again in this case it would not show highlighting. Alternatively, with a probability governed by a free parameter, RULEX could retain the early rule I→E and learn I.PL→L as an exception. In this case, RULEX would show highlighting. In summary, RULEX predicts that at least one half of the participants would learn the rules PE→E and PL→L, and therefore would not show highlighting, but at least some of the remaining half of the participants would learn the rules I→E and I.PL→L, and therefore would exhibit highlighting.

Thus, RULEX could qualitatively produce the basic blocking and highlighting effects. Unfortunately, RULEX might suffer when trying to simultaneously fit data from other probe items. For example, in highlighting, the test phase can include the case of PE by itself. People respond E in this case at very high rates (e.g., 92.5% in Experiment 2 of Kruschke, 1996). For RULEX to exhibit strong highlighting, a sizable proportion of its simulated participants do not know anything about PE, and therefore RULEX would probably not show high response rates for PE.

It is also not clear how RULEX would show correlations in blocking and highlighting. Whereas the degree of highlighting would be governed by the degree to which people look for exceptions to rules (rather than jettisoning imperfect rules), it is not clear how that would govern the degree of blocking. Indeed, the blocking procedure is so simple that it is unclear how RULEX would show any variance at all, except through its response-error parameter, which governs the extent to which individuals make a random, unintended response.

Despite these potential problems, it is likely that a reasonable variant of RULEX could account for the blocking and highlighting data we have presented in this article. If so, the essential explanatory principles in RULEX are fundamentally similar to those in EXIT: Cues are selectively attended to, and the selection of cues is driven by error reduction.

Variants of ALCOVE. The ALCOVE model (Kruschke, 1992) learns to differentially attend to psychological stimulus dimensions depending on their relevance to the category label, and the rapid attention shifts ‘n’ learning (RASHNL) model (Kruschke & Johansen, 1999) is an extension of ALCOVE that incorporates rapid shifts of attention (in contradistinction to the presumably more gradual shifts in ALCOVE). Both models assume that what is attended to is dimensions, as opposed to values within dimensions, whereas the CORNER model (Kalish & Kruschke, 2000) also selectively attends to values within dimensions. None of these models is specifically designed to address traditional blocking or highlighting procedures, however, which rely on the presence or absence of cues (cf. Kalish, 2001). Verguts, Ameel, and Storms (2004) incorporated into ALCOVE a different similarity function that can accommodate present and absent features, in a model they called *additive ALCOVE* (ADDCOVE). The ADDCOVE model has no attention-shifting mechanism, however, and therefore cannot account for highlighting. Perhaps future work could combine the similarity function of ADDCOVE with the rapid shifts of attention in RASHNL or EXIT to address the results reported in the present article.

Summary and Discussion

Previous research using choice data alone has suggested that highlighting and blocking involve learned shifts of cognitive attention. If people tend to look at what they are cognitively attending to, then they should tend to look more at highlighted cues and less at blocked cues. These predictions were confirmed.

To the extent that attentional shifting and learning vary across individuals and are stable individual differences, there should be covariation of attention-dependent behaviors across individuals. If attentional shifting and learning is an important factor in blocking, and if eye gaze is linked with cognitive attention, then individuals who show stronger blocking in choice should show stronger differences in eye gaze. This predicted correlation was confirmed. Analogously, if attentional shifting and learning is an important factor in highlighting, and if eye gaze is linked with cognitive attention, then individuals who show stronger highlighting in choice should show stronger differences in eye gaze. This predicted correlation was also confirmed (although marginal two-tailed). If attentional shifting and learning is an important factor in *both* blocking and highlighting, and if eye gaze is linked with cognitive attention, then individuals who show larger gaze differences in blocking should tend to show larger gaze differences in highlighting. This predicted correlation was confirmed. Finally, if attentional shifting and learning is an important factor in both blocking and highlighting, then individuals who show stronger blocking (measured by choice) should tend to show stronger highlighting (measured by choice). This predicted correlation was also confirmed. Note that this last prediction does not depend on eye gaze.

A formal model of attentional shifting and learning was fit simultaneously to choice and gaze data from blocking and highlighting experiments. The model could also naturally account for correlations of blocking, highlighting, and gaze by varying the attentional parameters across simulated participants. Other parameters in the model, such as choice decisiveness or learning rate, do not easily show such correlations. Other candidate models from the literature cannot account for the findings, but perhaps those models could be modified in future research.

As a whole, this research adds significant new evidence to the argument that highlighting and blocking involve learned attention. Not only do people tend to look less at blocked cues and more at highlighted cues, but covariation across people suggests enduring individual differences in attentional shifting and learning.

Attention in EXIT and Other Formalizations

Attention means different things to different researchers, depending in part on what is conceived to be the recipient of attention. Attention can be allocated to spatial locations, to objects, to features or dimensions (either spatially localized or spread across space and time), to response options, to outcomes, or to discrepancies between actual and predicted outcomes. Attention also means different things depending on its role in processing. Attention can refer to an enhanced effect on immediate generation of a response, or it can refer to enhanced learning that is only manifested in later behavior. In this section we briefly discuss what sort of attention is formalized in the EXIT model.

In EXIT (Kruschke, 2001a, 2001b), attention is formalized as multipliers on cue activations. The multipliers affect both immediate response generation and subsequent learning about the cues. Some previous theories have separated those functions; for example, Mackintosh (1975) emphasized attention to a cue as its learning rate or *associability,* distinct from its immediate impact on response generation. The associability of a cue increased if it was predictive of the outcome, and decreased otherwise. Kruschke (2001b) described formal relationships between Mackintosh's approach and the EXIT model. Pearce and Hall (1980) proposed an alternative perspective in which the associability of a cue is determined as the error it caused on its previous presentation. This approach implies that after sufficient learning, when there is little error, the associability of a cue is low. The different approaches address different sets of data, and a unified approach has yet to emerge. Pearce and Bouton (2001) provided a nice review of these issues. In EXIT, predictions of differential gaze come from the notion of attention as a multiplier for immediate response generation. It is not clear how the Mackintosh or Pearce and Hall approaches might be modified to address eye gaze (and, note that those approaches were designed with animal, not human, learning in mind).

The EXIT model attends to cues. In general, cues can be (circularly) defined as any psychological entity to which attention can be allocated. A cue can be a concept, a word, a color value (e.g., red), color as a dimension (as distinct from, say, the dimension of shape), objects, locations, and so on. Cues can be spatiotemporally localized or distributed. More specifically, the EXIT model attends to present and absent features (e.g., the word *ocean,* which can be present or absent). Other models, such as RASHNL (Kruschke & Johansen, 1999) and its predecessor ALCOVE (Kruschke, 1992), attend to dimensions (e.g., the dimensions of color vs. shape). The CORNER model (Kalish & Kruschke, 2000) extends the representation of EXIT to include continuous dimensions, so that the model simultaneously attends to dimensions and values within dimensions. Our aim here is simply to point out that these distinctions exist and to be clear that connections between these varieties of attention are not a foregone conclusion. In the present research, we relied on a hypothesized correspondence between (a) cognitive attention to a concept and (b) perceptual attention to the word designating that concept (although the location of the word varied across trials and people could not use location as an indicator of relevance, unlike the experiments of Rehder & Hoffman, 2005, in press). This correspondence does not preserve some aspects of attention in the two realms. For example, cognitive attention might be distributed across several concepts simultaneously, but eye gaze is directed at one location at a time.

Logan (2002) presented a unified view of attention to dimensions, response options, and objects, which he called the *instance theory of attention and memory* (ITAM; see also Logan, 2004). In ITAM, selective attention to an item in space is conceptualized as the same process as selective attention to a categorical response. The two selective processes take place simultaneously and in a unified formalization. This is unlike the approach taken by EXIT, in which competition between response options is distinct from competition between cues (or objects). ITAM learns by storing copies of whatever instances it has attended to, but it does not have a mechanism for shifting or learning attention analogous to EXIT. Perhaps an even broader unification of these attentional theories will emerge soon.

References

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye–mind hypothesis. *Psychological Science, 15,* 225–231.

Crookes, A. E., & Moran, P. M. (2003). An investigation into age and gender differences in human Kamin blocking, using a computerized task. *Developmental Neuropsychology, 24,* 461–477.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19–23.

Fagot, J., Kruschke, J. K., Depy, D., & Vauclair, J. (1998). Associative learning in humans (*Homo sapiens*) and baboons (*Papio papio*): Species differences in learned attention to visual features. *Animal Cognition, 1,* 123–133.

Gibbons, H., Rammsayer, T. H., & Lubow, R. E. (2001). Latent inhibition depends on inhibitory attentional learning to the preexposed stimulus: Evidence from visual search and rule-learning tasks. *Learning and Motivation, 32,* 457–476.

Juslin, P., Wennerholm, P., & Winman, A. (2001). High-level reasoning and base-rate use: Do we need cue competition to explain the inverse base-rate effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 849–871.

Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2002). Perspective effects on online text processing. *Discourse Processes, 33,* 159–173.

Kalish, M. L. (2001). An inverse base rate effect with continuously valued stimuli. *Memory & Cognition, 29,* 587–597.

Kalish, M. L., & Kruschke, J. K. (2000). The role of attention shifts in the categorization of continuous dimensioned stimuli. *Psychological Research, 64,* 105–116.

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 (Ed.), *Punishment* (pp. 279–296). New York: Appleton-Century-Crofts.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130,* 169–183.

Kappenman, E. S., Kruschke, J. K., & Hetrick, W. P. (2005). *Highlighting and illusory correlation in schizotypal personality disorder.* Manuscript in preparation.

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

Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science, 5,* 3–36.

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

Kruschke, J. K. (2001a). The inverse base rate effect is not explained by eliminative inference. *Journal of Experimental Psychology: Learning, Memory, and 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. (2003a). Attention in learning. *Current Directions in Psychological Science, 12,* 171–175.

Kruschke, J. K. (2003b). Attentional theory is a viable explanation of the inverse base rate effect: A reply to Winman, Wennerholm, and Juslin (2003). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1396–1400.

Kruschke, J. K. (2005). Learning involves attention. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (pp. 113–140). Hove, England: 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, and Cognition, 25,* 1083–1119.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review, 109,* 376–400.

Logan, G. D. (2004). Cumulative progress in formal theories of attention. *Annual Review of Psychology, 55,* 207–234.

Lohmeier, J. H. (1996). *The role of feature validity in categorization learning: Effects on attention to and access of featural information.* Unpublished doctoral dissertation, University of Massachusetts. (UMI No. AAM9606540)

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.

Medin, D. L., & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General, 117,* 68–85.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79.

Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology, 52,* 111–139.

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

Rehder, B., & Hoffman, A. B. (2005). Eye tracking and selective attention in category learning. *Cognitive Psychology, 51,* 1–41.

Rehder, B., & Hoffman, A. B. (in press). Thirty-something categorization results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

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.

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 & Motivation, 25,* 127–151.

Verguts, T., Ameel, E., & Storms, G. (2004). Measures of similarity in models of categorization. *Memory & Cognition, 32,* 379–389.

Webb, M. R., & Lee, M. D. (2004). Modeling individual differences in category learning. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th annual meeting of the Cognitive Science Society* (pp. 1440–1445). Mahwah, NJ: Erlbaum.

Young, M. E., & Wasserman, E. A. (2002). Limited attention and cue order consistency affect predictive learning: A test of similarity measures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 484–496.

Received February 5, 2005
Revision received April 20, 2005
Accepted April 20, 2005 ■