Journal of Experimental Psychology: General
2008, Vol. 137, No. 2, 226–243

0096-3445/08/$12.00 DOI: 10.1037/0096-3445.137.2.226

The Psychophysics of Contingency Assessment

Lorraine G. Allan, Samuel D. Hannah, Matthew J. C. Crump, and Shepard Siegel
McMaster University

The authors previously described a procedure that permits rapid, multiple within-participant evaluations of contingency assessment (the "streamed-trial" procedure, M. J. C. Crump, S. D. Hannah, L. G. Allan, & L. K. Hord, 2007). In the present experiments, they used the streamed-trial procedure, combined with the method of constant stimuli and a binary classification response, to assess the psychophysics of contingency assessment. This strategy provides a methodology for evaluating whether variations in contingency assessment reflect changes in the participant's sensitivity to the contingency or changes in the participant's response bias (or decision criterion). The sign of the contingency (positive or negative), outcome density, and imposition of an explicit payoff structure had little influence on sensitivity to contingencies but did influence the decision criterion. The authors discuss how a psychophysical analysis can provide a better understanding of findings in the literature such as mood and age effects on contingency assessment. They also discuss the relation between a psychophysical approach and an associative account of contingency assessment.

Keywords: contingency assessment, signal detection theory, associative learning, psychophysics, outcome density effect

There are many experiments in which a participant is asked to assess the contingency or causal relationship between two events, a cue and an outcome. Typically, a discrete trial format is used. On each trial, a cue may, or may not, be presented, after which an outcome may, or may not, occur. Various cues and outcomes have been used. The cue may consist of information that a hypothetical individual has or has not eaten shrimp, and the outcome may consist of information that the individual has or has not suffered an allergic reaction (Wasserman, 1990); the cue may consist of information that a fertilizer has or has not been applied to a plant, and the outcome may consist of information that the plant has or has not thrived (Spellman, Price, & Logan, 2001); the cue may consist of information that a potentially antibacterial chemical has or has not been infused into a bacterial culture, and the outcome may consist of information that the bacteria did or did not survive (Tangen & Allan, 2004).

Lorraine G. Allan, Samuel D. Hannah, Matthew J. C. Crump, and Shepard Siegel, Department of Psychology, Neuroscience, and Behaviour, McMaster University, Hamilton, Ontario, Canada.

Matthew J. C. Crump is now at Vanderbilt University.

Some of the data were presented in an invited address by Lorraine G. Allan at the 2006 meeting of the International Society of Psychophysics, which also appears in D. E. Kornbrot, R. M. Msetfi, & A. W. MacRae (Eds.), *Fechner Day 2006: Proceedings of the 22nd Annual Meeting of the International Society for Psychophysics*, published in 2006 by the International Society for Psychophysics, St. Albans, England. The preparation of this article was supported by research grants from the Natural Sciences and Engineering Research Council of Canada to Lorraine G. Allan and Shepard Siegel and by a Natural Sciences and Engineering Research Council of Canada Graduate Scholarship to Mathew J. C. Crump.

Correspondence concerning this article should be addressed to Lorraine G. Allan, Department of Psychology, Neuroscience, and Behaviour, McMaster University, Hamilton ON L8S 4K1, Canada. E-mail: allan@mcmaster.ca

More generally, the stimuli presented to a participant can be summarized as a 2 × 2 matrix (see Table 1). On each trial, the cue either is presented (C) or is not presented ($\sim C$), and then the outcome either does occur (O) or does not occur ($\sim O$). The letters in the cells (a, b, c, d) represent the joint frequency of occurrence of the four cue–outcome combinations in a block of trials. One commonly used measure of the contingency between the cue and the outcome is ΔP (Allan, 1980):

$$\Delta P = P(O|C) - P(O|\sim C) = \frac{a}{a+b} - \frac{c}{c+d}.$$

At the end of a series of trials consisting of many cue and outcome presentations, the participant is usually asked to rate the strength of the relationship between the two events. For example, on a 100-point scale, the participant must indicate the strength of the relationship between eating shrimp and the occurrence of an allergic reaction, or between applying fertilizer and the growth of the plant, or between infusing the chemical and the survival of the bacteria. Usually the rating is about the causal relationship between the cue and the outcome (e.g., "Rate the degree to which eating shrimp causes an allergic reaction") or about the contingency between the cue and the outcome (e.g., "Rate the strength of the association between eating shrimp and an allergic reaction").

There are many theoretical accounts of performance in the contingency assessment task (see De Houwer & Beckers, 2002). Some investigators have suggested that humans, and perhaps nonhuman animals, extract rules from the sequence of cue and outcome presentations; that is, participants are innate statisticians, computing and comparing the conditional probabilities of the outcome in the presence, and in the absence, of the cue (e.g., McKenzie & Mikkelsen, 2007; Peterson & Beach, 1967). Others have suggested that contingency assessment is really another instance of associative learning; that is, the cues and outcomes function as conditional and unconditional stimuli, respectively (e.g., Dickinson, Shanks, & Evenden, 1984). Finally, some have

Table 1
2 × 2 Matrix for the Cue–Outcome Pairings in a Contingency Assessment Task

	Outcome	
Cue	O	$\sim O$
C	a	b
$\sim C$	c	d

Note. The letters in the cells indicate the frequency of occurrence of each of the four cue (C)–outcome (O) combinations in a block of trials. $\sim O$ = absence of outcome; $\sim C$ = absence of cue.

suggested that contingency assessment can be understood as deductive reasoning; that is, organisms evaluate contingent relationships by applying rules of inferential logic (e.g., Beckers, De Houwer, Pineño, & Miller, 2005). Although these classes of models have been very useful in describing data from typical contingency tasks, we suggest that they do not address an important aspect of contingency assessments.

The contingency assessment task, as it typically is implemented in the laboratory, is quite different from the usual contingency assessment challenges faced by people. When we have experienced a series of cues and outcomes, we typically do not judge the statistical relationship between the events—rather, we make a categorical, and typically a binary, decision about the relationship: "Am I allergic to shrimp, or may I eat them with impunity?" "Is the fertilizer associated with plant growth or not associated with plant growth?" "Does the putative antibacterial chemical work or not?"[1] This is often the position faced by participants in a psychophysics task. They are commonly given different levels of a stimulus and must make a categorical decision, such as whether the signal is present or absent or whether it is loud or soft. Recently, we (Allan, Siegel, & Hannah, 2007; Allan, Siegel, & Tangen, 2005) and others (Perales, Catena, Shanks, & González, 2005) have discussed the value of a psychophysical analysis of contingency assessment (specifically, signal detection theory, or SDT; Green & Swets, 1966) to more accurately characterize contingency assessment.

SDT was developed to understand how organisms decide whether a signal (originally an auditory stimulus) has, or has not, been presented when there is uncertainty because the signal is presented in a noisy environment. Similarly, the participant in a contingency assessment task must make a judgment under conditions of uncertainty. Consider the typical experiment. The participant is passively exposed to a (usually lengthy) sequence of cue and outcome presentations and, at the end of the presentations, must rate the strength of the contingency. Such assessment will be uncertain because of "noise" caused by, for example, vagaries in the participant's memory of the events in the series and shifts in attention during the cue and outcome presentations.

Recasting the contingency assessment task as a signal detection task means that the participant's binary decision about the contingent relationship may be correct in one of two ways. In the shrimp example, the participant (a) could conclude that the hypothetical individual is allergic to shrimp, and in fact the sequence of cue and outcome presentations was such that the two events were positively correlated—a *hit*, in the language of SDT, or (b) could conclude that the hypothetical individual is not allergic to shrimp, following a noncontingent presentation of the events—a *correct rejection*. The participant also may be incorrect in one of two ways in the shrimp example. The participant (a) could conclude that the hypothetical individual is allergic to shrimp, following a noncontingent presentation of the events—a *false alarm*, or (b) could conclude that the hypothetical individual is not allergic to shrimp, when the cue and outcome presentations were, in fact, positively correlated—a *miss*.

On the basis of SDT, two separable factors contribute to the participant's response indicating whether or not a signal was presented. One factor is the participant's sensitivity to detecting the signal, which is determined by the magnitude of the signal and also by the participant's ability to process the signal. The second is the participant's response bias or decision criterion. The criterion parameter is determined by nonsensory or cognitive variables, such as the participant's analysis of the costs of making each of the two types of mistakes (false alarms and misses). Such an analysis may bias the participant toward certain conclusions concerning the presence or absence of a signal. An individual incorrectly told that he is allergic to shrimp might needlessly avoid eating a tasty food. That is not good, but it is probably not as bad as suffering a severe allergic reaction as a result of being incorrectly informed that he is not allergic to shrimp. Thus, a participant experiencing a sequence of events that includes eating shrimp and suffering an allergic reaction may reasonably be biased toward concluding that he is allergic to shrimp. According to SDT, the participant's response is determined by the independent actions of sensitivity and criterion. The primary distinction between SDT and all alternative accounts of contingency assessment is that only SDT incorporates these two distinct processes in determining the participant's behavioral response.

There are several challenges to evaluating a SDT analysis of contingency assessment. One is the need for a categorical behavioral response. For example, the task should be one in which the participant makes a binary decision. In the traditional signal detection task, the participant must conclude that the auditory signal has, or has not, been presented. In the contingency assessment task, the participant must conclude that ingestion of shrimp is or is not followed by an allergic reaction, or that the fertilizer is or is not effective, or that the chemical does or does not have antibacterial action. In the typical implementation of the contingency assessment task, the participant does not make a categorical response but rather indicates the strength of the contingent relationship on an analogue scale. Both Allan et al. (2005) and Perales et al. (2005) additionally included a categorical prediction response. Each trial started with the information about the cue status (presented or not presented), and then the participant had to predict whether or not the outcome would occur on that trial. Allan et al. and Perales et al. evaluated the usefulness of a SDT analysis of contingency assessment by examining these prediction responses averaged over participants. Their results indicated the value of such an analysis of prediction responses. However, the application of a SDT analysis to the trial prediction responses is indirect in that it assumes that the prediction of the outcome on C and on $\sim C$ trials reflects the

[1] The same is true of the contingency assessment task faced by the nonhuman animal. For example, when an odor associated with a potential predator is detected, the question for the animal is not, "What is the magnitude of the relationship between the odor and the presence of a predator?" Rather, the question is, should the animal stay or flee?

participant's assessment of the contingency on those trials. A binary dependent measure in which the participant is explicitly asked about the contingency would be a more direct approach. Other concerns about the use of trial prediction responses as reflecting contingency assessment have been raised by Winman and Gredebäck (2006). Additionally, the pitfalls of estimating parameters and evaluating models on the basis of averaged data are well known (see Wickens, 2002). One purpose of the present experiments was to evaluate the usefulness of the SDT view of contingency assessment with a categorical response that would allow individual (not group) estimates of sensitivity and criterion and that would not be subject to other interpretative problems inherent in the prediction response.

SDT typically is used in the context of psychophysical procedures that demand extensive within-participant measures of performance. A second challenge in evaluating a SDT analysis of contingency assessment is the development of procedures for presenting a participant with a sequence of cues and outcomes. The discrete-trial contingency task is poorly suited for a SDT approach. Many cue–outcome presentations must be provided to the participant to ensure that sufficient information is given about the actual contingency. Depending on the nature of the visuals used to represent cues and outcomes, a series of trials can take many minutes. For example, with presentation times of 3 s for both the cue and the outcome and an interpair interval of 2 s, viewing a block of 40 pairings takes more than 5 min. Requiring trial prediction responses adds to the length of the trial. Thus, few ratings can be obtained from a participant during a typical session, greatly limiting the experimenter's ability to make within-participant comparisons.

Crump, Hannah, Allan, and Hord (2007) recently described a new procedure that permits rapid measurement of contingency assessment—the streamed-trial task. The present experiments used this task to address both of the challenges to analyzing contingency assessment within the context of SDT. With the streamed-trial procedure, it takes only a few seconds to define a contingency value. The rapid sequential presentation of cue–outcome pairs allows an entire block of trials to be telescoped into a single streamed trial. A presentation stream is depicted schematically in Figure 1. The cue and the outcome are colored geometric forms. Each 100-ms presentation consists of one of four cue–outcome combinations (see insets, Figure 1), and presentations are separated by a 100-ms black screen. The contingency value is defined by a presentation stream of these cue–outcome combinations. For example, if each of the four cue–outcome combinations occurred equally often, the contingency between cue and outcome would be 0.

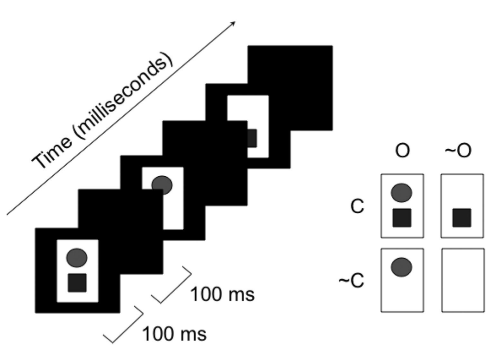

Figure 1. On the left is a schematic illustrating the structure of a streamed trial in Experiment 1A. On the right (insert) are the four possible cue–outcome combinations in a streamed trial. Squares represent cues (*C*) and were presented in blue. Circles represent outcomes (*O*) and were presented in red.

In the Crump et al. (2007) experiment, following the rapid stream of cue and outcome presentations, participants were asked to assess the contingency between the events using the traditional rating procedure. In the present experiments, we modified the streamed-trial task to provide an analogue for a commonly used psychophysical procedure, the method of constant stimuli. Rather than asking for a rating at the end of a streamed trial, we asked participants to categorize the contingency between the two geometric forms either as *strong* or *weak*. This binary dependent measure, unlike ratings, provides information about the participant's sensitivity to the contingencies. For example, one can determine how different two contingencies have to be in order to be discriminated (i.e., the just-noticeable difference) and whether sensitivity to positive contingencies differs from sensitivity to negative contingencies. Most important, the binary response, unlike the rating response, is amenable to theoretical analyses that dissociate the participant's ability to assess the actual contingency from any biases that might exist in the decision process.

Experiment 1

Experiment 1A: Geometric Forms

In Experiment 1A, we used the method of constant stimuli to explore contingency sensitivity. In a typical psychophysical experiment using the method of constant stimuli, there are *k* possible values of the independent variable, and the participant's task is to make a binary decision. For example, in loudness discrimination, one of *k* intensities is presented on each trial, and the participant's task is to categorize the perceived loudness as either *loud* or *soft*. On each trial of Experiment 1A, we presented one of 11 ΔP values and asked the participant to judge the contingency between the cue and the outcome as either *strong* or *weak*.

Method

Participants, apparatus, and stimuli. Four participants completed (see Table 2). The experiments were controlled by eMac computers (Apple Corp., Cupertino, CA) on 17–in. (43.18-cm) cathode-ray-tube displays running Metacard software (Metacard Corp., Boulder, CO). Participants sat approximately 60 cm from the computer screen. There were four possible cue–outcome pairs, and each of these pairs is depicted in Figure 1. A cue–outcome pair was presented for 100 ms in a gray frame (height × width, 6.4 cm × 5.0 cm) displayed in the center of a black screen.[2] The cue was a blue square (1.6 cm in height and width) centered at the bottom of the frame. The outcome was a red circle (1.6 cm in diameter) centered at the top of the frame. Cue–outcome pairs were separated by a 100-ms black screen. Note that the cue absent–outcome absent combination (cell *d*) was an empty gray

[2] The refresh rate of the displays was 89 Hz. The timing of the raster display was not synchronized to the timing of stimulus presentations. This would have introduced some flicker, but any degradation would have been random and would not have differed systematically across conditions or participants.

frame clearly discriminable from the black screen that appeared between frames.

Procedure. There were 11 values of ΔP ranging from 0 to 1.0 in increments of 0.1. A stream of 60 cue–outcome presentations, with a total duration of approximately 12 s, defined a value of ΔP. Table 3 shows the ΔP, $P(O|C)$, and $P(O|{\sim}C)$ values, and also the frequencies in cells a and c of the 2 × 2 contingency matrix. For all ΔP values, the probability of the cue, $P(C)$, was .5, and the probability of the outcome, $P(O)$, was close to .5; (referring to Table 1) the values were derived as follows:

$$P(C) = \frac{a+b}{a+b+c+d}.$$

and

$$P(O) = \frac{a+c}{a+b+c+d}.$$

At the end of each streamed trial, the participant was required to make a binary decision. Two clickable response (R) buttons, one labeled *Weak* (an R_W response) and one labeled *Strong* (an R_S response), appeared on the screen. The participant's task was to select the button that best represented the strength of the contingency between the square and the circle on that streamed trial.

Each of the 11 ΔP values was presented four times in a randomized order during each block of 44 streamed trials. A session consisted of five blocks, resulting in 20 presentations of each of the 11 ΔP values. Each participant completed 10 sessions.

Instructions. At the beginning of first session, the following instructions appeared on the monitor:

> On each trial, you will be presented with a rapid sequence of squares and circles. The square represents a cue that may or may not occur, and the circle represents an outcome that may or may not occur. Taken together, there are four important events that can occur in each sequence.
>
> 1. The square and the circle occur together.
>
> 2. The square appears, but the circle does not.
>
> 3. The square does not appear, but the circle does appear.
>
> 4. Both the square and the circle do not appear.
>
> Each sequence that you see will contain many of each of these four events. After viewing each sequence, you will be asked to judge whether the contingency between the square and the circle was weak or strong. A strong contingency, in this case, means that the appearance of the square positively predicted the appearance of the circle. In other words, most of the time when the square appeared, the circle appeared, and most of the time when the square did not appear, the circle did not appear. A weak contingency means that the appearance of the square did not predict the appearance of the circle. In other words, the square appeared with the circle just as often as it appeared without the circle. Your task, after viewing each sequence, is to decide whether the contingency between the square and the circle was weak or strong.

Table 2
Summary of Participants in the Experiments

Participant	Status	Experiment 1A	1B	2	3	4
MC	Graduate student	√	√	√	√	√
JB	Research assistant	√				
AC	Graduate student	√				
CT	Graduate student	√				
ER	Graduate student		√			
SV	Volunteer		√			
AS	Research assistant		√	√		
AB	Graduate student			√	√	√
XG	Graduate student			√	√	√
AAS	Graduate student			√		
GM	Graduate student				√	√
KS	Graduate student					√

Note. Graduate students and volunteers were paid $10 per session. Research assistants participated as part of the job requirement.

Table 3
Experiment 1: Values for ΔP, $P(O|C)$, $P(O|{\sim}C)$, and $P(O)$ and the Frequencies of the 2 × 2 Contingency Matrix in Cells a and c

ΔP	$P(O\|C)$	$P(O\|{\sim}C)$	a	c	$P(O)$
0	.500	.500	15	15	.500
.1	.567	.467	17	14	.517
.2	.600	.200	18	12	.500
.3	.633	.333	19	10	.483
.4	.700	.300	21	9	.500
.5	.767	.267	23	8	.517
.6	.800	.200	24	6	.500
.7	.833	.133	25	4	.483
.8	.900	.100	27	3	.500
.9	.933	.033	28	1	.483
1.0	1.00	.000	30	0	.500

Note. ΔP = contingency between cue and outcome, O = outcome; C = cue; ${\sim}C$ = absence of cue.

Results and Discussion

The participant's decision problem in the streamed-trial procedure, as conceptualized within a SDT framework, is presented schematically in Figure 2.[3] Repeated presentations of a constant ΔP value do not result in a constant subjective value. Rather, the resulting subjective value is variable. To simplify the presentation in Figure 2, we have shown only four ΔP values (.4, .5, .6, and .7) rather than the 11 values used in the experiment. The x axis is a random variable, X representing subjective contingency values. The left y axis shows values of the probability density, $f(X)$, for the four values of ΔP. Figure 2 illustrates the simplest version of SDT, which assumes that the distribution of subjective values generated by a constant ΔP is normal, with a mean equal to the physical ΔP value and a standard deviation, σ, that is constant across all ΔP values. The participant's task is to place the subjective value experienced on each streamed-trial into one of two categories, R_W or R_S. The participant does so by setting a decision criterion value, λ. If the subjective value is larger than λ, the response is R_S and if the subjective value is less than λ, the response is R_W. The area to the right of λ under each distribution represents the probability

[3] Many other psychophysical models have been proposed since the introduction of SDT. We have used the SDT framework in the present article because SDT is the simplest and the most familiar of these models. Our purpose is to evaluate a psychophysical approach, rather than to argue for a specific psychophysical model.

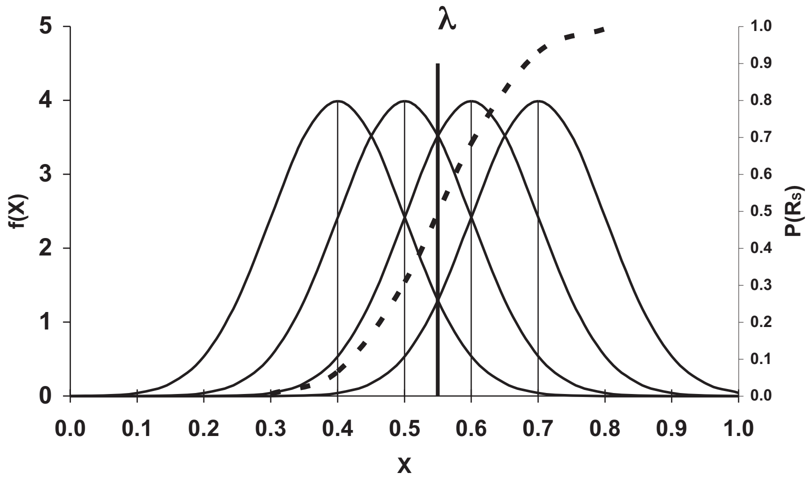

Figure 2. The participant's decision problem for the streamed-trial task with four values of ΔP (cue–outcome contingency). Probability density, *f(X)*, is on the left *y* axis; probability of a *Strong* response, $P(\mathrm{R_S})$, is on the right *y* axis. The *x* axis is a random variable, with *X* representing subjective contingency values The dashed line is the psychometric function. λ = decision criterion value.

that the subjective value was larger than λ. Thus, the probability of a strong response, $P(\mathrm{R_S})$, generated by a particular value of ΔP provides an estimate of the proportion of the area under the distribution to the right of λ. The function in Figure 2 that plots $P(\mathrm{R_S})$ on the right *y* axis as a function of ΔP is referred to as the psychometric function. When the underlying distributions are normal and have a constant variance, the psychometric function is the cumulative normal function (see Killeen, Fetterman, & Bizo, 1997). The slope of the cumulative normal function is $1/\sigma$ and provides an estimate of the participant's sensitivity to discriminating among the ΔP values. The value of ΔP—for which $P(\mathrm{R_S}) = .5$, often referred to as the point of subjective equality (*PSE*)—provides an estimate of λ.

Figure 3 displays $P(\mathrm{R_S})$ as a function of ΔP for each of the four participants. The streamed-trial procedure clearly produces orderly psychometric functions—$P(\mathrm{R_S})$ increases with increasing ΔP. We fitted the cumulative normal to the data from each participant using pro Fit (QuantumSoft, Uetikon am See, Switzerland).[4] One measure of the goodness of fit of a function to the data is R^2—the proportion of the variance in the obtained values of $P(\mathrm{R_S})$ accounted for by the fitted function. The R^2 values for each participant are available in Table 4a. The cumulative normal function provides an excellent description of each participant's data. For three participants, the fitted functions accounted for over 99% of the variance in $P(\mathrm{R_S})$.

The λ and σ values for the individual psychometric functions are shown in Table 4a. The values of σ indicate that the participants varied in their ability to discriminate among the ΔP values. Similarly, the values of λ indicate that the participants varied in the placement of their criterion for $\mathrm{R_S}$. Participants AC and MC tended to be more conservative about responding *Strong* ($\lambda > .5$) than were Participants CT and JB ($\lambda < .5$).

Experiment 1B: Emoticon Stimuli

In Experiment 1A, the cue–outcome stimuli were geometric forms that were presented simultaneously, and the participant was asked to categorize the contingency as weak or strong. To evaluate the generality of our findings, in Experiment 1B, we used the cues and outcomes from a different stimulus set, made the representation of event present and event absent perceptually symmetrical, made the timing for the presentation of the cue and the outcome sequential rather than simultaneous, and used a different wording for the binary response. The cues and outcomes were schematic face drawings indicating different emotional expressions. These schematic face drawings, popularly termed *emoticons*, are commonly used in e-mails and chat rooms to express emotions during text-based conversations. The presence of the cue and the outcome was represented by a smiling expression, and the absence of the cue of the outcome was represented by a neutral expression. The cue emoticon was presented before the outcome emoticon, and the participant was required to categorize the predictive relationship as weak or strong.

Method

Participants, apparatus, and stimuli. There were four participants (see Table 2). The four cue–outcome pairs are shown in Figure 4. A cue–outcome pair was presented in the center of a black monitor. Both cue and outcome faces were in a white square (3 cm × 3 cm) and consisted of a circular schematic drawing (3 cm in diameter). Cue faces were yellow and were presented on the left; outcome faces were orange and were presented on the right. Faces were either smiling (*C* and *O* events) or neutral ($\sim C$ and $\sim O$ events). The only difference between smiling and neutral faces was the emotion expressed on the mouth. Smiling faces displayed a mouth in the form of a conspicuous crescent-moon shape (2.1 cm in length), and neutral faces displayed a mouth in the form of a flat line (1.1 cm in length).

Procedure. Each frame consisted of the staggered presentation of cues and outcomes. At the beginning of each cue–outcome

[4] The software is available at the Web site: http://www.quansoft.com/

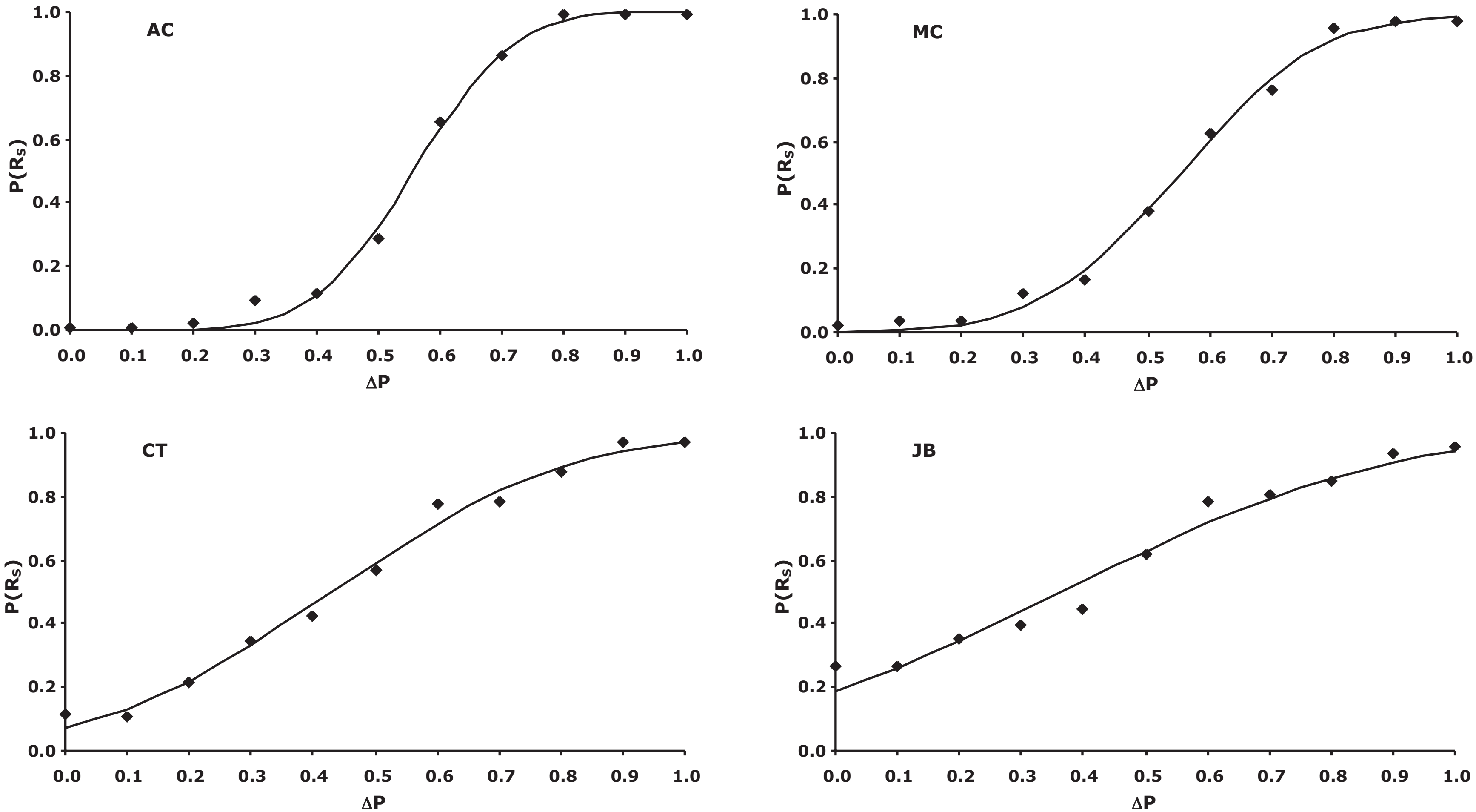

Figure 3. Probability of a *Strong* response, $P(R_S)$, as a function of ΔP (cue–outcome contingency) for each participant (AC, MC, CT, and JB) in Experiment 1A. The black diamonds represent the data, and the lines are the best-fitting Gaussian functions.

Table 4a
Experiment 1A: σ, λ, and R^2 Values for Each Participant

Participant	σ	λ	R^2
AC	.13	.56	.9969
MC	.18	.55	.9964
CT	.30	.43	.9907
JB	.41	.37	.9690

Note. σ = standard deviation; λ = decision criteria; R^2 = strength of relationship.

presentation, a left face was presented in isolation for 100 ms, followed by a right face that remained on the screen with the left face for 200 ms. The interpair interval was 50 ms. At the end of the stream, the participant was asked to indicate whether the happiness of Person A was a strong or weak predictor of Person B's happiness. As in Experiment 1A, a streamed trial consisted of 60 cue–outcome presentations. However, the staggered presentation and the longer interpair interval increased the duration of a stream to approximately 21 s. Because the duration of the stream was longer than in Experiment 1A, a session consisted of three (rather than five) blocks, resulting in 12 presentations of each of the 11 ΔP values. Each participant completed 15 sessions. In all other details, the procedure of Experiment 1B was the same as that of Experiment 1A.

Instructions. At the beginning of the first session, the following instructions appeared on the monitor:

> In this experiment, you will be asked to judge whether one person's emotional state is a strong or a weak predictor of another's emotional state. On each trial, you will be shown a summary of a conversation that took place on an Internet chat line. To summarize the emotional content of each conversation, we will be presenting you with the emoticons that each person used during the conversation. There are always two people in each conversation, each represented by an emoticon that appears on the left or the right (example pictures to follow). The emoticon for Person A on the left always represents the person who initiated the chat, and the emoticon for Person B on the right always represents the person who agreed to the chat. For each conversation, we will present you with the emoticons used by each person as the conversation progressed. Your task will be to judge whether the happiness of Person A is a strong or weak predictor of Person B's happiness. On each trial, you will be presented with a rapid sequence of emoticons for Person A (on the left) and Person B (on the right). Person A on the left either smiles or does not. Person B on the right either smiles or does not. Taken together, there are four important events that can occur in each sequence:
>
> 1. Person A and B both smile.
>
> 2. Person A smiles, but Person B does not.
>
> 3. Person A does not smile, but Person B smiles.
>
> 4. Person A and B do not smile.
>
> Each sequence that you see will contain many of each of these four events. At the end of each stream of emoticons, you will be presented with a button marked *Weak* and a button marked *Strong*. If you think that Person A's happiness weakly predicted Person B's happiness or did not predict it at all, then press the button marked *Weak*. If you think that Person A's happiness strongly predicted Person B's happiness, then press the button marked *Strong*.

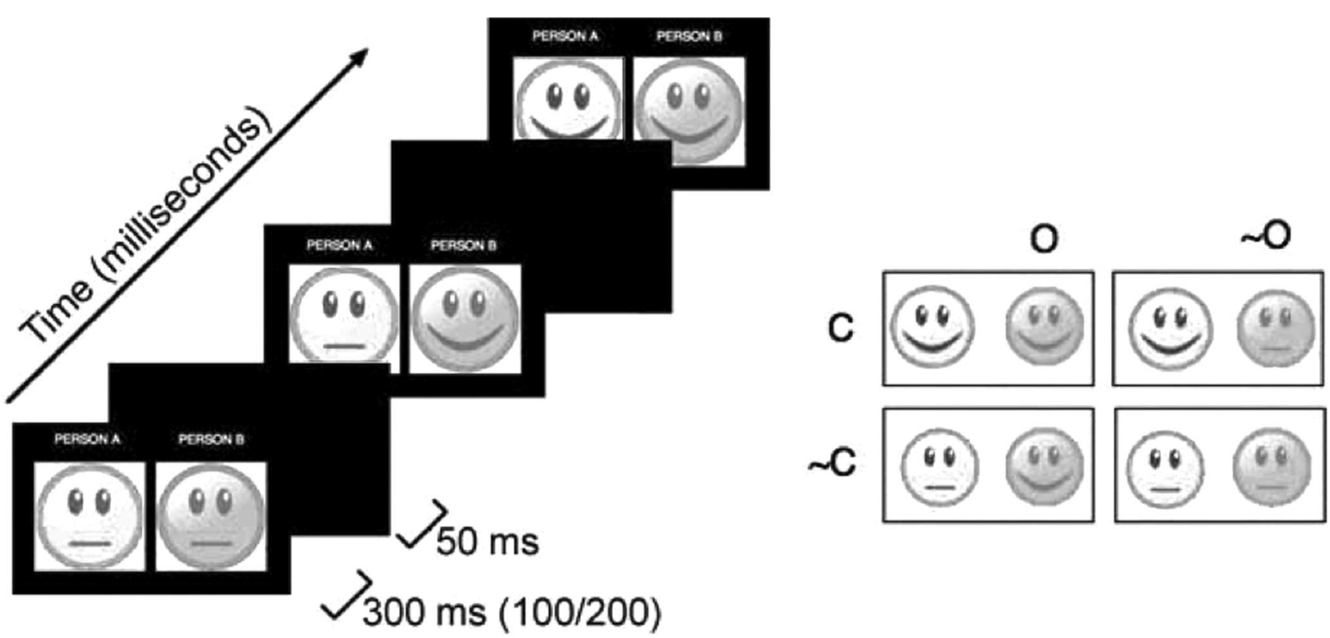

Figure 4. On the left is a schematic illustrating the structure of a streamed trial in Experiment 1B. At the beginning of each cue–outcome presentation, the left face was presented in isolation for 100 ms, followed by the right face that remained on the screen with the left face for 200 ms. The interpair interval was 50 ms. On the right are the four possible cue–outcome combinations in a streamed trial. Cue faces were yellow and were presented on left; outcome faces were orange and were presented on the right. Faces were either smiling (C and O events) or neutral ($\sim C$ and $\sim O$ events).

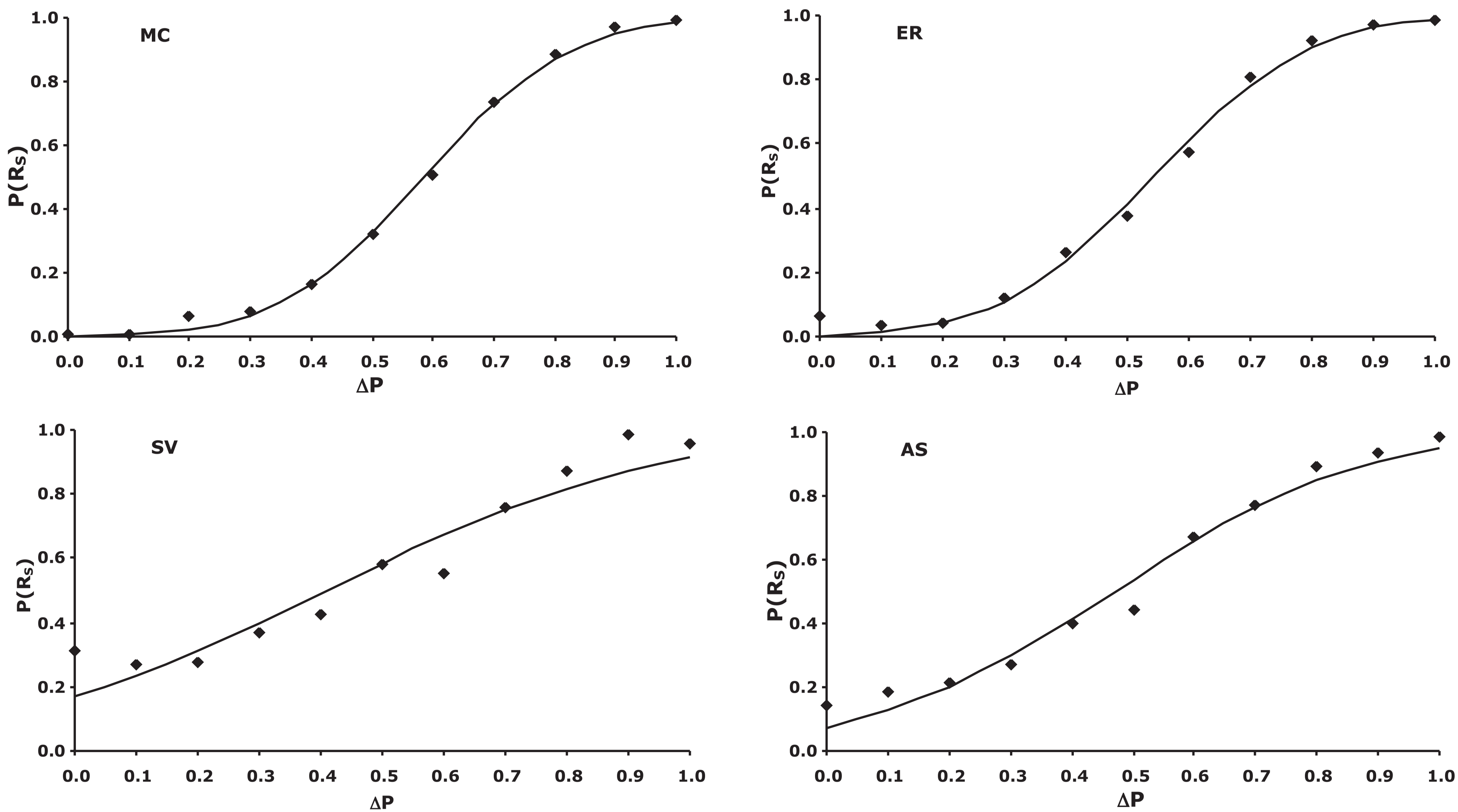

Figure 5. Probability of a *Strong* response, $P(R_S)$, as a function of ΔP (cue–outcome contingency) for each participant (MC, ER, SV, and AS) in Experiment 1B. The black diamonds represent the data, and the lines are the best-fitting Gaussian functions.

Results and Discussion

Figure 5 displays $P(R_S)$ as a function of ΔP for each of the four participants. As in Experiment 1A, $P(R_S)$ increased with increasing ΔP, and we fitted the cumulative normal to the data from each participant. The R^2, λ, and σ values are available in Table 4b. As in Experiment 1A, the participants differed in their ability to discriminate among the contingency values and in their criterion placement. Participant MC took part in both Experiment 1A and 1B. His parameter values were very similar in the two experiments. The data from Experiment 1B indicated that the streamed-trial procedure is not dependent on the geometric forms originally used by Crump et al. (2007), nor is it dependent on simultaneous presentation of cues and outcomes.

The results of Experiments 1A and 1B are relevant to understanding how the instructions to the participants affect contingency ratings. A number of researchers have suggested that the wording of the rating question in the traditional contingency assessment task is an important variable. For example, Vadillo, Miller, and Matute (2005) compared three wordings: *causal* ("To what extent do you think the medicine is the cause of the allergic reaction?"), *predictive value* ("To what extent do you think that taking the medicine is a good predictor of the allergic reaction"), and *prediction* ("If a patient has taken the medicine, to what extent do you think it is likely that this patient will develop the allergic reaction?). They found that prediction ratings differed from the ratings for the other two question types but that the predictive value ratings and causal ratings did not differ. Vadillo et al. (2005) hypothesized that question type had its effect through "postacquisition processes that modulate participants' responses in a flexible way" (p. 172). The streamed-trial procedure with sequentially presented emoticons provides a methodology for the direct investigation of this hypothesis. Our questions in Experiments 1A ("Is the contingency strong or weak?") and 1B ("Was the emotional state of Person A a strong or weak predictor of the emotional state of Person B?") fall into the predictive-value question type described in Vadillo et al. (2005). A prediction categorization at the end of the emoticon stream would ask, "Given that Person A smiled, what is the likelihood that Person B smiled—weak or strong?" A causal categorization at the end of the emoticon stream would ask, "Was Person A's mood a strong or weak cause of Person B's mood?" One would expect, on the basis of the hypothesis put forth by Vadillo et al. (2005), that question type would have its effect on the placement of the criterion (and not on sensitivity to the contingency).

Experiment 1C: Group Data

The results of Experiments 1A and 1B indicated that the streamed-trial procedure yields orderly psychometric functions for individual participants who are well practiced on the task. It was argued earlier that psychometric functions based on individual participant data should be used for the evaluation of psychophysical models, especially when the goal is to unconfound the effects of variables on sensitivity and response bias. However, there might be situations when the experimenter would prefer a more conventional approach, yielding data from one session averaged across a group of participants. Experiment 1C was conducted to explore whether the streamed-trial procedure would yield orderly psychometric functions obtained by averaging across inexperienced participants.

Table 4b
Experiment 1B: σ, λ, and R^2 Values for Each Participant

Participant	σ	λ	R^2
MC	.19	.58	.9990
ER	.20	.54	.9979
SV	.43	.41	.9617
AS	.32	.47	.9895

Note. σ = standard deviation; λ = decision criteria; R^2 = strength of relationship.

Method

Seventeen McMaster University students either received partial course credit or were paid $10 (Canadian) for their participation. Experiment 1C was identical to Experiment 1A, except that each participant completed only one session.

Results and Discussion

Figure 6 displays $P(R_S)$ as a function of ΔP averaged over the 17 participants. We fitted the cumulative normal to the data of each of the 17 participants. The mean of the 17 values of σ was .33, the mean of the 17 values of the *PSE* was .51, and the function in Figure 6 was the mean of the 17 individual fitted functions. The streamed-trial procedure, in conjunction with the method of constant stimuli, does generate an orderly psychometric function based on group data. However, caution should be exercised in the interpretation of the mean function. While on average, there does not appear to be any bias for a particular response (i.e., the mean *PSE* = .51), the 17 *PSE* values ranged from −.084 to .845 (*SD* = .200). Also there was considerable variability in the sensitivity parameter σ, ranging from 0.05 to 1.85 (*SD* = 0.41). Finally, although some of the individual fits between the observed data and the predicted values were very good, others were not. The mean R^2 value was .9113, ranging from .4193 to .9983 (*SD* = .1400).

Experiment 2: Positive and Negative Contingencies

The previous experiments indicated that the streamed-trial procedure, in conjunction with the method of constant stimuli, yields psychometric functions that provide information about contingency sensitivity (σ) and about criterion placement (λ). In those experiments, all contingencies were positive. A number of reports in the literature have provided for a comparison of negative and positive contingencies. A frequent finding is that the sign of the contingency affects ratings. For example, the data reported by Dickinson et al. (1984) and by Perales et al. (2005) indicated that for a fixed ΔP value, a larger value on the rating scale was used when the sign was positive than when it was negative. Wasserman, Elek, Chatlosh, and Baker (1993)[5] and Maldonado, Catena, Cándido, and Garcia (1999) concluded that ratings were better calibrated to the actual ΔP values when they were positive than when

[5] Wasserman et al. (1993) used the free operant version of the active task in which trials are not delineated.

they were negative. Mutter and Williams (2004)[6] reported that under some conditions, the effect of the sign of the contingency on ratings was age dependent. Older adults' ratings of negative contingencies, compared with young adults' ratings, were less affected by the size of the negative contingency. The age effect was not present for positive contingencies. In Experiment 2, we compared psychometric functions for positive and negative contingencies. We were especially interested in determining whether the asymmetry in the ratings that others have reported was attributable to contingency sensitivity or to a biased criterion.

Method

There were five participants (see Table 2). Experiment 2 was the same as Experiment 1A, except that in some sessions, the values of ΔP were negative. The sign of the 11 ΔP values was constant in a session and alternated between sessions, and the participant was reminded about the sign at the beginning of each session.[7] Each participant completed 20 sessions, 10 with the set of positive ΔPs and 10 with the set of negative ΔPs. The instructions were similar to those used in Experiment 1A except that both a positive contingency and a negative contingency were described.

Results and Discussion

Figure 7 displays $P(R_S)$ for each participant as a function of ΔP for the two signs. Except in the case of Participant XG, the two functions appear to have similar slopes (i.e., σ) but differ in *PSE* (i.e., λ). For each participant, separate psychometric functions were fitted for the two signs with the restriction that σ was the same across the two functions. The σ, λ, and R^2 values for each participant are available in Table 5. So that the effect of the common σ constraint on the fit can be evaluated, R^2 values are also shown for fits that allowed different values of σ for the two signs. The constraint of a common σ had a negligible effect on R^2. For all participants except XG, λ was smaller for positive contingencies than for negative contingencies, suggesting that the asymmetry in the effect of sign on ratings that has been reported in the literature is likely attributable to a biased criterion placement.

Biased criterion placement may be relevant to understanding the effect of age on contingency assessment reported by Mutter and Williams (2004). They concluded that "detecting causal contingency apparently becomes more difficult with age, especially . . . when the relationship between a causal event and an outcome is negative" (p. 13). The design of the experiments in Mutter and Williams, however, do not allow for locating the age difference in the detection of the contingency. The experiments reported in the present article could be repeated with older adults. This would allow one to determine whether older and young adults differ in the detection of the contingency (σ) or in the manner in which they respond (λ). As reviewed by Allan and Siegel (2002), it is now well established that some findings in the literature (such as short-term memory performance, gustatory assessment, and flicker fusion) that had been attributed to sensitivity deficits actually reflected different decision strategies across participant groups.

Experiment 3: Outcome Density

The purpose of Experiment 3 was to use the streamed-trial procedure and the method of constant stimuli to investigate a well-established but poorly understood phenomenon of contingency judgments—the outcome density effect (ODE). The ODE refers to the finding that for a fixed ΔP, ratings of contingency usually are not constant but rather increase with the probability of the outcome, $P(O)$.

Allan et al. (2005) suggested that SDT is relevant to understanding the ODE. In their experiment, they used the traditional contingency task and based their SDT analysis on the trial prediction responses. Allan et al. (2005) concluded that outcome density does not affect sensitivity to the contingency; rather, it affects the participant's willingness to predict that the outcome will occur.

As discussed previously, the traditional contingency task is not well suited for evaluating SDT interpretations of contingency assessment. Crump et al. (2007) recently demonstrated the ODE in the streamed-trial task, with ratings as the dependent measure. In Experiment 3, we used the streamed-trial task with a binary behavioral response. We manipulated outcome density. If the ODE is indeed a criterion effect, one would expect $P(O)$ to have its affect on λ rather than on σ.

Method

There were four participants (see Table 2).[8] A stream of 80 presentations of the geometric form stimuli used in Experiment 1A defined a value of ΔP. The value of $P(C)$ was .5, and the value of $P(O)$ was either .3 or .7. Table 6 shows the ΔP, $P(O|C)$, and $P(O|{\sim}C)$ values and also the frequencies in cells *a* and *c* of the 2×2 contingency matrix for each value of $P(O)$. The range of possible ΔP values was constrained by the number of presentations in the streamed trial and the values of $P(C)$ and $P(O)$. With 80 presentations in a stream, $P(C)$ value of .5, and $P(O)$ values of .3 and .7, nonnegative ΔP values were constrained to values equal to or less than .6. Seven values of ΔP, ranging from .0 to .6 in increments of .1, were used.

$P(O)$ was constant throughout a session and was randomized between sessions. Each of the seven ΔP values was presented four times in a randomized order during each block of 28 streamed trials. A session consisted of five blocks, resulting in 20 presentations of each of the seven ΔP values. Each participant completed 20 sessions, 10 with each value of $P(O)$ in a randomly determined order. In all other details, the procedure of Experiment 3 was the same as that of Experiment 1A.

Results and Discussion

Figure 8 displays $P(R_S)$ for each participant as a function of ΔP for the two values of $P(O)$. Except for data from Participant AB, the two functions differ in λ (i.e., the *PSE*) and appear to have similar slopes. For each participant, separate psychometric func-

[6] Mutter and Williams (2004) also used the free operant version of the active contingency task in which trials are not delineated.

[7] It is commonplace in psychophysical experiments with well-practiced participants to fully inform the participants about the conditions they will experience in a session. Providing participants with such information decreases variability due to uncertainty, since once the participants have experienced an easy-to-discriminate value, they would know which session type was in effect (i.e., whether they were in the positive or in the negative session).

[8] There was 1 additional participant, but we could not fit her data. She made few R_S responses, even to the larger ΔP values.

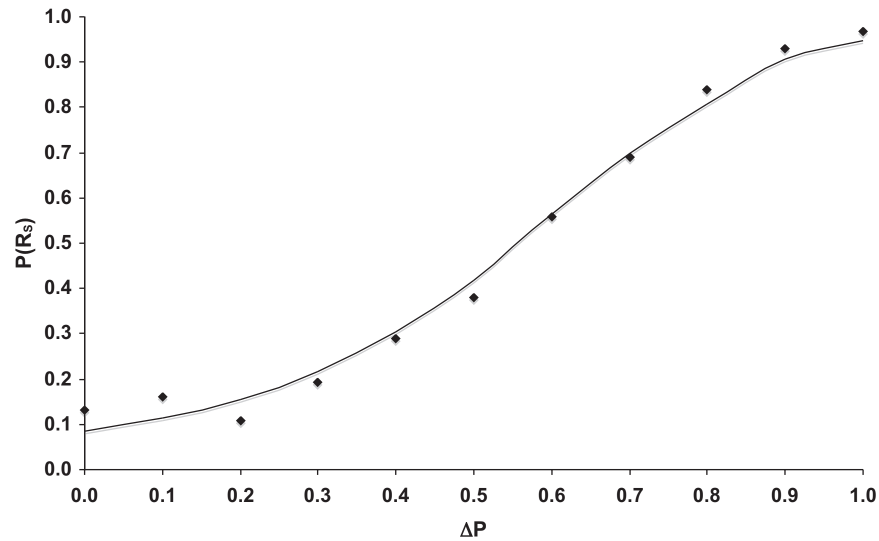

Figure 6. Mean probability of a *Strong* response, $P(R_S)$, as a function of ΔP (cue–outcome contingency) in Experiment 1C. The black diamonds represent the data, and the line is mean of the Gaussian functions fitted to each participant's data.

tions were fitted for the two values of $P(O)$, with the restriction that σ was the same across the two functions. The σ, λ, and R^2 values for each participant are available in Table 7. To evaluate the effect of the common σ constraint on the fit, we also determined values of R^2 for fits that allowed different values of σ for the two $P(O)$ values. The constraint of a common σ had little effect on R^2. For all participants, λ is smaller for $P(O) = .7$ than for $P(O) = .3$, indicating that $P(R_S)$ increases with $P(O)$.

These data are consistent with the suggestion of Allan et al. (2005) that P(O) does not affect the ability to perceive the strength of the relationship between the cue and the outcome but does affect the tendency to categorize the relationship as strong. The results not only indicate a mechanism for the ODE but also are relevant to understanding individual differences in susceptibility to the phenomenon.

Although most participants (like those in the present experiment) displayed the ODE, there has been considerable interest in those who do not. Specifically, there are reports that depressed individuals (in contrast with nondepressed individuals) do not display the ODE. That is, people with depression do not inflate their ratings when $P(O)$ is increased. This apparent knack for people with depression not to be misled by outcome density in their contingency judgments has been termed *depressive realism* and the absence of an ODE has led to the characterization of depressed individuals as "sadder but wiser" (Alloy & Abramson, 1979). Allan et al. (2007) suggested that depressive realism might best be understood from a psychophysical analysis of contingency judgment. That is, depressed and nondepressed participants may not differ in their perception of contingency (depressed participants are not "wiser"), but rather nondepressed participants respond to the increased salience of an outcome—as implemented, for example, by increasing $P(O)$—by a change in decision criterion.[9]

Experiment 4: Payoffs

The SDT account of contingency assessment is different from alternative accounts primarily because it makes a distinction between sensitivity and criterion. The criterion is determined predominantly by the participant's analysis of the costs of making each of the two types of mistakes (false alarms and misses). The traditional way of manipulating decision criterion in psychophysical tasks is to explicitly manipulate the consequences of these mistakes (the *payoff structure*). Perales et al. (2005) manipulated payoff structure in the traditional, discrete-trial contingency task and concluded that payoffs affected criterion placement but not sensitivity. As we noted earlier, their SDT analysis was conducted on the prediction responses, and their conclusions were based on averaged data. In Experiment 4, we examined the role of payoff structure on performance in the streamed-trial task. To extend the generality of our findings, we modified the psychophysical task from that used in the previous experiments. In Experiment 4, there were only two possible ΔP values.

Method

Participants and procedure. There were five participants (see Table 2). The two values of ΔP were .4 and .6 for all participants except KS. Participant KS was less sensitive than the other participants, and the values for KS were .3 and .7. Each value of ΔP was defined by 60 streamed presentations, using the geometric forms from Experiment 1A. The cell frequencies for these contin-

[9] Alloy and Abramson (1979) varied outcome valence, as well as outcome density, in their experiments. An outcome was made either desirable or undesirable (rather than frequent or infrequent). In the "win" condition, the participant gained points on each trial on which the outcome occurred, and in the "lose" condition, the participant lost points on each trial on which the outcome did not occur. As with outcome density, Alloy and Abramson found that outcome valence influenced ratings of nondepressed but not of depressed individuals. That is, they observed depressive realism with an outcome valence effect, as well as with an outcome density effect. It is likely that the outcome valence effect, like the outcome density effect, reflects a decision strategy.

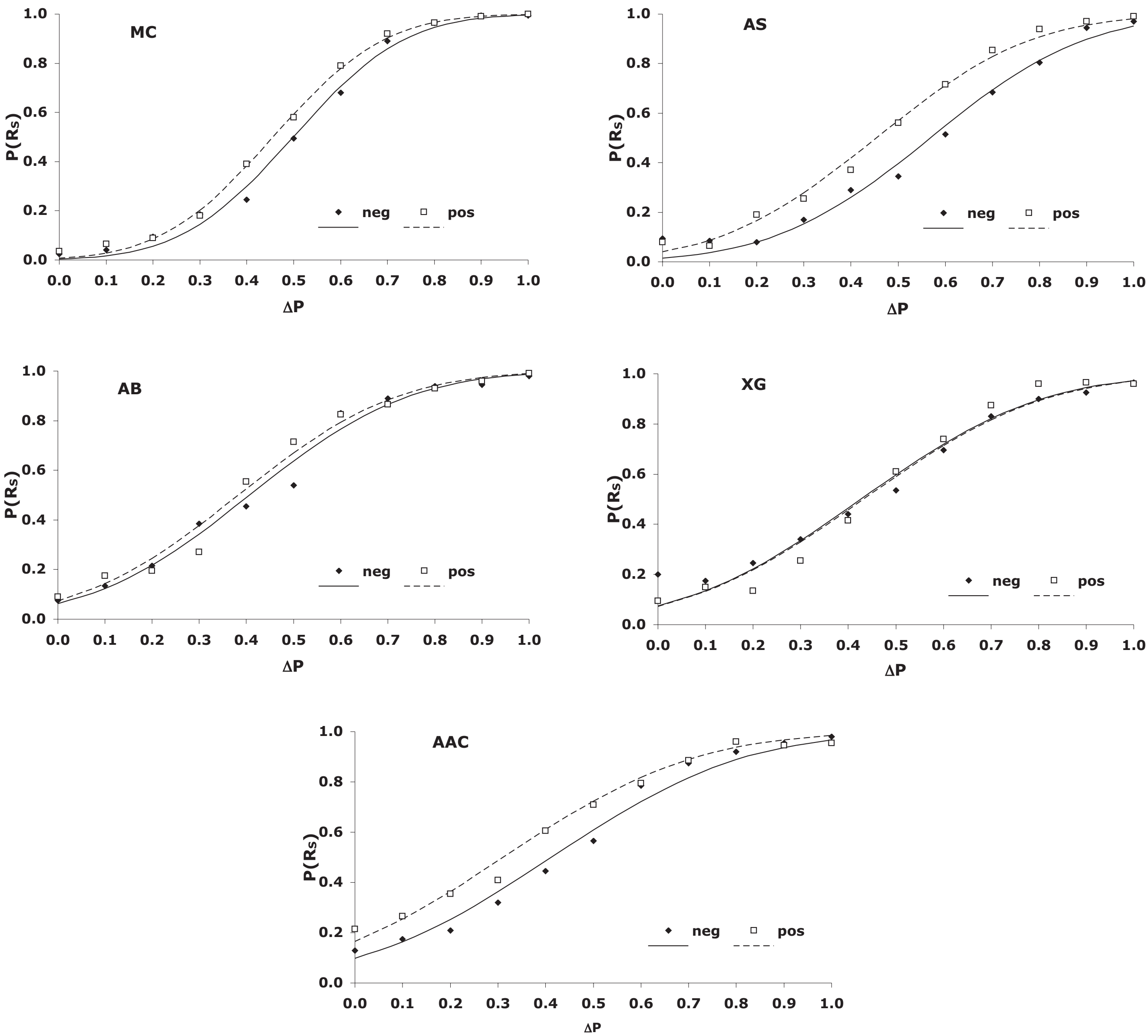

Figure 7. Mean probability of a *Strong* response, $P(R_S)$, as a function of ΔP (cue–outcome contingency) for each participant (MC, AS, AB, XG, and AAC) for each sign in Experiment 2. The black diamonds (negative) and white squares (positive) represent the data, and the lines (continuous for negative and hatched for positive) are the best fitting Gaussian functions fitted with the constraint that σ (common standard deviation) is constant across the two signs.

gencies are available in Table 3. As in the previous experiments, the participant made either a R_S or a R_W response at the end of each streamed trial. Each ΔP value was presented 14 times in a randomized order during each block of 28 streamed-trials. A session consisted of five blocks, resulting in 70 presentations of each ΔP value.

The 2 × 2 matrix relating the two ΔP values to the two response categories is shown in Table 8. There are two types of correct responses: R_S to the larger ΔP value (a hit, H) and R_W to the smaller ΔP value (a correct rejection, CR). There are also two types of errors: R_S to the smaller ΔP value (a false alarm, FA) and R_W to larger ΔP value (a miss, M). Participants won points for correct responses and lost points for errors.[10] There were two payoff matrices that are illustrated in Table 9. In the weak (W) condition, there were greater gains and smaller losses for R_W than for R_S. A CR earned 50 points, while an H earned only 10 points. Moreover, an FA lost 50 points, while an M lost only 10 points. This payoff structure should bias participants to respond *Weak*,

[10] These points were converted to cash at the end of the experiment. The actual amount accumulated by a participant over the 20 sessions was determined by his or her sensitivity and by the effect of the payoff matrix on his or her criterion. The amount ranged from \$17.97 to \$30.60.

moving λ toward a more conservative location. The payoffs were reversed in the strong (S) condition, where an H earned 50 points and a CR earned only 10, while an M cost 50 points, and an FA cost only 10. This payoff structure should bias participants to respond *Strong*, moving λ toward a more liberal location. Each participant completed 20 sessions, 10 with each payoff matrix in a randomly determined order. In all other details, the procedure of Experiment 4 was the same as that of Experiment 1.

Instructions. The following was added to the instructions from Experiment 1A:

> After you make your decision, you will be told whether you are correct or incorrect. If you are correct, you will win some points. If you are incorrect, you will lose some points. These points will accumulate across blocks. At the end of the experiment, you will be able to exchange your points for a small amount of money. The exchange rate will be 1,000 points to $1. After learning whether you are correct or not, you can press a button to return to the task for another trial.

Results and Discussion

The participant's decision problem in Experiment 4 is illustrated in Figure 9. There were only two distributions, and again the participant adopted a criterion λ. With two distributions, the participant's sensitivity, d', was defined as the difference between the means of the two distributions normalized by the common standard deviation σ (Green & Swets, 1966):

$$d' = \frac{\text{difference between means}}{\sigma}.$$

Estimates of d' and λ can be obtained from the data by converting the obtained values of P(H) and P(FA) to z scores, Z(H) and Z(FA), respectively: $d' = Z(\text{FA}) - Z(\text{H})$, and $\lambda = Z(\text{FA})$. Estimates of d' and λ are shown in Table 10 for each participant under each payoff condition. Also shown, for each parameter, is the difference between the estimates under the two payoffs. For two participants (AB and KS), payoffs had little effect on either d' or λ. For the other three participants (MC, XG, and GM), payoffs had a dramatic effect on λ and little if any effect on d'. For these participants, the criterion was more liberal for R_S when the payoffs favored H and FA compared with when the payoffs favored CR and M. Thus, when payoffs do affect behavior, they do so at the decision. Overall, the data are consistent with the conclusion reached by Perales et al. (2005) that payoffs affect decision strategy and do not affect the ability to perceive the strength of the relationship between the cue and the outcome.

Table 5
Experiment 2: Values for Each Participant

		λ			
Participant	σ	$-\Delta P$	$+\Delta P$	R^2	(R^2)
MC	.19	.50	.46	.9967	.9967
AB	.27	.41	.38	.9847	.9855
XG	.30	.43	.43	.9789	.9879
AS	.26	.57	.45	.9913	.9920
AAS	.32	.41	.31	.9859	.9894

Note. Values for σ (standard deviation) and λ (decision criteria) were derived with the constraint that σ was constant across the two signs for ΔP (contingency between cue and outcome). The first column of R^2 values describes the goodness of fit with σ constant. The next column, (R^2) values, describes the goodness of fit when σ for the two signs was allowed to assume different values.

Table 6
Experiment 3: P(O) and ΔP Values, P(O|C) and P(O|~C) Values for Each Combination of P(O) and ΔP, and the Frequencies of the 2 × 2 Contingency Matrix in Cells a and c

$P(O)$	ΔP	$P(O\|C)$	$P(O\|{\sim}C)$	a	c
.3	0	.30	.30	12	12
	.1	.35	.25	14	10
	.2	.40	.20	16	8
	.3	.45	.15	18	6
	.4	.50	.10	20	4
	.5	.55	.05	22	2
	.6	.60	.00	24	0
.7	0	.70	.70	28	28
	.1	.75	.65	30	26
	.2	.80	.60	32	24
	.3	.85	.55	34	22
	.4	.90	.50	36	20
	.5	.95	.45	38	18
	.6	1.0	.40	40	16

Note. ΔP = contingency between cue and outcome; O = outcome; C = cue; $\sim C$ = absence of cue.

General Discussion

The study of contingency assessment involves the study of the relationship between physical events (the statistical contingency between cue and outcome) and the participant's internal experience of these events. Inasmuch as "psychophysics is the study of the relationship between physical events and our internal experience of these physical events" (Allan & Siegel, 2002, p. 419), it would seem that contingency assessment would be a topic of considerable interest to psychophysicists. It is not. With very few exceptions, research concerned with contingency assessment and research concerned with psychophysics have progressed independently, each with its own traditions and each motivated by different theoretical perspectives and models. We have suggested that a particular psychophysical model, SDT, can be profitably applied to the contingency assessment situation.

Originally SDT was developed to examine how organisms decide whether an auditory signal has, or has not, been presented (see Green & Swets, 1966). Subsequently, it has been found to be applicable to many areas in addition to this limited domain. For example, SDT has been applied to medical diagnoses (reviewed by Swets, 1996), clinical psychological assessment (reviewed by McFall & Treat, 1999), responses of depressed people (reviewed by Allan et al., 2007), and the placebo effect (reviewed by Allan & Siegel, 2002). In all these cases, a participant's judgment was determined both by the participant's sensitivity to the stimuli and by the participant's decision criterion. The present experiments were designed to evaluate the applicability of applying SDT to contingency assessments.

The traditional contingency assessment task does not readily lend itself to a SDT analysis—most forms of the task do not involve a categorical response and do not permit extensive within-

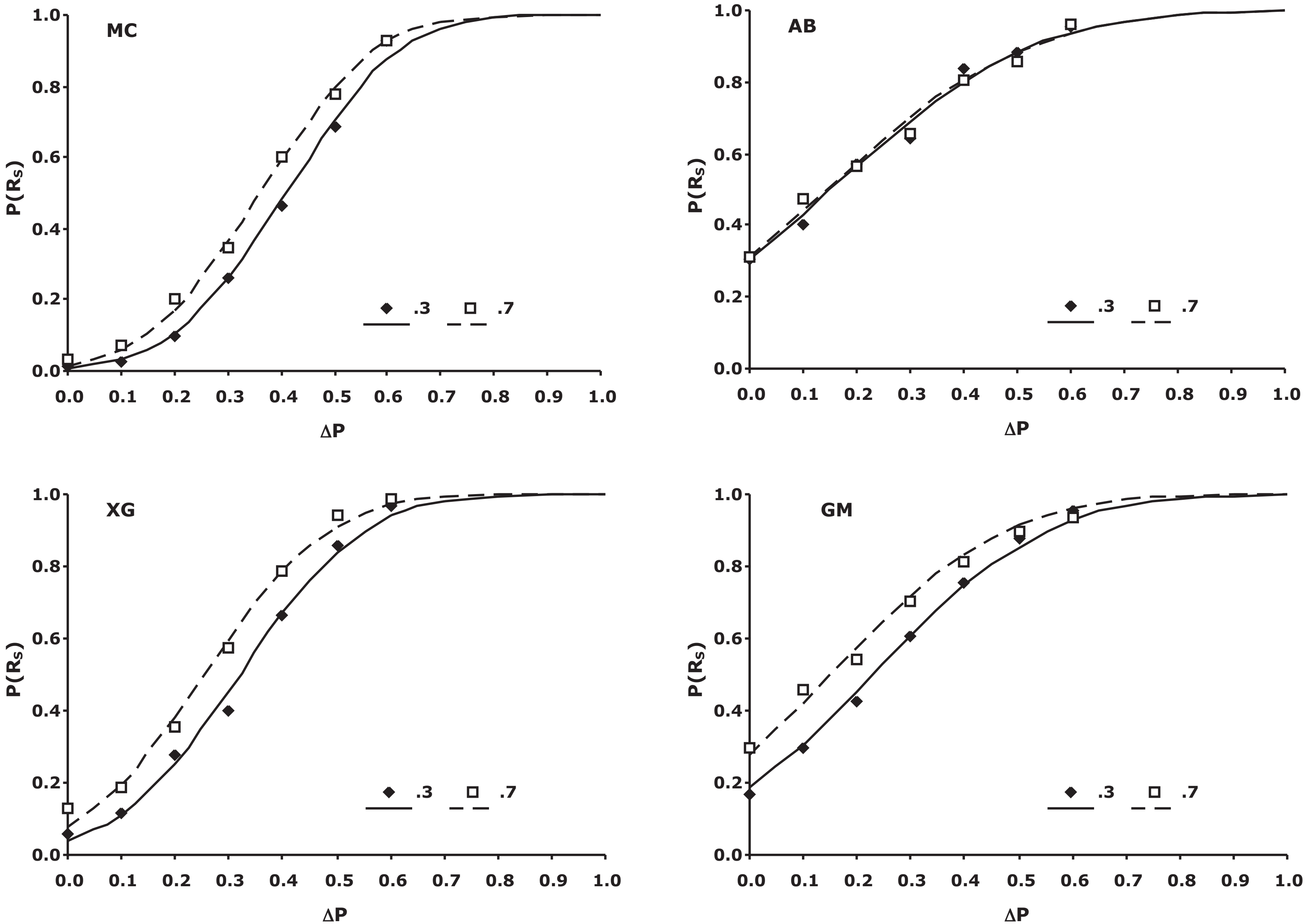

Figure 8. Mean probability of a *Strong* response, $P(\mathrm{R_S})$, as a function of ΔP (cue–outcome contingency) for each value of probability of outcome, $P(O)$, for participants (MC, AB, XG, and GM) in Experiment 3. The black diamonds (.3 value) and white squares (.7 value) represent the data, and the lines (continuous for .3 value and hatched for .7 value) are the Gaussian functions fitted with the constraint that σ (common standard deviation) is constant across the two values of $P(O)$.

Table 7
Experiment 3: Values for Each Participant

		λ			
Participant	σ	$P(O) = .3$	$P(O) = .7$	R^2	(R^2)
MC	.17	.41	.36	.9964	(.9972)
AB	.30	.15	.14	.9865	(.9881)
XG	.18	.32	.26	.9940	(.9940)
GM	.26	.23	.15	.9919	(.9966)

Note. Values for σ (standard deviation) and λ (decision criteria) were derived with the constraint that σ was constant across the two values of probability of outcome, $P(O)$. The first column of R^2 values describes the goodness of fit with σ constant. The next column, (R^2) values, describes the goodness of fit when σ for the two $P(O)$ values was allowed to assume different values.

participant measures of performance. We designed the experiments reported here so that we could evaluate contingency assessment using a categorical response (relationship judged either as strong or weak) and a task suitable to the SDT analysis (the streamed-trial procedure).

The results of Experiment 1 indicated that contingency assessments generate orderly psychometric functions and that participants' assessments of positive contingencies are determined by the independent actions of contingency sensitivity and individual response criteria. This was seen in data from well-practiced participants (Experiments 1A and 1B) and in group data (Experiment 1C). It was also the case for arbitrary geometric forms that were presented simultaneously (Experiment 1A) and for faces expressing different emotions that were presented sequentially (Experiments 1B). The results of Experiment 2 indicated that the psychophysical procedure used in Experiment 1 is useful for comparing the roles of contingency sensitivity and individual response criteria in the assessment of positive and negative contingencies.

Experiment 3 was designed to evaluate a well-established but poorly understood phenomenon of contingency assessment—the ODE. The results of Experiment 3 indicated that the new contingency assessment task could be used to demonstrate the ODE. That is, most participants assess two statistically equal contingencies as being unequal because they judge the contingency with the higher outcome density, $P(O)$, as being greater than the contingency with the lower outcome density. In contrast to the traditional contingency assessment task, the version of the task that we used permits evaluation of the source of the ODE: Does increasing P(O) modulate the participant's ability to detect the contingency, or does it bias the participant toward reporting that a given contingency is higher than it really is? Consistent with our previous suggestion (Allan et al., 2005), the results of Experiment 3 indicated that increasing P(O) does not affect the participant's ability to detect the relationship between cue and outcome—rather, it biases the participant to report that the contingency is inflated. Inasmuch as the ODE is attributable to the participant's decision criterion, rather than to the sensitivity to the contingency, it is possible that individuals who are apparently immune to the ODE (i.e., depressed individuals) may be distinguished from those displaying the ODE (i.e., nondepressed individuals) on the basis of their adoption of a conservative decision criterion (see Allan et al., 2007).

Table 8
2 × 2 Matrix in Experiment 4

	Response	
ΔP	R_S	R_W
Large	H	M
Small	FA	CR

Note. The large and small ΔP (contingency between cue and outcome) values were .6 and .4, respectively, for all participants except for KS, whose ΔP values were .7 and .3, respectively. R_S = *Strong* response; R_W = *Weak* response; H = hit; M = miss; FA = false alarm; CR = correct rejection.

Table 9
Experiment 4: Payoff Matrices

		Response	
Condition	ΔP	R_S	R_W
Weak	Large	10	−10
	Small	−50	50
Strong	Large	50	−50
	Small	−10	10

Note. The large and small ΔP (contingency between cue and outcome) values were .6 and .4, respectively, for all participants except for KS, whose ΔP values were .7 and .3, respectively.

Experiments 1–3 evaluated the effects of various manipulations on contingency sensitivity and the participants' decision criterion. In Experiment 4, we experimentally manipulated the criterion by the imposition of an explicit payoff structure. The results were consistent with a SDT analysis of the contingency assessment task: For participants whose contingency assessments are affected by the payoffs, the locus of the effect is in the criterion parameter—not in the sensitivity parameter.

When fitting the psychometric functions and when calculating d', we used the simplest SDT model. We assumed that there was a monotonic relationship between mean perceived contingency and actual contingency, that the variability in perceived contingency had a Gaussian distribution, and that this variability was

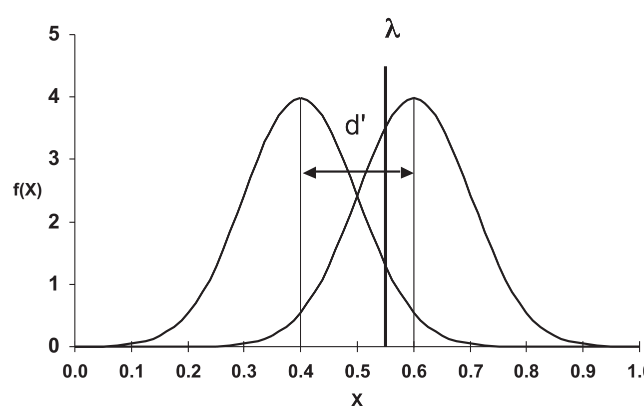

Figure 9. The participant's decision problem for the streamed-trial task with two values of ΔP (cue–outcome contingency). $f(X)$ = probability density; X = subjective contingency, λ = decision criterion value. Line with arrowheads = the difference between the mean values of .4 and .6 (d', or participant's sensitivity).

Table 10
Experiment 4: Values for Each Participant

Participant	d' S	d' W	d' Diff	λ S	λ W	λ Diff
MC	1.56	1.56	0.00	0.37	1.01	−0.64
AB	1.15	1.14	0.01	0.34	0.40	−0.06
XG	1.76	1.64	0.12	0.45	1.23	−0.78
GM	1.60	1.41	0.19	−0.10	1.37	−1.47
KS	2.51	2.50	0.01	1.13	1.04	0.09

Note. d' = participant's sensitivity; λ = decision criteria; S = strong condition; W = weak condition; Diff = difference between the parameter estimates for the two payoff matrices.

constant across ΔP values.[11] The fits of this simple model to the data were good, and there did not appear to be any systematic departures of the fitted functions from the data.

In summary, sign, outcome density, and payoffs influenced the decision parameter (λ) and had little influence on the ability to assess the contingency (σ and d'). Thus, manipulations that have been shown to be decision effects in the psychophysical literature (see Macmillan & Creelman, 2005) have similar effects in the streamed-trial contingency task. Moreover, as in the psychophysical literature, these manipulations have little effect on the participant's sensitivity to the contingencies.

Although there are clear trends in the data, individual differences do exist. Not surprisingly, as in other psychophysical experiments (e.g., Allan, 2002), participants differed in sensitivity and also with regard to the impact of an independent variable on the decision criterion. For example, in Experiment 2, for Participant XG (and only this participant), sign had a clear effect on σ and no effect on λ. One advantage of the streamed-trial procedure is that it allows individual differences to be seen, rather than to be hidden, as is the case with the conventional contingency task, where averaging over participants is the norm.

The SDT approach incorporating the streamed-trial procedure and the method of constant stimuli provides a methodology that would be useful in obtaining a better understanding of conditions such as mood (e.g., Alloy & Abramson, 1979) and age (Mutter & Williams, 2004), which have been shown to affect contingency assessment. Researchers have tended to conclude that these performance differences reflect differences in the ability to detect contingencies. As we (Allan & Siegel, 2002; Allan et al., 2007) and others (e.g., McFall & Treat, 1999; Swets, 1996) have documented, a SDT analysis reveals that effects that have been attributed to sensitivity differences often reflect different decision strategies. Inasmuch as the streamed-trial procedure, in conjunction with the method of constant stimuli, provides independent measures of contingency sensitivity and decision criterion, it provides a methodology that may be of interest to the study of other clinical conditions that affect contingency assessment.

In apparent contrast to the SDT analysis advocated here, most current research in contingency assessment focuses on associative models (see Allan & Tangen, 2005). Such an approach was inspired by Dickinson et al.'s (1984) associative account of contingency assessment. According to an associative account of contingency assessment, a human participant learns over trials to associate cues with outcomes in the same manner as organisms learn to associate conditional and unconditional stimuli in a Pavlovian learning experiment. Associative models specify one process that describes the learning algorithm for the growth of associative strength (V). These models assume that the behavioral response (judged magnitude of contingency) is a monotonic function of V. In recent years, a number of researchers have questioned this assumption (e.g., Allan & Tangen, 2005; Vadillo & Matute, 2007; Vadillo et al., 2005).

It is intriguing to consider the possibility that the two-process SDT model is not an alternative to associative interpretations of contingency assessment but rather that it is a useful way of integrating the acquisition of associations with a separate decision process. A similar suggestion was made some years ago by Schmajuk (1987) in the context of classical conditioning of nonhuman organisms. Consider that on each streamed trial, an associative algorithm (e.g., Bush & Mosteller, 1951; Rescorla & Wagner, 1972) results in a value of associative strength, V. In current conceptualizations of associative learning rules, for fixed parameter values, there is no variability in the value of V. Within the SDT framework, V would be embedded in a noisy background. Thus, for fixed parameter values, there would be variability in V. The x axis in Figures 2 and 9 would be subjective V. On each trial, the subjective value of V would be compared with an internal standard or criterion strength value, λ.

The merging of an associative learning rule with SDT might provide for a better understanding of cue interaction effects that have dominated the contingency assessment literature since the first report by Dickinson et al. (1984). When multiple cues are paired with a common outcome, the cues usually are not rated independently. For example, when two cues, a target cue C_T and a companion cue C_C, are paired with a common outcome, the typical finding is that the rating of the relationship between C_T and the outcome depends on the strength of the relationship between C_C and the outcome. Cue interaction has been shown using a variety of paradigms including two-phase blocking (e.g., Shanks, 1985), overshadowing (e.g., Waldmann, 2001), relative cue validity (e.g., Wasserman, 1990) and one-phase blocking (e.g., Baker, Mercier, Vallee-Tourangeau, Frank & Pan, 1993; Tangen & Allan, 2004). Cue interaction effects have been central to the evaluation of competing theoretical accounts of contingency assessment (for recent reviews, see Allan & Tangen, 2005; De Houwer & Beckers, 2002). For the binary-response streamed-trial task to be a useful alternative to traditional contingency assessment tasks, it must be amenable to the study of cue interaction.

Consider the one-phase blocking paradigm in which one of four cue combinations is possible on a given trial: Both cues may be present ($C_T\ C_C$), one cue may be present and the other absent (C_T $\sim C_C$ or $\sim C_T\ C_C$), or both cues may be absent ($\sim C_T \sim C_C$). For each cue combination, the outcome either occurs (O) or does not occur ($\sim O$), resulting in eight possible cue–outcome combina-

[11] Stevens's psychophysical law (1957) specifies that the relationship between subjective magnitude and physical magnitude is a power function. Weber's law specifies that the standard deviation of the subjective magnitudes is proportional to the mean of the subjective values (see Kling & Riggs, 1971). Our data indicate that the Stevens's exponent for contingency perception is 1.0 (i.e., mean perceived contingency is monotonic with actual contingency) and that Weber's law does not apply to contingency perception (i.e., the standard deviation remains constant with variation in ΔP).

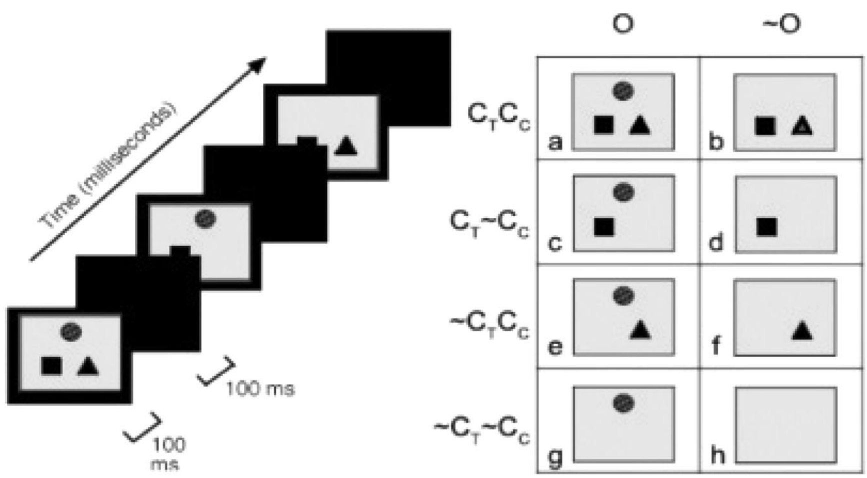

Figure 10. On the left is a schematic illustrating the structure of a streamed trial in the one-phase blocking task. On the right (a–h) are the eight possible cue–outcome combinations in a streamed trial. Triangles and squares are presented in blue and function as target (C_T) and companion cues (C_C), respectively. The circle is presented in red and functions as the outcome (*o*). Both cues may be present ($C_T\ C_C$), one cue may be present and the other absent ($C_T \sim C_C$ or $\sim C_T\ C_C$), or both cues may be absent ($\sim C_T \sim C_C$). For each cue combination, the outcome either occurs (*O*) or does not occur ($\sim O$), resulting in eight possible cue–outcome combinations.

tions. The usual finding is that ratings of C_T depend on the contingency between C_C and the outcome. For example, for a fixed contingency of .5 between C_T and the outcome, ratings of C_T are lower when the contingency between C_C and the outcome is perfect ($\Delta P = 1.0$) than when there is no contingency between C_C and the outcome ($\Delta P = .0$) (Tangen & Allan, 2004). A central issue in the contingency assessment literature is whether the effect of C_C on ratings resides in the assessment of the contingency between C_T and the outcome or whether it resides in the decision process.

The streamed-trial task can be readily modified to address this question. The eight cue–outcome combinations are presented in Figure 10. Triangles and squares would be presented in blue and function as target and companion cues, respectively. The circle would be presented in red and function as the outcome. On a streamed trial, the companion cue could have a perfect relationship with the outcome ($companion_{1.0}$) or could be unrelated to the outcome ($companion_{0.0}$), and the ΔP value for the target cue could vary between 0 and 1.0 (as in Experiment 1A). At the end of a stream, the participant would be signaled to make a binary response (*Weak* or *Strong*) either about the target's relationship to the circle or about the companion's relationship to the circle. Psychometric functions for the target cue could be generated for each companion cue contingency (0.0 and 1.0). If cue interaction is due to cue competition during learning, the two functions should differ in slope—the slope should be steeper for $companion_{0.0}$ (no cue competition and accurate assessment of the contingency) than for $companion_{1.0}$ (cue competition and inaccurate assessment of the contingency). In contrast, if cue interaction is a criterion effect, then the two functions should have the same slope and should differ in *PSE*.

We previously suggested that there were similarities between SDT and the associative model developed by Rescorla and Wagner (1972). As Siegel & Allan (1996) noted:

> . . . signal detection theory has been found to be applicable to many areas in addition to the limited domain in which it was first developed; that is, it has become a way to think about issues in areas other than psychophysics . . . . Similarly, the Rescorla–Wagner model has provided a basis for thinking about issues in areas other than Pavlovian conditioning. There are only a precious few such inspirational contributions in experimental psychology (p. 319).

We suggest that the streamed-trial task, as implemented in the present experiments, provides a methodology for generating the type of data that could lead to the integration of these two "inspirational" models.

References

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15,* 147–149.

Allan, L. G. (2002). The location and interpretation of the bisection point. *Quarterly Journal of Experimental Psychology Section B, 55,* 43–60.

Allan, L. G., & Siegel, S. (2002). A signal detection analysis of the placebo effect. *Evaluation & the Health Professions, 25,* 410–420.

Allan, L. G., Siegel, S., & Hannah, S. (2007). The sad truth about depressive realism. *Quarterly Journal of Experimental Psychology, 60,* 482–495.

Allan, L. G., Siegel, S., & Tangen, J. M. (2005). A signal detection analysis of contingency data. *Learning & Behavior, 33,* 250–263.

Allan, L. G., & Tangen, J. M. (2005). Judging relationships between events: How do we do it? *Canadian Journal of Experimental Psychology, 59,* 22–27.

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General, 108,* 441–485.

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

Beckers, T., De Houwer, J., Pineño, O., & Miller, R. R. (2005). Outcome additivity and outcome maximality influence cue competition in human causal learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 238–249.

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review, 58,* 313–323.

Crump, M. J. C., Hannah, S. D., Allan, L. G., & Hord, L. K. (2007). Contingency judgments on the fly. *Quarterly Journal of Experimental Psychology, 60,* 753–761.

De Houwer, J., & Beckers, T. (2002). A review of recent developments in research and theories on human contingency learning. *Quarterly Journal of Experimental Psychology, 55B,* 289–310.

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

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Killeen, P. R., Fetterman, J. G., & Bizo, L. A. (1997). Time's causes. In C. M. Bradshaw & E. Szabadi (Eds.), *Time and behaviour: Psychological and neurological analyses* (pp. 79–131). Amsterdam: Elsevier Science.

Kling, J. W., & Riggs, L. A. (1971). Woodworth & Schlossberg's experimental psychology (pp. 47–86). New York: Holt, Rinehart and Winston.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed., pp. 71–72). Mahwah, NJ: Erlbaum.

Maldonado, A., Catena, A., Cándido, A., & García, I. (1999). The belief revision model: Asymmetrical effects of noncontingency on human covariation learning. *Animal Learning & Behavior, 27,* 168–180.

McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology, 50,* 215–241.

McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology, 54,* 33–61.

Mutter, S. M., & Williams, T. W. (2004). Aging and the detection of contingency in causal learning. *Psychology and Aging, 19,* 13–26.

Perales, J. C., Catena, A., Shanks, D. R., & González, J. A. (2005). Dissociation between judgments and outcome–expectancy measures in covariation learning: A signal detection theory approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1105–1120.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68,* 29–46.

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. Prokosy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Schmajuk, N. A. (1987). Classical conditioning, signal detection, and evolution. *Behavioural Processes, 14,* 277–289.

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

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

Spellman, B. A., Price, C. M., & Logan, J. M. (2001). How two causes are different from one: The use of (un)conditional information in Simpson's paradox. *Memory & Cognition, 29,* 193–208.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64,* 153–181.

Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers.* Mahwah, NJ: Erlbaum.

Tangen, J. M., & Allan, L. G. (2004). Cue-interaction and judgments of causality: Contributions of causal and associative processes. *Memory & Cognition, 32,* 107–124.

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

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

Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review, 8,* 600–608.

Wasserman, E. A. (1990). Attribution of causality to common and distinctive elements of compound stimuli. *Psychological Science, 1,* 298–302.

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

Wickens, T. D. (2002). *Elementary signal detection theory* (pp. 195–198). New York: Oxford University Press.

Winman, A., & Gredebäck, G. (2006). Inferring causality assessments from predictive responses: Cue interaction without cue competition. *Quarterly Journal of Experimental Psychology, 59,* 28–37.

Received March 13, 2007
Revision received October 11, 2007
Accepted October 15, 2007 ■

Correction to Condry and Spelke (2008)

In the article, "The Development of Language and Abstract Concepts: The Case of Natural Number," by Kirsten F. Condry and Elizabeth S. Spelke (*Journal of Experimental Psychology: General,* 2008, Vol. 137, No. 1, p. 22), the DOI for the supplemental materials was printed incorrectly. The correct DOI is as follows:
http://dx.doi.org/10.1037/0096-3445.137.1.22.supp

DOI: 10.1037/0096-3445.137.2.243