PSYCHOPHYSICS OF CAUSALITY: DETECTING CONTINGENCIES IS LIKE DETECTING SIGNALS

Lorraine G. Allan, Samuel D. Hannah, Mathew J. C. Crump, & Shepard Siegel
Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton ON, Canada, L8S 4K1
allan@mcmaster.ca

Abstract

The contingency assessment situation and the signal detection situation are similar in that the information on which the decision is based is uncertain. Nevertheless the two research endeavours have progressed independently, each with its own traditions and each motivated by different theoretical perspectives. Recently, some researchers have integrated these two lines of research and have demonstrated the value of applying a signal detection analysis to contingency assessment. In the present paper, we describe a new method, the streamed-trial procedure, that is better suited to the study of contingency detection and discrimination than is the traditional contingency judgment task, and we present data from experiments using this new procedure. We also discuss future research with the streamed-trial procedure that could provide a better understanding of intriguing findings in the contingency judgment literature such as "depressive realism".

We must often make a decision even though the information we have is ambiguous or uncertain. One such situation is illustrated by a patient being treated by an allergist. The patient sometimes, but not always, develops hives after eating strawberries. Moreover, the patient sometimes develops hives even when strawberries are not eaten. Although the relationship between eating strawberries and developing hives is uncertain, the allergist must decide whether or not to recommend that the patient stop eating strawberries. Another type of ambiguous situation is illustrated by the task confronted by the radiologist. The radiologist must decide whether or not an x-ray indicates the presence of lung cancer. The signals seen in the x-ray are ambiguous, some consistent with lung cancer and others inconsistent with lung cancer. Even though the correct diagnosis is unclear, the radiologist must decide whether or not to recommend treatment. Despite the obvious similarities between the tasks, they have been treated quite differently. The allergy task has often been used by researchers interested in contingency assessment; that is, how humans judge that a cue (strawberry ingestion) imperfectly signals an outcome. The cancer task has often been used by researchers interested in signal detection; that is, how humans make decisions about the presence of a signal (cancer symptoms) in a noisy background.

The contingency assessment situation and the signal detection situation are similar in that the information on which the decision is based is uncertain. Yet research concerned with contingency assessment and research concerned with signal detection have progressed independently, each with its own traditions and each motivated by different theoretical perspectives and models. Recently, Allan, Siegel, and Tangen (2005) and Perales, Catena, Shanks, and González (2005) demonstrated the value of applying a psychophysical analysis to contingency assessment. In the present paper, we describe a new procedure that is better suited to the study of contingency detection and discrimination than is the traditional contingency judgment task.

Table 1: The 2x2 matrix for the cue-outcome pairings in a contingency assessment task.

	O	~O
C	a	b
~C	c	d

On each trial of the traditional contingency task, a cue either is presented (C) or is not presented (~C), and then the outcome either occurs (O) or does not occur (~O). For example, in the previously-described allergy task, the observer is shown whether the patient consumed or did not consume strawberries, and then is shown whether the allergic reaction occurred or did not occur. After a series of trials on which each of the four cue-outcome combinations are presented with a pre-defined probability, the observer is asked to rate the strength of the relationship between the cue and the outcome. Table 1 presents the 2 x 2 matrix for the cue-outcome pairings. 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 measure of the contingency between the cue and the outcome is ΔP (Allan 1980),

$$\Delta P = P(O \mid C) - P(O \mid \sim C) = \frac{a}{a+b} - \frac{c}{c+d}. \tag{1}$$

Many early studies of contingency assessment concentrated on exploring the relationship between the observer's ratings of the contingency between the two binary variables and ΔP. While these studies often reported a high correlation between rating judgment and ΔP, systematic departures from ΔP were frequently noted (see Allan, 1993; Shanks, 1993). One such departure has been termed the outcome density effect, ODE. For a fixed ΔP, ratings of contingency often are not constant but increase with the probability of the outcome, P(O),

$$P(O) = \frac{a+c}{a+b+c+d}. \tag{2}$$

Most explanations of the ODE have placed the effect in the perception of the contingency: P(O) affects the observer's ability to detect the actual contingency. The psychophysical analysis in Allan et al. (2005) suggested otherwise. This analysis was based on the prediction responses that observers are often asked to make on each trial. After the cue is presented or not, the observer is asked to predict whether the outcome will occur (Y) or will not occur (~Y), and then the outcome information is presented. Originally, these trial prediction responses were inserted into the trial sequence as a means of keeping the observer engaged in the task, and usually were not reported or analyzed. Allan et al. applied a signal-detection theory (SDT) analysis to these trial prediction responses. A 2 x 2 matrix can be constructed from the trial prediction responses as shown in Table 2, resulting in two conditional probabilities, P(Y|C) and P(Y|~C). Allan et al. plotted these conditional probabilities to generate ROC curves. They found that for a constant ΔP value performance moved along an iso-sensitivity ROC as P(O) changed. That is, P(O) had no effect on sensitivity (d′) but did systematically affect the observer's criterion (λ) for predicting that the outcome would occur. Allan et al. concluded that the ratings at the end of a block of trials reflected the shift in λ and did not provide accurate information about the perception of the contingency. The ODE is a bias effect: P(O) affects how observers respond but not their ability to detect the actual contingency.

Table 2: The 2x2 matrix for the cue-response pairings in a contingency assessment task.

	Y	~Y
C	a	b
~C	c	d

In their experiments, Perales et al. (2005) varied P(C) rather than P(O),

$$P(C) = \frac{a + b}{a + b + c + d}. \tag{3}$$

and showed that P(C) affected bias and not perception. Moreover, they introduced a payoff matrix into their contingency task and found that payoff manipulations also affected bias and not the detection of contingency.

The research and analyses presented in Allan et al. (2005) and Perales et al. (2005) are supportive of the hypothesis that detecting contingencies is like detecting signals. However, the traditional contingency task is not well suited to psychophysical analyses. In particular, it generally takes about 5 min to present enough trials to define a particular contingency value to an observer; thus, little data can be obtained in a session. Researchers are faced with investing an inordinate amount of time to collect relatively few estimates of performance across different conditions. For example, it is often necessary to run large numbers of observers in order to ensure that ratings are statistically reliable. Furthermore, because each observer contributes a small number of ratings over the course of an entire experiment, researchers are often forced to conduct less powerful between-subject designs. Indeed, there would be little advantage to running multiple session, within-subject designs, as observers would need to spend an unreasonable number of hours in the lab in order for researchers to obtain reliable, within-subject estimates of performance.

Our new procedure was constructed to avoid previous limitations due to task length. We reduced the length of a block of trials from several minutes to a few seconds using a method involving the rapid sequential presentation of cue-outcome pairs, telescoping an entire block of trials into a single streamed-trial. A presentation stream is depicted schematically in Figure 1. The cue and the outcome are coloured geometric forms presented on a grey frame (6.4 cm in height x 5 cm in width) in the centre of a black monitor screen. Each 100 msec presentation consists of one of the four cue-outcome combinations (see inset), and presentations are separated by a 100 msec black screen. The cue was a blue square (1.6 cm in height and width). When it was presented it was centered at the bottom of the frame. The outcome was a red circle (1.6 cm in diameter). When it was presented it was centered at the top of the frame. A contingency value is defined by a presentation stream of cue-outcome combinations. For example, if each of the four cue-outcome combinations occurred equally often, the ΔP value on that streamed-trial would be zero.

Hord (2006) validated the streamed-trial procedure by replicating two of the central findings in the contingency judgment literature. In her experiments, the observer provided a rating at the end of each streamed-trial. She varied ΔP and P(O). She showed that observer's ratings of contingency correlated highly with the programmed ΔP values. She also showed that the ratings were influenced by P(O); that is, she demonstrated the ODE with the streamed-trial task.

Another limitation of the traditional contingency task is that it is not designed for making direct comparisons between different contingency values. For example, it does not provide a means for determining how large ΔP has to be in order to be detected as different from zero, or how different two ΔP values have to be in order to be discriminated, or whether the discriminability of positive contingencies differs from that of negative contingencies. Our streamed-trial procedure allows us to address these issues. Moreover, our streaming procedure can be readily adapted so that it is amenable to analyses that provide separate estimates of perceptual and response bias parameters. In the present paper, we report data from four psychophysical experiments using the streamed-trial procedure.

Experiment 1

In this experiment, 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 observer's task is to make a binary decision. For example, in loudness discrimination, one of k intensities is presented on each trial, and the observer's task is to categorize the perceived loudness as either "loud" or "soft". In Experiment 1, on each trial we presented one of 11 ΔP values and asked the observer to respond "strong" or "weak".

Method

Four observers (O1, O2, O3, O4) participated. O4 was an undergraduate research assistant in the laboratory. The other three observers were graduate students who were paid for their participation.

In Experiment 1, there were 60 presentations in a streamed-trial, with a total duration of approximately 12 sec. At the end of each streamed-trial, the observer was required to make a binary decision. Two clickable buttons, one labeled "weak" (a R_W response) and one labeled "strong" (a R_S response) appeared on the screen. The observer's task was to select the button which best represented the strength of the contingency on that trial.

There were 11 values of ΔP ranging from 0.0 to 1.0 in increments of 0.1. For all contingencies, $P(O) = .5$ and $P(C) = .5$. Each of the 11 values was presented four times in a randomized order during each block of trials. A session consisted of five blocks, resulting in 20 presentations of each of the 11 ΔP values. Each observer participated for 10 sessions.

Results and Discussion

Figure 2 displays the probability of a strong response, $P(R_S)$, as a function of ΔP averaged over the four observers. The streamed-trial procedure clearly produces orderly data. Gaussian psychometric functions were determined for each observer. We assumed that the exponent of the psychophysical function was 1.0 – i.e., that the mean subjective contingency was equal to the presented contingency. We also assumed that the variance was independent of the mean – i.e., that the variance was constant for the 11 ΔP values. Observers did not always use the full $P(R_S)$ range, and therefore in fitting the functions we did not impose the restriction that the psychometric function asymptote at the lower and upper bounds of $P(R_S)$, 0 and 1.0 respectively. The standard deviation (σ) of the psychometric function provides a measure of the observer's sensitivity to discriminating among the contingencies. The mean of the function defines the value of ΔP at $P(R_S) = .5$ and is often referred to as the point of subjective equality (PSE). The PSE can be interpreted as a measure of the observer's criterion or bias for responding R_S. The σ values and the PSEs of the individual psychometric functions are shown in Table 3, and the mean function is plotted in Figure 2. The PSE values indicate that the observers varied in the placement of their criterion for R_S. O1 and O2 tended to be somewhat conservative (PSE > .5) whereas O3 and O4 tended to be somewhat liberal (PSE < .5).

Table 3: σ and PSE values in Experiment 1.

O	σ	PSE	d'
1	.12	.56	.83
2	.15	.55	.67
3	.26	.44	.26
4	.25	.40	.25

The σ of the psychometric function can be converted to the SDT sensitivity measure d',

$$d' = \frac{\Delta P_j - \Delta P_i}{\sigma}, \tag{4a}$$

and for a difference of .1

$$d' = \frac{.1}{\sigma}. \tag{4b}$$

The d' values based on Eq. 4b are available in Table 3 for each observer. The d' values for discriminating between two ΔP values which differ by .1 range from .38 to .83 across the four observers.

Experiment 2

In Experiment 2, we compared psychometric functions for positive and negative contingencies.

Method

Four observers (O2, O3, O5, O6) participated. O2 and O3 participated in Experiment 1, and the other two were new graduate students. There were 11 values of ΔP ranging from –1.0 to +1.0 in increments of 0.2. The observers were instructed to base their decision on the absolute size of the contingency. In all other respects, the procedure was the same as in Experiment 1.

Results and Discussion

Figure 3 displays $P(R_S)$ as a function of ΔP averaged over the four observers. Separate psychometric functions were fit for the positive and negative contingencies. The σ values and the PSEs are shown in Table 4 for each observer. Overall, the sign of the contingency affected both the PSE and σ. For each observer, the absolute value of the positive PSE was larger than the absolute value of the negative PSE. The observer was more likely to respond

R_S to a negative contingency than to a positive contingency. For three of the observers, σ for the positive ΔP values was smaller than for the negative ΔP values suggesting greater sensitivity in discriminating among positive contingencies than among negative contingencies.

Table 4: σ and PSE values in Experiment 2.

O	σ		PSE	
	positive	negative	positive	negative
2	.18	.16	.59	.48
3	.19	.25	.62	.44
5	.14	.24	.64	.49
6	.17	.28	.70	.61

Experiment 3

If ODE were a bias effect, as suggested by Allan et al. (2005), then one would expect P(O) to affect the PSE, but not σ, of the psychometric function. In Experiment 3, we varied outcome density.

Method

Four observers (O2, O5, O7, O8) participated. O2 and O5 had participated previously and the other two were new graduate students. A stream of 80 presentations defined a value of ΔP. P(C) = .5 and P(O) was either .3 or .7. Table 5 shows the values of ΔP, P(O|C), and P(O|~C) and also the frequencies in cells a and c of the 2 x 2 contingency matrix for each value of P(O). With 80 presentations in a stream, P(C) = .5, and P(O) values of .3 and .7 the range of possible positive ΔP values is constrained to values ≤ .6. Seven values of ΔP, ranging from 0.0 to 0.6 in increments of 0.1, were used.

Table 5: ΔP and P(O) values in Experiment 3.

P(O)	ΔP	P(O|C)	P(O|~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

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 trials. A session consisted of five blocks resulting in 20 presentations of each of the seven ΔP values. Each observer participated for 20 sessions, 10 with each value of P(O) in a randomly determined order.

Results and Discussion

Figure 4 displays P(R_S) as a function of ΔP averaged over the four observers. Separate psychometric functions were fit for the two values of P(O). The σ and PSE values are available in Table 6 for each observer. The σ values are similar for the two values of P(O) and do not vary systematically across observers. In contrast, for every observer the PSE is smaller for P(O) = .7 than for P(O) = .3. Increasing P(O) increases the tendency for the observer to categorize the relationship as "strong". While R_S was more likely when the outcome occurred frequently, the ability to discriminate among the contingencies remains constant for the two values of P(O). Increasing P(O) does not affect the observer's ability to detect the strength of the relationship between the cue and the outcome.

Table 6: σ and PSE values in Experiment 3.

O	σ		PSE	
	P(O) = .3	P(O) = .7	P(O) = .3	P(O) = .7
2	.15	.16	.40	.35
5	.15	.21	.19	.15
7	.19	.17	.24	.18
8	.15	.13	.33	.27

Experiment 4

We noted earlier that Perales et al. (2005) had manipulated payoff structure in the traditional contingency task, and found that payoffs affected bias but not sensitivity. In Experiment 4, we examined the role of payoff structure on performance in streamed-trial task.

Method

The four observers from Experiment 3 participated. There were two values of ΔP, .4 and .6, each defined by a stream of 60 presentations. As in the previous experiments, the observer made either an R_S or an R_W response on each trial. Each ΔP value was presented 14 times in a randomized order during each block of 28 trials. A session consisted of five blocks.

Table 7: 2 x 2 matrix in Experiment 4.

ΔP	Response	
	R_S	R_W
.6	H	M
.4	FA	CR

The 2 x 2 matrix relating the two ΔP values to the two response categories is shown in Table 7. There are two types of correct responses: R_S to ΔP = .6 which we will call a Hit (H) and R_W to ΔP = .4 which we will call a Correct Rejection (CR). There are also two types of errors: R_S to $\Delta P = .4$ which we will call a False Alarm (FA) and R_W to $\Delta P = .6$ which we will call a Miss (M). Observers won points for correct responses and lost points for errors. The two payoff matrices are illustrated in Table 8. In the Weak (W) condition, there were greater gains and smaller losses for R_W than for R_S. CR earned 50 points, while H earned only10 points. Moreover, FA lost 50 points, while M lost only 10 points. These payoffs were reversed in the Strong (S) condition, where H earned 50 points and CR earned only 10, while M cost 50 points, and FA cost only 10. Each observer participated for 20 sessions, 10 with each payoff matrix in a randomly determined order.

Table 8: Payoff Matrices in Experiment 4.

Condition	ΔP	Response	
		R_S	R_W
Weak	.6	10	-10
	.4	-50	50
Strong	.6	50	-50
	.4	-10	10

Results and Discussion

The probability of a hit, P(H), and the probability of a false alarm, P(FA), was determined for each payoff condition. Converting these probabilities to z-scores allows values of d'

$$d' = Z(FA) - Z(H) \tag{5}$$

and λ

$$\lambda = Z(FA) \tag{6}$$

to be estimated. These estimates are shown in Table 9 for each observer under each payoff condition. Payoffs had little effect on d' but did affect λ. The criterion was more liberal for R_S when the payoffs favoured H and FA compared to when the payoffs favoured CR and M.

Table 9: d' and λ in Experiment 4.

O	d'		λ	
	S	W	S	W
2	1.56	1.56	.37	1.01
5	1.15	1.14	.34	.40
7	1.60	1.41	-.10	1.37
8	1.76	1.64	.45	1.23

General Discussion

Our streamed-trial procedure generates orderly psychometric functions which indicate that observers are able to discriminate ΔP differences as small as 0.1. We found that negative contingencies were more difficult to discriminate than were positive contingencies and that observers are more likely to categorize a negative contingency as strong relative to a positive contingency. Of special interest is the clear evidence that outcome density and payoffs have large and systematic effects on the placement of the criterion. Our data establish that manipulations that have been shown to be criterion effects in the psychophysical literature have similar effects on contingency discrimination performance. Moreover, as in the psychophysical literature, these variables have little effect on ability to discriminate differences in ΔP.

In analyzing our data and fitting our functions, we made a number of assumptions. We assumed a monotonic relationship between perceived contingency and actual contingency (i.e., that the exponent of the psychophysical function was 1.0). We also assumed that the variability in perceived contingency had a Gaussian distribution and was constant across ΔP values. When fitting the psychometric functions with these assumptions we were able to account for at least 98% of the variance (usually 99%), and there do not appear to be any systematic departures of the fitted functions from the data. Interestingly, Weber's Law does not appear to hold for contingency discrimination; we were able to fit the data with a constant value of σ which was independent of the mean. Future research needs to be directed at evaluating these assumptions.

Most importantly, our streaming-trial procedure provides a tool for isolating perceptual effects from response effects in contingency assessment. It could be particularly useful in providing an understanding of "depressive realism".

Depressive Realism

Alloy and Abramson (1979) concluded that depressed college students are more accurate than nondepressed college students in the traditional contingency assessment task (see Allan, Siegel, & Hannah, in press for a recent review). In particular, nondepressives were influenced by manipulations that affect the salience of the outcome, especially outcome density. In contrast, depressives displayed little or no ODE. This apparent knack for depressives 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 depressives as "sadder but wiser".

It is not entirely clear in the writings of Alloy, Abramson, and their colleagues whether they thought the locus of the difference in the judgments of depressives and nondepressives to be in the perception of the contingency or in the response criterion. On the one hand, they talk about nondepressives displaying an "illusion", and illusions are usually considered to be perceptual. On the other hand, they talk about "biases" and suggest that both depressed and nondepressed are susceptible to biases, but in opposite directions, "with nondepressives distorting environmental information optimistically and depressives distorting it pessimistically." (Alloy & Abramson, 1988, p. 255). Understanding depressive realism requires some precision in distinguishing whether mood differences are the result of perceptual differences or response differences.

In fact there are psychophysical data that document response criterion differences between depressives and nondepressives in a variety of quite diverse tasks, such as short-term memory, gustatory sensitivity, flicker fusion, and judgment of line length (for a summary see Allan et. al., in press). In these studies, depressives did not differ from nondepressives with regard to sensitivity. Rather depressives and nondepressives adopted different criteria.

We showed in Experiment 3 that the ODE is a criterion effect. Increasing the outcome density does not affect the observer's perception of the relationship between the cue and the outcome. Rather, for any ΔP value, increasing P(O) increases the tendency for the observer to categorize the relationship as "strong". It might be said that, when the outcome density is increased, the typical observer displays irrational optimism. When outcome density increases, they increase their bias for saying that the cues and the outcomes are related. Depressed individuals are not optimistic individuals. Compared to nondepressives, depressives are “nay-sayers”. They must be very confident that a relationship exists before they accede to responding in a manner that indicates that a relationship exists. As noted above, there is evidence that the performance differences between depressed individuals and nondepressed individuals on a variety of tasks is located in the decision process, not the perception process. Similarly, we suggest that the differences in depressive and nondepressive responding to outcome density and to reward feedback is a manifestation of the depressed individual’s relative (compared to the nondepressed individual) pessimism. Depressives may be sadder but they are *not* wiser; rather, so-called “depressive realism” results from the bias of depressives to say “no”.

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. (1993). Human contingency judgments: Rule-Based or associative? *Psychological Bulletin*, *114*, 435-448 .

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

Allan, L. G, Siegel, S., & Hannah, S. (in press). The sad truth about depressive realism. *Quarterly Journal of Experimental Psychology*.

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General, 108*, 41-485.

Alloy, L. B., & Abramson, L. Y. (1988). Depressive realism: Four theoretical perspectives. In, L. B. Alloy (Ed.), *Cognitive processes in depression*, 223-265. New York: Guilford Press.

Hord, L. (2006). Exciting new method of exploring contingency judgments. Bachelor’s Thesis, Department of Psychology, Neuroscience and Behaviour, McMaster University.

Perales, J. C., Catena, A., Shanks, D. R., and Jose´ A. 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.

Shanks, D. R. (1993). Human instrumental learning: A critical review of data and theory. *British Journal of Psychology*, 84, 319-354.

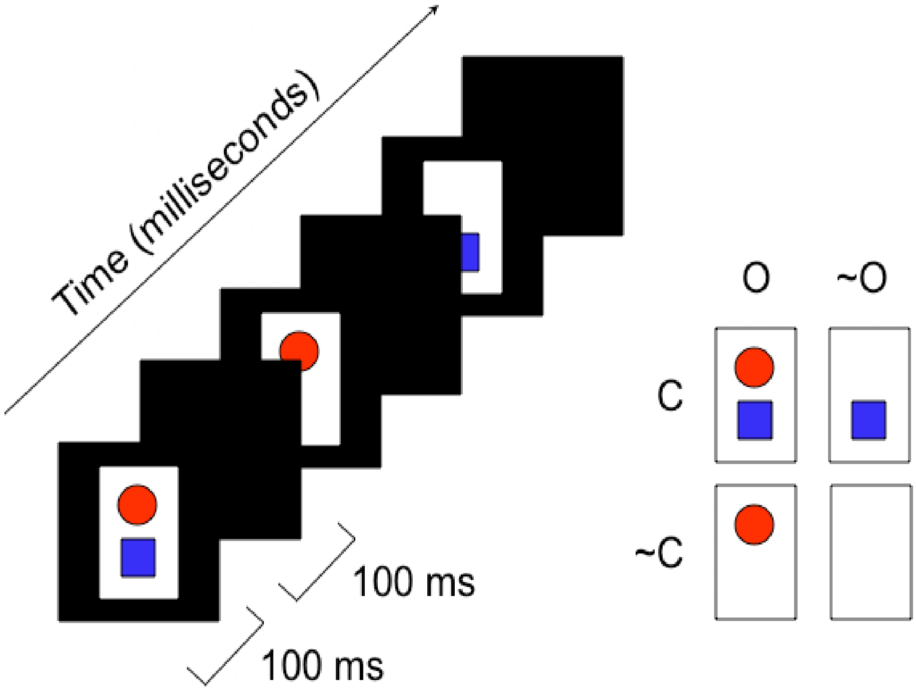

Figure 1: Schematic of a presentation stream.

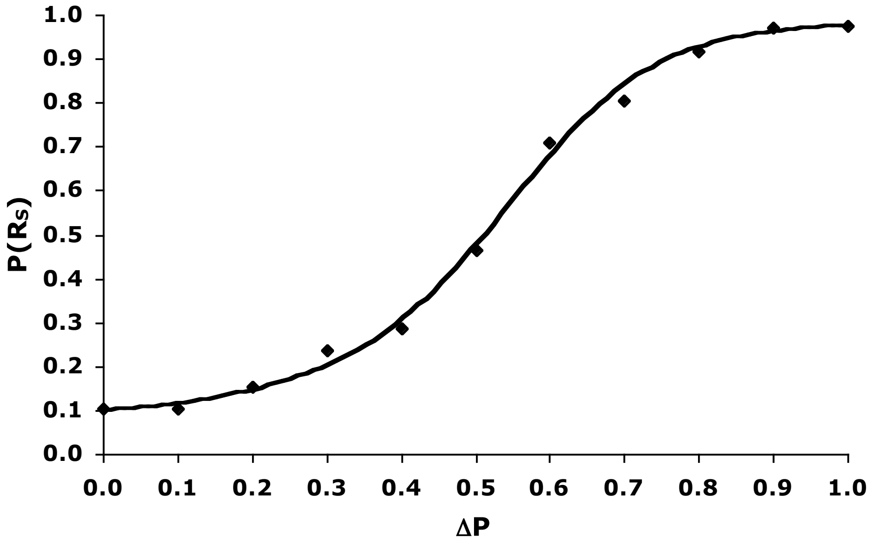

Figure 2. Mean $P(R_S)$ as a function of ΔP in Experiment 1.

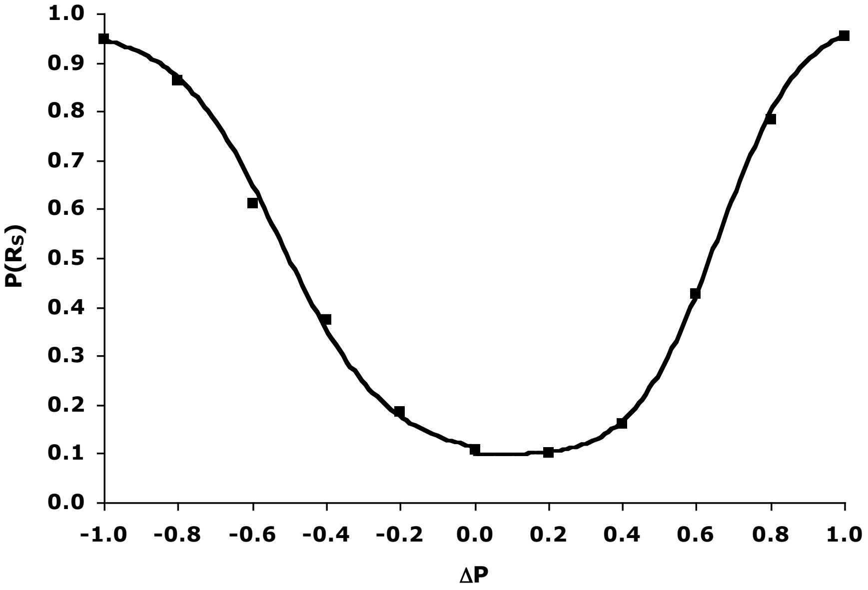

Figure 3. Mean $P(R_S)$ as a function of ΔP in Experiment 2.

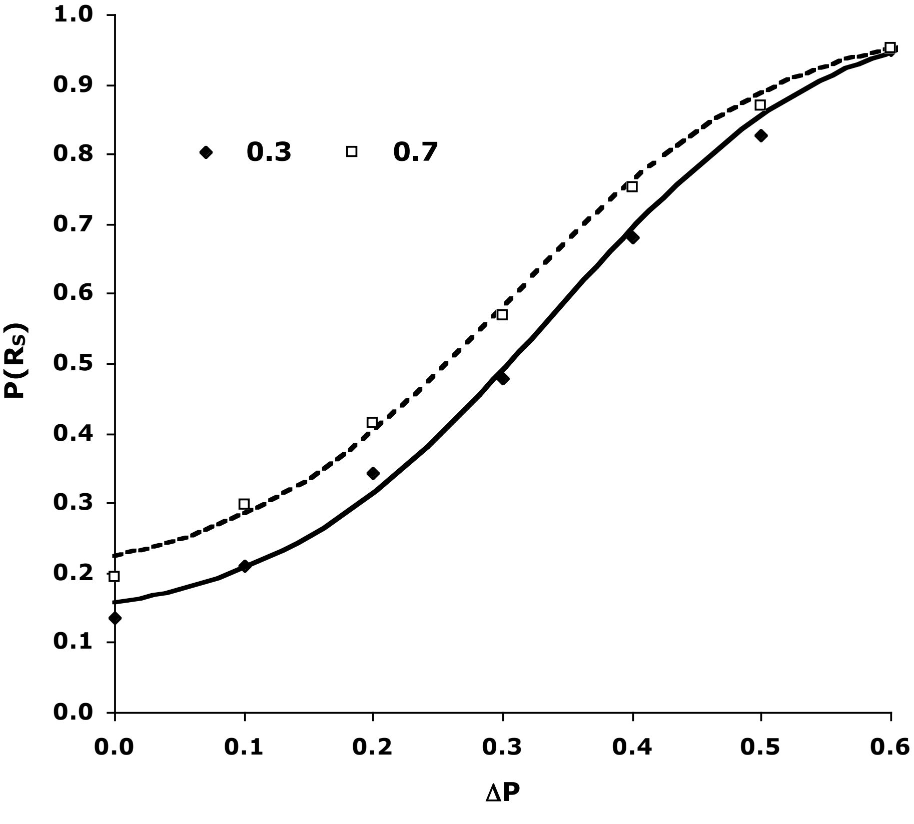

Figure 4. Mean $P(R_S)$ as a function of ΔP for each value of $P(O)$ in Experiment 3.