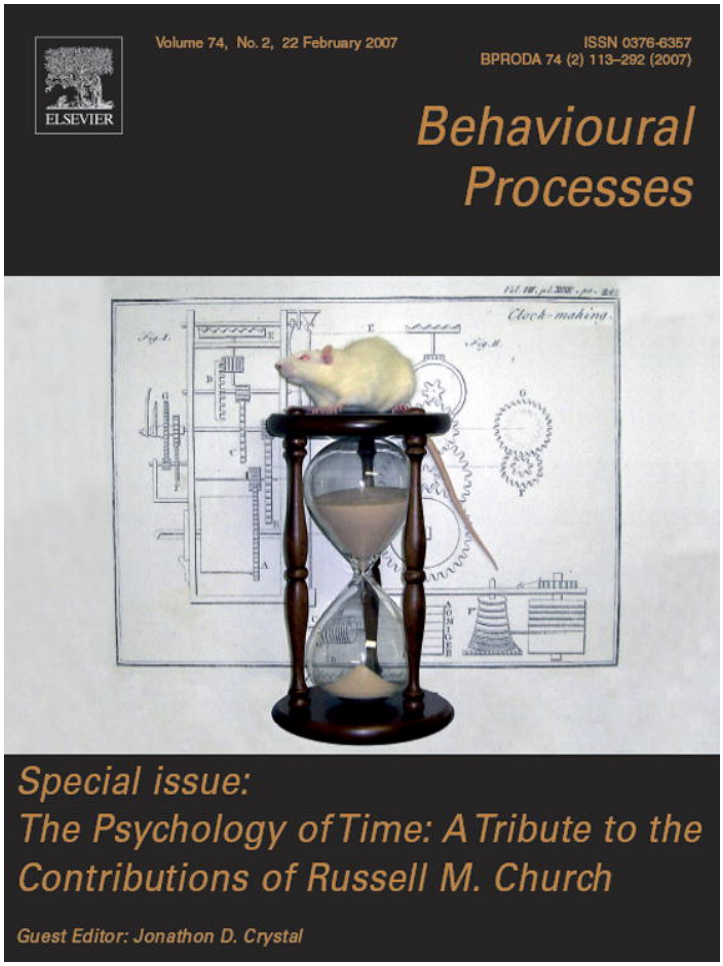

Available online at www.sciencedirect.com
ScienceDirect
Behavioural Processes 74 (2007) 265–273
BEHAVIOURAL PROCESSES
www.elsevier.com/locate/behavproc

The consequences of surrendering a degree of freedom to the participant in a contingency assessment task

Samuel Hannah, Lorraine G. Allan *, Shepard Siegel

Department of Psychology, Neuroscience, & Behaviour, McMaster University, Hamilton, Ontario, Canada L8S 4K1

Received 15 May 2006; received in revised form 18 September 2006; accepted 20 September 2006

Abstract

Many studies of contingency judgments have used a task in which, on each trial, the participant is free either to respond or not to respond, and an outcome may, or may not, be presented. Typically, the experimenter specifies a nominal value for the contingency between responding and outcome, but the actual values of a variety of variables experienced by a particular participant depend on that participant's frequency of responding. The results of computer simulations of various strategies for implementing the contingency manipulation, and the results of an experiment, indicate that the same nominal contingency value will lead to considerable variability in the actual contingency experienced by participants. Moreover, nominal contingency manipulations are confounded with the probability that the subject experiences an outcome. While researchers might be aware of these issues, not enough attention has been paid to their potential impact.

Keywords: Contingency assessment; Active task; Percentile schedules

> "Darwin's legacy includes evidence for impressive continuity of both body and mind among animal species, and experimental psychologists have attempted to identify principles of generality." (Church, 1993, p. 171)

There is considerable interest in the study of contingency assessment. Some investigators have applied the contingency assessment task to understanding medical diagnoses (e.g., Allan et al., 2005) and depression (e.g., Alloy and Abramson, 1979; Allan et al., in press). Others have been interested in theoretical analyses of the task (e.g., Allan and Tangen, 2005; De Houwer and Beckers, 2002). Contingency assessment is also an area that has drawn the attention of researchers sympathetic to Church's (1993) argument for the mutual benefits that human and nonhuman research offer one another. Researchers, following the lead of Dickinson et al. (1984) realized that contingency assessment tasks (studied in humans) are structurally similar to traditional learning tasks (typically studied in nonhuman animals), and theoretical analyses of these learning tasks may be profitably applied to understanding contingency assessment (see Allan, 1993). The purpose of this article is to elaborate the implications of one facet of learning research that has not been incorporated into the contingency assessment literature.

Operant learning researchers have noted that the manipulation of contingencies of reinforcement is complicated by the fact that the participant – not the experimenter – controls the occurrence of the response. Thus, the experimenter's objective of presenting (or not presenting) a reinforcement when a response does (or does not) occur is a challenge. Contingency assessment researchers have used similar manipulations but generally have been oblivious to the challenge.

Contingency assessment tasks can be divided into two categories, passive and active. Cues predict the presence (or absence) of outcomes in the passive task, whereas actions produce (or preclude) outcomes in the active task. An example of a passive task is the pairing of the consumption of a food (e.g., strawberries) with the appearance of an allergic reaction. On each trial, the participant is shown whether a hypothetical patient consumed strawberries or did not consume strawberries, and then is shown whether the allergic reaction occurred or did not occur. The participant's task is to rate the strength of the relationship between the consumption of strawberries and the occurrence of the allergic reaction. In an active task, the participant has the option of making, or not making, a designated response on each trial.

* Corresponding author. Tel.: +1 905 525 9140; fax: +1 905 529 6225.
E-mail addresses: hannahsd@mcmaster.ca (S. Hannah), allan@mcmaster.ca (L.G. Allan), siegel@mcmaster.ca (S. Siegel).

0376-6357/$ – see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.beproc.2006.09.007

Table 1
The 2 × 2 matrix for the input–outcome pairings in the contingency task

	O	$\sim O$
I	a	b
$\sim I$	c	d

The letters in the cells (a, b, c, d) represent the joint frequency of occurrence of the four input–outcome combinations in a block of trials.

For example, on each trial the participant decides to press a button or not to press a button, and then a light bulb is illuminated or is not illuminated. The participant's task is to rate how much control he or she has over the illumination of the light. This active task most commonly involves discrete trials. In another version of the active task, often referred to as the free-operant procedure, trials are not delineated (e.g. Wasserman et al., 1983). A critical analysis of inherent problems with the free-operant task has previously been provided (e.g., Buehner and May, 2003). In the present paper, we restrict our discussion to the discrete-trial version of the active task not only because of greater complexities confronting the free-operant version, but also because much of the research uses the discrete-trial format.

Table 1 presents the 2 × 2 contingency matrix applicable to both task-types. The input variable (the cue in the passive task or the response in the active task) either occurs (I) or does not occur ($\sim I$), and the outcome variable (the allergic reaction in the example of the passive task or the illumination of the bulb in the example of the active task) either occurs (O) or does not occur ($\sim O$). The passive and active tasks differ with regard to who has control over the input variable. In the passive task, the experimenter determines how often the cue is presented, whereas in the active task the participant determines how often the response is made. The letters in the cells of the matrix (a, b, c, d) represent the joint frequency of occurrence of the four input-outcome combinations. The probability of the input:

$$P(I) = \frac{a+b}{a+b+c+d} \tag{1}$$

and the probability of the outcome:

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

and the contingency between the input and the outcome:

$$\Delta P = P(O|I) - P(O|\ I) = \frac{a}{a+b} - \frac{c}{c+d} \tag{3}$$

can be varied in any experiment. However, $P(I)$, $P(O)$, and ΔP cannot be varied independently of each other. Knowing how these variables interact is important in assessing their effect on the participant's evaluation of the relationship between the two binary variables.

Consider the nine matrices in Table 2. In the three matrices in the top row, $P(I)$ is constant at 0.5, ΔP is constant at 0.5, and $P(O)$ varies [$P(O) = 0.3$ in the left matrix, $P(O) = 0.5$ in the middle matrix, and $P(O) = 0.7$ in the right matrix]. In the middle row, $P(I)$ is increased to 0.8, and ΔP is maintained at 0.5. The constraint of increasing $P(I)$ and maintaining ΔP forces $P(O)$ to increase relative to the top row. In the bottom row, $P(I)$ is again increased to 0.8, and now $P(O)$ is maintained (0.3, 0.5, and 0.7). The constraint of increasing $P(I)$ and maintaining $P(O)$ forces ΔP to decrease relative to the top row. We prove in Appendix A that $P(O)$ is a linear function of $P(I)$ with slope ΔP and intercept $P(O|{\sim}I)$:

$$P(O) = \Delta P \times P(I) + P(O|\sim I) \tag{4}$$

In the passive task, the experimenter is in *control* of $P(I)$ and must decide whether ΔP or $P(O)$ should be maintained as $P(I)$ is varied. Consider experiment 3 in Allan and Jenkins (1983) where the passive task was used and there were two values of $P(I)$, 0.5

Table 2
Matrices illustrating the relations among $P(I)$, ΔP, and $P(O)$

P(I) = .5, ΔP = .5

	O	~O			O	~O			O	~O	
I	55	45	100	I	75	25	100	I	95	5	100
~I	5	95	100	~I	25	75	100	~I	45	55	100
	60	140	200		100	100	200		140	60	200
	P(O|I)	P(O|~I)	P(O)		P(O|I)	P(O|~I)	P(O)		P(O|I)	P(O|~I)	P(O)
	0.55	0.05	0.30		0.75	0.25	0.50		0.95	0.45	0.70

P(I) = .8, ΔP = .5

	O	~O			O	~O			O	~O	
I	88	72	160	I	120	40	160	I	152	8	160
~I	2	38	40	~I	10	30	40	~I	18	22	40
	90	110	200		130	70	200		170	30	200
	P(O|I)	P(O|~I)	P(O)		P(O|I)	P(O|~I)	P(O)		P(O|I)	P(O|~I)	P(O)
	0.55	0.05	0.45		0.75	0.25	0.65		0.95	0.45	0.85

P(I) = .8, ΔP = .31

	O	~O			O	~O			O	~O	
I	58	102	160	I	90	70	160	I	122	38	160
~I	2	38	40	~I	10	30	40	~I	18	22	40
	60	140	200		100	100	200		140	60	200
	P(O|I)	P(O|~I)	P(O)		P(O|I)	P(O|~I)	P(O)		P(O|I)	P(O|~I)	P(O)
	0.36	0.05	0.30		0.56	0.25	0.50		0.76	0.45	0.70

and 0.7. In that experiment, ΔP was maintained across $P(I)$, and therefore as $P(I)$ increased from 0.5 to 0.7, $P(O)$ also increased. However, the selected values of $P(I)$, $P(O)$, and ΔP were the same for each participant. The situation is more complicated in the active task. Because the participant, not the experimenter, is in control of $P(I)$, there likely will be between-participant variability in $P(I)$. As Eq. (4) makes clear, $P(O)$ and ΔP are affected by $P(I)$. That is, between-participant variability in $P(I)$ will result in between-participant variability in $P(O)$ and ΔP. Moreover, because the participant is in control of $P(I)$, the row frequencies of the 2×2 matrix cannot be preprogrammed, and the experimenter must adopt an algorithm to program the trial events that are dependent on the participant's trial responses. As we will show, the selection of a particular algorithm influences the variability in $P(O)$ and ΔP.

Many of the early studies concerned with contingency assessment used versions of the active task. In more recent research, the passive task has become more prevalent because of the control it provides the experimenter over the value of the input variable (see Dickinson, 2001). However, there are research questions that require the active task.

For example, an interest in "depressive realism" has resurfaced in the contingency literature (Allan et al., 2005, in press; Msetfi et al., in press, 2005). It has been known for some time (see Allan, 1993) that for a fixed ΔP, ratings of the relationship between two binary variables often are not constant but increase with $P(O)$ – the phenomenon is termed the "outcome density effect". Alloy and Abramson (1979) concluded that the outcome density effect is seen in nondepressed, but not depressed individuals. Such resistance by depressives to concluding that the contingency between a response and outcome is increased when just the outcome density is increased was termed "depressive realism". Alloy et al. (1985) concluded that this mood difference is found only in the active task where participants are required to assess how much control they had over the outcome. That is, in the active task, nondepressives display the outcome density effect and depressives do not, and in the passive task both mood groups show the outcome density effect. Therefore, when investigating depressive realism, it becomes especially important to understand the effect $P(I)$ has on the actual ΔP and $P(O)$ values, because in the active task the participant controls $P(I)$ and the values can vary among participants.

Examination of the literature indicates that few papers which used the active task report actual ΔP.[1] Rather, most only report the programmed values. Those that do provide information about actual ΔP report only mean ΔP values, and provide no information about the variability in ΔP. Information about actual $P(O)$ and $P(I)$ is even less frequently given, and again variability is not addressed. Also, the algorithm used to present the trial events is usually not explicitly described. Thus, in many studies using the active task, although we are told the nominal values of ΔP and $P(O)$, we know little about the actual values.

Table 3
Values of ΔP_N, $P(O|I)_N$, $P(O|{\sim}I)_N$, and $P(O)_N$ used in the experiment

ΔP_N	$P(O\|I)_N$	$P(O\|{\sim}I)_N$	$P(O)_N$
0	0.2	0.2	L
	0.5	0.5	M
	0.8	0.8	H
0.5	0.55	0.05	L
	0.75	0.25	M
	0.95	0.45	H

In the present paper we describe two algorithms and we present simulated data produced by these algorithms. We then present data from an active task which used the algorithm that produced the least variability in ΔP and $P(O)$. For both the simulated and the experimental data, we provide mean and variability information about actual ΔP and actual $P(O)$.

1. Simulations of the active task

We use the subscript "N" to denote a nominal value and the subscript "A" to denote an actual value. Our simulations are based on 40-trial sequences. $P(I)$ was varied from 0.1 to 0.9 in steps of 0.1. We used two values for ΔP_N (0 and 0.5). For each ΔP_N, there were two pairs of conditional probabilities, $P(O|I)_N$ and $P(O|{\sim}I)_N$. These values are shown in Table 3. As the pairs of conditional probabilities increase in value so does $P(O)_N$ (outcome density). Thus for each value of ΔP_N, there are three levels of outcome density, low (L), medium (M), and high (H).

Since the task requires the participant to judge the contingency between their input and the outcome, the algorithms focus on constraining ΔP_A with variations in $P(I)$ and $P(O)_N$.[2] For both algorithms, on I trials the selection of whether O or ${\sim}O$ occurs is determined by the value of $P(O|I)_N$, and on ${\sim}I$ trials the selection of whether O or ${\sim}O$ occurs is determined by $P(O|{\sim}I)_N$. The two algorithms differ in how this selection is made.

1.1. Constrained algorithm

This algorithm is based on the procedure described by Alloy and Abramson (1979). It was also used by Allan and Jenkins (1980, 1983) and Msetfi et al. (2005, accepted).[3] In these experiments, two outcome arrays, consisting of randomly ordered O and ${\sim}O$ values, were generated at the beginning of a block of trials *prior* to the start of data collection. The $(O|I)$ outcome array was determined by the value of $P(O|I)_N$, and the $(O|{\sim}I)$ outcome array was determined by the value of $P(O|{\sim}I)_N$. The number of elements in each array is equal to the total number of

[1] A notable exception is the research reported by Shanks and Dickinson (e.g., Dickinson et al., 1984; Shanks, 1985, 1987, 1989; Shanks and Dickinson, 1987; Shanks et al., 1989). These authors present mean actual ΔP values which differ little from the nominal values. However, variability in the actual values of ΔP was not reported in these studies.

[2] We have also explored a method that constrains $P(O)$ perfectly. However, it is of limited use because it applies only to $\Delta P = 0$.

[3] Personal communication. The algorithm is not described in the published papers.

trials. As an example, consider a condition with 40 trials where $P(O|I)_N = 0.55$, and $P(O|\sim I)_N = 0.05$ (and therefore $\Delta P_N = 0.5$). There would be 22 O values and 18 $\sim O$ values in the $(O|I)$ outcome array, and 2 O values and 38 $\sim O$ values in the $(O|\sim I)$ outcome array. Whether an outcome did, or did not, occur on a particular trial was determined by a rule that depended on input eventuality (i.e., the participant chose to respond or to not respond). The rule specifies that on I trials the $(O|I)$ array is consulted and on $\sim I$ trials the $(O|\sim I)$ array is consulted. Specifically, if the participant responded on trial y then the value at index y of the $(O|I)$ array was selected, and if the participant did not respond on trial y then the value at index y of the $(O|\sim I)$ array was selected. This can be also described as a *sampling without replacement* algorithm.

In our simulation of the constrained algorithm, the computer, rather than the participant, generated the input. That is, at the beginning of each simulation-run, an input array, determined by the value of $P(I)$, was generated. For example, if $P(I)$ were 0.6, there would be 24 I events and 16 $\sim I$ events, randomly ordered, in the input array. The simulation was run 1000 times for each of the 54 combinations of $P(I)$, ΔP_N, and $P(O)_N$: nine values of $P(I) \times 2$ values of ΔP_N, $\times 3$ values of $P(O)_N$.

Fig. 1 presents mean ΔP_A as a function of $P(I)$ for each level of $P(O)_N$. The bars indicate ± 1 standard deviation. Fig. 1A shows the data for $\Delta P_N = 0$ and Fig. 1B shows the data for $\Delta P_N = 0.5$. It is clear that while mean $\Delta P_A = \Delta P_N$, there is considerable variability. Fig. 2 illustrates the nature of this variability by plotting the standard deviation ($SD_{\Delta P}$) as a function of $P(I)$. When $\Delta P_N = 0$ (Fig. 2A), $SD_{\Delta P}$ is a U-shaped function of $P(I)$. The minimal variability occurs when there is no response bias $[P(I) = P(\sim I)]$. Interestingly $SD_{\Delta P}$ is highest for $P(O)_N = 0.5$. When $\Delta P_N = 0.5$ (Fig. 2B), $SD_{\Delta P}$ is a U-shaped function of $P(I)$ for $P(O)_N = 0.5$. For $P(O)_N = 0.3$, $SD_{\Delta P}$ tends to decrease with increases in $P(I)$, and for $P(O)_N = 0.7$, $SD_{\Delta P}$ tends to increase with increases in $P(I)$.

Fig. 3 presents mean $P(O)_A$ as a function of $P(I)$ for each level of $P(O)_N$. The bars indicate ± 1 standard deviation. Fig. 3A

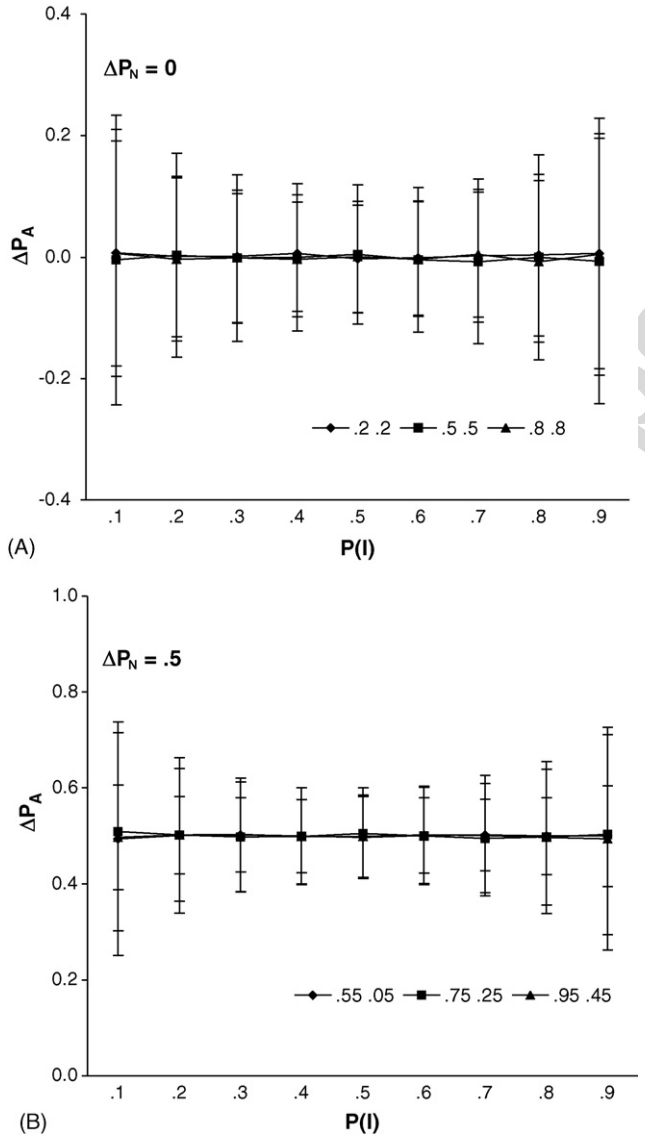

Fig. 1. Mean ΔP_A as a function of $P(I)$ for the three levels of $P(O)_N$, for $\Delta P_N = 0$ (A), and $\Delta P_N = 0.5$ (B). The bars indicate $\pm$standard deviation.

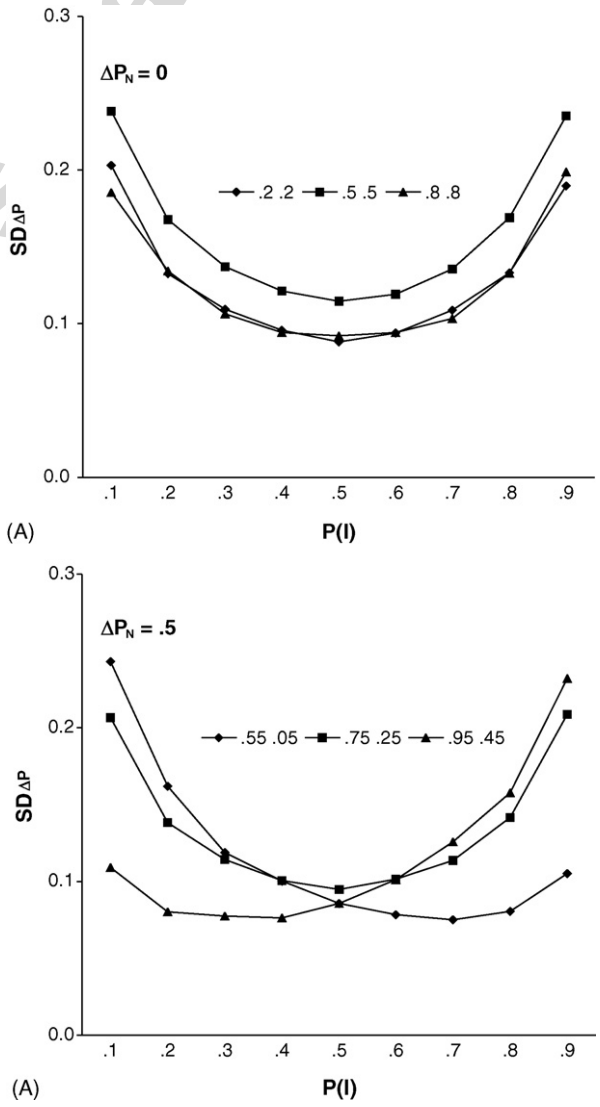

Fig. 2. Standard deviations of mean ΔP_A ($SD_{\Delta P}$) as a function of $P(I)$ for the three levels of $P(O)_N$, for $\Delta P_N = 0$ (A), and $\Delta P_N = 0.5$ (B).

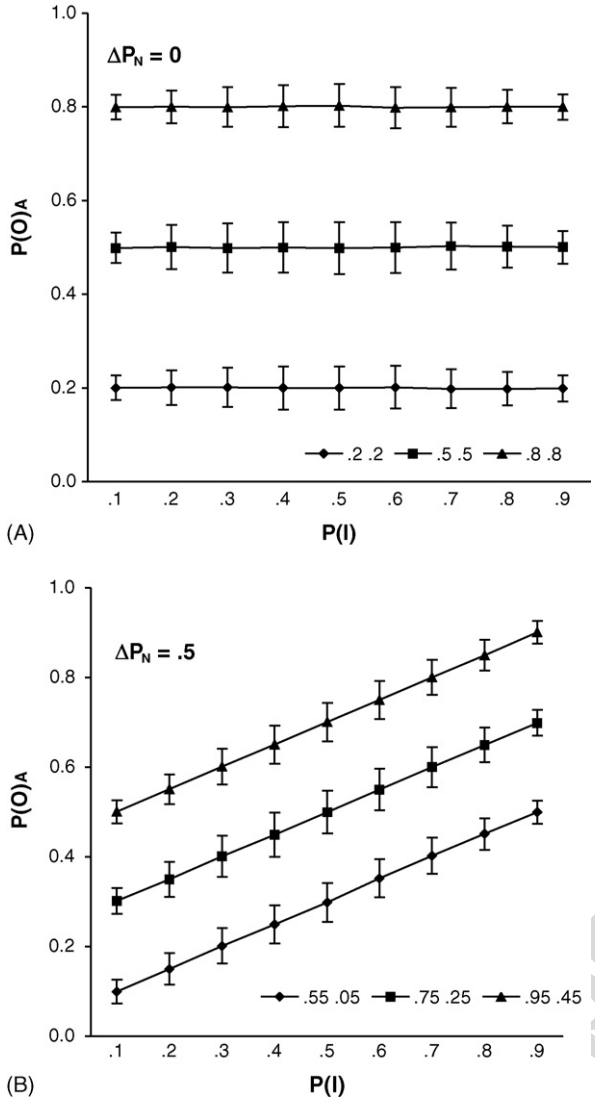

Fig. 3. Mean $P(O)_A$ as a function of $P(I)$ for the three levels of $P(O)_N$, for $\Delta P_N = 0$ (A), and $\Delta P_N = 0.5$ (B). The bars indicate ±standard deviation.

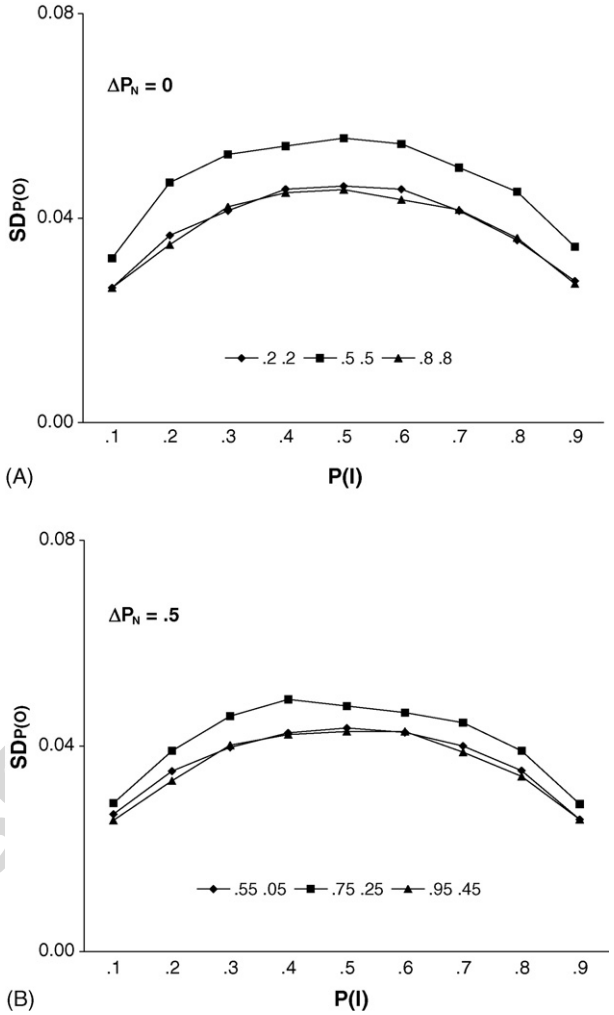

Fig. 4. Standard deviations of mean $P(O)_A$ ($SD_{P(O)}$) as a function of $P(I)$ for the three levels of $P(O)_N$, for $\Delta P_N = 0$ (A), and $\Delta P_N = 0.5$ (B).

shows the data for $\Delta P_N = 0$ and Fig. 3B shows the data for $\Delta P_N = 0.5$. As specified in Eq. (4), mean $P(O)_A$ is a linear function of $P(I)$ with slope ΔP_N and intercept $P(O|{\sim}I)_N$. The variability is less than that observed in ΔP_A. Fig. 4 illustrates the nature of the variability in $P(O)_A$ by plotting the standard deviation ($SD_{P(O)}$) as a function of $P(I)$. For both values of ΔP_N, $SD_{P(O)}$ is a inverted U-shaped function of $P(I)$. $SD_{\Delta P}$ is highest when $P(O)_N = 0.5$.

1.2. Unconstrained algorithm

This algorithm is similar to that used by Shanks and Dickinson (e.g., Shanks, 1985, 1987, 1989; Shanks and Dickinson, 1991; Shanks et al., 1989).[4] In contrast with the constrained algorithm, the unconstrained algorithm determines for each trial, *de noveau,* whether or not an outcome will be presented. For each trial a random number is generated, and (depending on the value of the number), an outcome is, or is not, presented. Whether or not a particular randomly selected number is an instruction to present an outcome is determined by the values of $P(O|I)_N$ and $P(O|{\sim}I)_N$. Consider again the $\Delta P_N = 0.5$ example in which $P(O|I)_N = 0.55$, and $P(O|{\sim}I)_N = 0.05$. On each trial, a random number between 0.01 and 1.00 is generated. If the participant has responded, an outcome will be presented only if that random number is $\leq$0.55. If the participant has not responded, an outcome will be presented only if that random number is $\leq$0.05. This can be thought of as a *sampling with replacement* algorithm.

As with the constrained algorithm, for each simulation-run a 40-trial input array, determined by the value of $P(I)$, was constructed. On each trial, a random number S, between 0.01 and 1.00 was generated. On an I trial, O was selected if $S < P(O|I)_N$; otherwise $\sim O$ was selected. On an $\sim I$ trial, O was selected if $S \leq P(O|{\sim}I)_N$; otherwise $\sim O$ was selected. The simulation was run 1000 times for each of the 54 combinations of $P(I)$, ΔP_N, and $P(O_N)$.

[4] Personal communication. The algorithm is not described in detail in the published papers. For example, Shanks and Dickinson (1991) state "These probabilities acted as parameters for a software probability generator. . ." p. 355.

The mean values of ΔP_A and $P(O)_A$ were very similar to those generated by the constrained algorithm. However, the variability was always greater.[5]

2. Experiment

We evaluated data collected from participants in an active form of the contingency task to verify the variability trends apparent in the simulations. Since the constrained algorithm resulted in less variability in ΔP_A and $P(O)_A$, it was used in the experiment.

This experiment is part of a larger project concerned with our theoretical formulation which views the detection of a contingency as similar to the detection of a signal (Allan et al., 2005). This is not the focus of the present article, and only data that relate to the simulations are reported here.

3. Method

3.1. Participants

The participants were 33 undergraduate students enrolled in psychology courses at McMaster University who received course credit. They had not been in other experiments concerned with contingency judgments.

3.2. Apparatus

Macintosh computers were used to present stimuli and record responses. The outcome variable was represented by a picture of a light bulb centered on the top portion of the computer monitor. The input variable was represented as a button located below the light bulb.

3.3. Procedure

The instructions for the experiment were presented to the participant on the computer monitor. In brief, participants were told that the task was to determine how much control they had over the illumination of a light bulb. On each trial they were to decide whether or not to click the button. They would then see whether or not the light bulb was illuminated. They were urged to sample both response alternatives (click and no click). They were told that after a series of 40 trials, they would be asked to rate how much control they thought they had over the bulb lighting up.

At the beginning of a block of 40 trials, an unilluminated light bulb appeared on the computer monitor. A trial began with the appearance of a button below the bulb for a maximum duration of 3 s. The participant had the option of pressing the button by clicking on it with the mouse or not pressing the button. A button click was scored as an I event and no click within the 3-s input period was scored as an $\sim I$ event. On I trials, the button disappeared as soon as it was clicked, and the 2-s outcome interval began. The light bulb either turned on for 2 s (O trials) or remained dark ($\sim O$). On $\sim I$ trials, the button disappeared at the end of the 3-s input interval, and the 2-s outcome interval began. Again, the light bulb either turned on for 2 s (O trials) or remained dark ($\sim O$). The dark light bulb remained on the monitor during the intertrial interval. The beginning of the next trial was signaled by the appearance of the button. At the end of a block, participants rated how much control they had over the illumination of the light bulb. The ratings were made on a horizontal scrollbar that ranged from 0 (no control) on the left to +100 (complete control) on the right. Participants made their ratings by moving the scrollbar left and right with the mouse.

There were two values of ΔP_N (0 and 0.5), and three values of $P(O)_N$ for each ΔP_N value (see Table 3). ΔP_N was varied between participants. There were 16 participants in the $\Delta P_N = 0$ group and 17 participants in the $\Delta P_N = 0.5$ group. Within each group, participants experienced the three values of $P(O)_N$ in a random order. A different light bulb was used for each level of $P(O)_N$.

4. Results and discussion

Table 4 shows the mean values of ΔP_A, $P(O|I)_A$, $P(O|\sim I)_A$, $P(I)_A$, and $P(O)_A$ for each of the six conditions in the experiment. As others have reported, the mean values of ΔP_A, $P(O|I)_A$, and $P(O|\sim I)_A$ are similar to the nominal values (ΔP_N, $P(O|I)_N$, and $P(O|\sim I)_N$, respectively). Overall there is a bias for choosing to press the button; i.e., $P(I)$ tends to be larger than $P(\sim I)$. As Eq. (4) specifies for $\Delta P_N > 0$, a bias for $P(I)$ results in increased values of $P(O)_A$ relative to unbiased responding (i.e., $P(I) = P(\sim I) = 0.5$).

Fig. 5 plots $P(I)$ for each participant for the three levels of $P(O)_N$. The data for $\Delta P_N = 0$ are in Fig. 5A and the data for $\Delta P_N = 0.5$ are in Fig. 5B. Each symbol represents the value for an individual participant. What is striking about Fig. 5 is the considerable variability in $P(I)$ both among participants and within a participant for the three levels of $P(O)_N$. Fig. 6 shows similar plots for ΔP_A. Note that the range of the y-axis is different in the two panels. Although mean ΔP_A deviated little from ΔP_N (Table 4), there is considerable variability among the participants. For example, in each outcome density condition some participants in the $\Delta P_N = 0$ group experienced a moderate positive contingency and others experienced a moderate negative contingency. Similarly, in the $\Delta P_N = 0.5$ group, some participants experienced a very high contingency (+0.95), and others a considerably lower-than-programmed contingency. This between-subject variability is also seen in the $P(O)_A$ data which are presented in Fig. 7.

The problems posed by the dependency of $P(O)$ and ΔP on $P(I)$ can be critical in the interpretation of the data. Differences in contingency ratings between conditions that have nominally identical levels of ΔP or $P(O)$, for example, may be artifacts due to differing levels of $P(I)$ leading to actual differences in ΔP or $P(O)$ or both. For example, Alloy and Abramson's (1979) seminal work on depressive realism used the constrained method for controlling trial events. However, their findings have had a mixed history of replication (see Allan et al., in press for a review). It is possible that this mixed history reflects some experiments in

[5] The outputs of both algorithms is available from the authors.

Table 4
Mean values of $P(I)_A$, ΔP_A, and $P(O)_A$ for each of the six conditions in the experiment

Nominal			Actual				
ΔP_N	$P(O\|I)_N$	$P(O\|\sim I)_N$	$P(I)_A$	ΔP_A	$P(O\|I)_A$	$P(O\|\sim I)_A$	$P(O)_A$
0	0.2	0.2	0.59	−0.01	0.17	0.18	0.18
	0.5	0.5	0.64	0.02	0.53	0.51	0.52
	0.8	0.8	0.64	0.03	0.80	0.77	0.79
0.5	0.55	0.05	0.67	0.51	0.55	0.04	0.39
	0.75	0.25	0.62	0.48	0.77	0.29	0.58
	0.95	0.45	0.68	0.57	0.95	0.38	0.77

which ΔP and $P(O)$ are equivalent across comparison groups, and some in which they are not.

Such concerns can be partially surmounted by measuring and reporting actual mean ΔP and $P(O)$ values, along with some measure of variability. If actual levels of ΔP and $P(O)$ do not vary significantly across conditions where they are supposed to be constant, then concerns about the artifice of results is minimized. The constrained method for controlling trials should

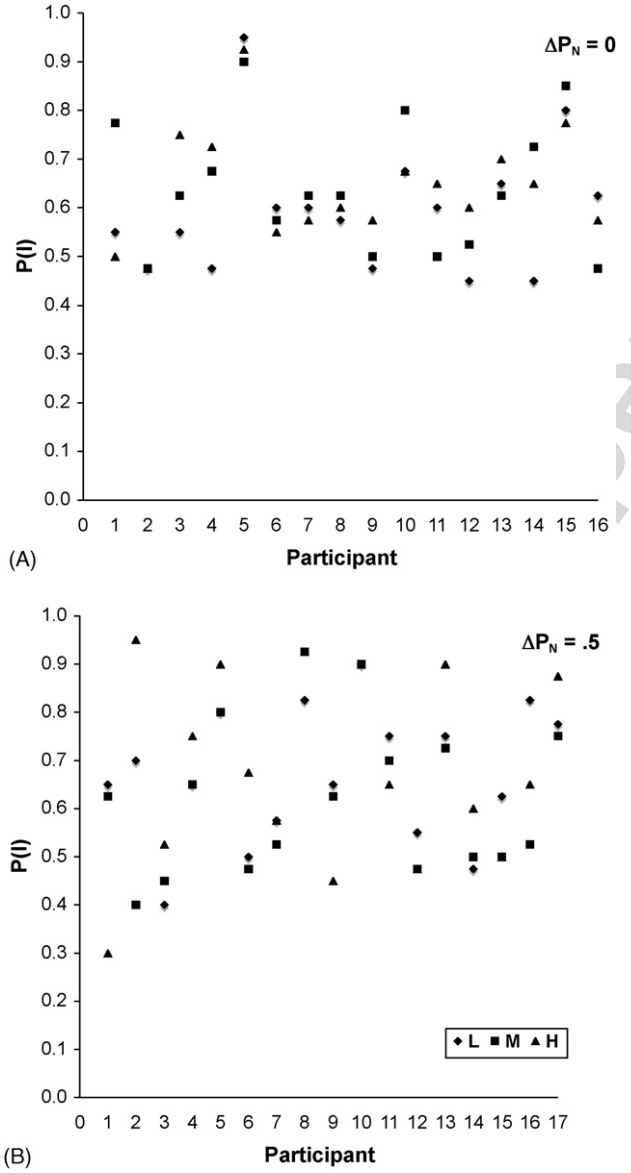

Fig. 5. $P(I)$ for each participant for the three levels of $P(O)_N$. Each symbol represents the value for an individual participant. The data for $\Delta P_N = 0$ are in (A), and the data for $\Delta P_N = 0.5$ are in (B).

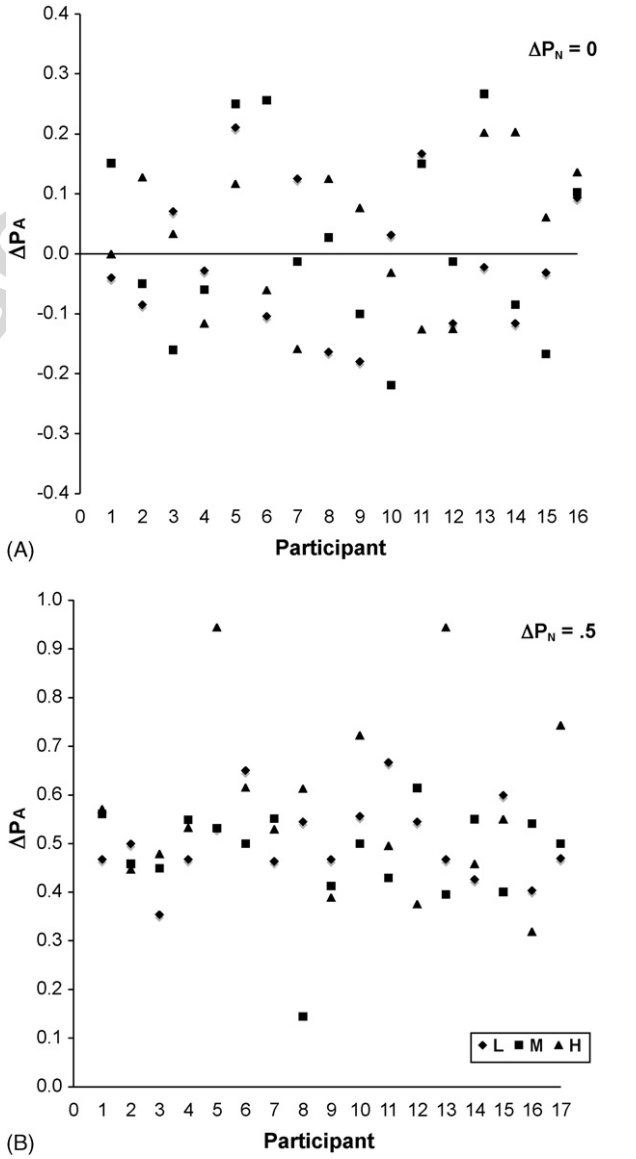

Fig. 6. ΔP_A for each participant for the three levels of $P(O)_N$. Each symbol represents the value for an individual participant. The data for $\Delta P_N = 0$ are in (A), and the data for $\Delta P_N = 0.5$ are in (B). The range on the y-axis is different in the two panels.

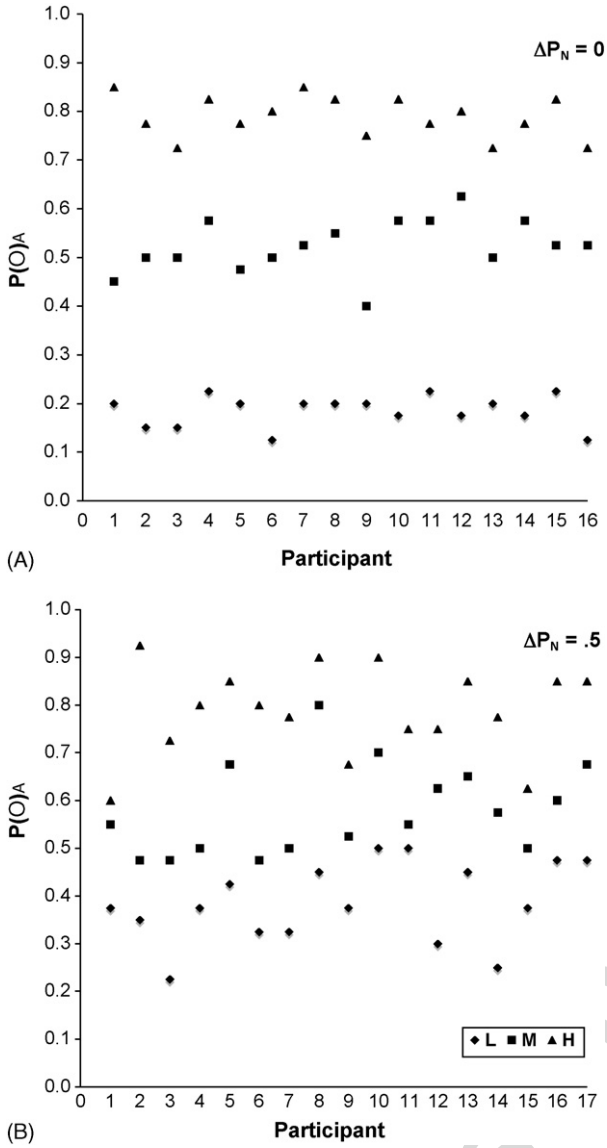

Fig. 7. $P(O)_A$ for each participant for the three levels of $P(O)_N$. Each symbol represents the value for an individual participant. The data for $\Delta P_N = 0$ are in (A), and the data for $\Delta P_N = 0.5$ are in (B).

be used rather than the unconstrained, as the former minimizes variability in ΔP and $P(O)$. Researchers should also consider determining whether the removal of deviant values of ΔP_A, $P(O)_A$, and/or $P(I)$ has a noticeable effect on the mean values. Another possibility is to instruct the participant, at the beginning of each trial, whether or not to respond. While such "command performance" would allow the experimenter to control $P(I)$, it remains an empirical question as to whether the participant would interpret the task as one of control rather than prediction.

4.1. Reclaiming the degree of freedom from the participant

The issues discussed in this paper are similar to those raised in the nonhuman animal conditioning literature many years ago. The operant conditioning situation is similar to the active contingency task. The organism either responds (for example, pecks a key) or does not respond, and then the outcome (for example, food reinforcement) occurs or does not occur. The Pavlovian conditioning situation is similar to the passive contingency task. A cue (for example, a tone) is presented or not, and then the outcome (for example, food reinforcement) occurs or does not occur. Gibbon et al. (1974) noted that operant conditioning departs from Pavlovian conditioning in that "one degree of freedom – response probability – is yielded to the subject" (p. 585). Platt and his colleagues (e.g., Platt, 1973; Galbicka and Platt, 1986; Scott and Platt, 1985) developed and evaluated a procedure that provides experimental control over the probability of a criterion operant response – percentile schedules. Part of the approach to percentile scheduling is the online updating of outcome schedules to maintain a relatively constant level of $P(O)$ and ΔP as response levels fluctuate. This aspect of percentile schedules suggests an intriguing solution to the problems outlined in this article. An online updating algorithm for human contingency learning has yet to be developed, but we would like to sketch out the main lines of the idea in hopes that someone will take up the challenge and generate a concrete solution.

Implementing the online-updating method is only possible using a constrained algorithm as it requires a fixed schedule of events. As responding unfolds, the current values of ΔP_A and $P(O)_A$ are evaluated at some fixed schedule, every ten trials, for example, and the unused portion of the outcome schedules (i.e., $O|I$, $O|{\sim}I$) are adjusted if the current ΔP_A and $P(O)_A$ values deviate beyond some tolerance threshold, e.g., 0.05. However, while the concept is relatively simple, there are many practical questions regarding implementation, especially surrounding what is an optimum scheme for reconfiguring the remaining outcome schedules after target values deviate. Do we re-populate both $O|I$ and $O|{\sim}I$ arrays if they deviate from some threshold, or only the most deviant of the two? Do we reconfigure the array if either ΔP or $P(O)$ values deviate from target, or only when both do, or do we just track the most critical and assume the other will tend to respond favourably? Do we reconfigure the entire remaining schedule arrays, or only the next X events, where X is some number of trials equivalent to that of the test window?

Going even further, researchers could implement a full percentile scheduling approach. This requires that reinforcement delivery is contingent on responses meeting a criterion defined with respect to some response property that can be measured on at least an ordinal scale (e.g., a pigeon is required to peck a key some number of times, or for some length of time). A percentile schedule identifies a criterion response by reference to a distribution of the organism's recent response values. For example, a criterion lever press may be defined as one with a relatively long duration – for example, longer than 80% of the durations generated by the organism's recent lever presses. This percentile schedule would maintain a fixed probability of the criterion response (0.2 in this example).

The active contingency task usually requires a discrete response, and is not amenable to a percentile schedule of outcomes. Abramson et al. (1981) used a task that could be modified to be compatible with a percentile schedule. In their experiment, the response period was 5 s. A button press during the first 2 s was considered a response and a button press during the last

3 s was considered a non-response. The response requirement could be controlled by a percentile schedule targeted on timing of the button press during the response period. Thus, only the fastest 25% of responses emitted in the response period could produce an outcome [or whatever percentage corresponding to $P(O|I)$].

Percentile schedules gave the control over response probability to the experimenter in operant conditioning. The application of percentile schedules in the active contingency task would also give the experimenter, not the participant, control over $P(I)$, and would reclaim the degree of freedom the experimenter typically surrenders to the participant. Church (1993) made an eloquent plea on behalf of a greater symbiosis between human and nonhuman based research. Although percentile schedules may be more difficult to implement in the human contingency task than in animal learning situation, we think they offer another opportunity for human-centered researchers to benefit from their nonhuman-centered colleagues.

Acknowledgements

The preparation of this paper was supported by research grants from the Natural Sciences and Engineering Research Council to LGA and SS, and by a grant from the United States National Institute on Drug Abuse to SS.

Appendix A

$$
\begin{aligned}
P(O) &= \Delta P \times\ P(I) + P(O|\sim I) \\
&= \left(\frac{a}{a+b} - \frac{c}{c+d}\right)\left(\frac{a+b}{a+b+c+d}\right) + \frac{c}{c+d} \\
&= \left(\frac{a}{a+b}\right)\left(\frac{a+b}{a+b+c+d}\right) - \left(\frac{c}{c+d}\right) \\
&\quad \times \left(\frac{a+b}{a+b+c+d}\right) + \frac{c}{c+d} \\
&= \left(\frac{a}{a+b+c+d}\right) + \left(\frac{c}{c+d}\right)\left(1 - \frac{a+b}{a+b+c+d}\right) \\
&= \left(\frac{a}{a+b+c+d}\right) + \left(\frac{c}{c+d}\right)\left(\frac{c+d}{a+b+c+d}\right) \\
&= \left(\frac{a}{a+b+c+d}\right) + \left(\frac{c}{a+b+c+d}\right) = P(O)
\end{aligned}
$$

References

Abramson, L.Y., Alloy, L.B., Rosoff, R., 1981. Depression and the generation of complex hypotheses in the judgment of contingency. Behav. Res. Ther. 19, 35–45.

Allan, L.G., 1993. Human contingency judgments: rule-based or associative? Psychol. Bull. 114, 435–448.

Allan, L.G., Jenkins, H.M., 1980. The judgment of contingency and the nature of the response alternatives. Can. J. Psychol. 34, 1–11.

Allan, L.G., Jenkins, H.M., 1983. The effects of representations of binary variables on judgment of influence. Learn. Motiv. 14, 381–405.

Allan, L.G., Siegel, S., Hannah, S., in press. The sad truth about depressive realism. Quart. J. Exp. Psychol.

Allan, L.G., Siegel, S., Tangen, J.M., 2005. A signal detection analysis of contingency data. Learn. Behav. 33, 250–263.

Allan, L.G., Tangen, J.M., 2005. Judging relationships between events: how do we do it? Can. J. Exp. Psychol. 59, 22–27.

Alloy, L.B., Abramson, L.Y., 1979. Judgment of contingency in depressed and nondepressed students: sadder but wiser? J. Exp. Psychol.: Gen. 108, 41–485.

Alloy, L.B., Abramson, L.Y., Kossman, D., 1985. The judgment of predictability in depressed and nondepressed college students. In: Overmier, J.B., Brush, F.R. (Eds.), Affect, Conditioning, Cognition: Essays on the Determinants of Behavior. Erlbaum, Hillsdale, NJ, pp. 230–246.

Buehner, M.J., May, J., 2003. Rethinking temporal contiguity and the judgment of causality: effects of prior knowledge, experience, and reinforcement procedure. Quart. J. Exp. Psychol. 56A, 865–890.

Church, R.M., 1993. Human models of animal hehavior. Psychol. Sci. 4, 170–173.

De Houwer, J., Beckers, T., 2002. A review of recent developments in research and theory on human contingency learning. Quart. J. Exp. Psychol. 55B, 289–310.

Dickinson, A., Shanks, D.R., Evenden, J.L., 1984. Judgement of act-outcome contingency: The role of selective attribution. Quart. J. Exp. Psychol. 36A, 29–50.

Dickinson, A., 2001. Causal learning: an associative analysis. Quart. J. Exp. Psychol. 54B, 3–25.

Galbicka, G., Platt, J.R., 1986. Parametric manipulation of interresponse-time contingency independent of reinforcement rate. J. Exp. Psychol.: Anim. Behav. Processes 12, 371–380.

Gibbon, J., Berryman, R., Thompson, R.L., 1974. Contingency spaces and measures in classical and instrumental conditioning. J. Exp. Anal. Behav. 21, 585–605.

Msetfi, R.M., Murphy, R.A., Simpson, J., in press. Depressive realism and contingency judgement: further support for the contextual processing hypothesis. Quart. J. Exp. Psychol.

Msetfi, R.M., Murphy, R.A., Simpson, J., Kornbrot, D.E., 2005. Depressive realism and outcome density bias in contingency judgements: the effect of the context and inter-trial interval. J. Exp. Psychol.: Gen. 134, 10–22.

Platt, J.R., 1973. Percentile reinforcement: paradigms for experimental analysis of response shaping. In: Bower, G.H. (Ed.), The Psychology of Learning and Motivation: Advances in Theory and Research, vol. 7. Academic Press, New York, pp. 271–296.

Scott, G.K., Platt, J.R., 1985. Model of response-reinforcer contingency. J. Exp. Psychol.: Anim. Behav. Processes 11, 152–171.

Shanks, D.R., 1985. Continuous monitoring of human contingency judgment across trials. Memory Cogn. 13, 158–167.

Shanks, D.R., 1987. Acquisition functions in contingency judgment. Learn. Motiv. 18, 147–166.

Shanks, D.R., 1989. Selectional processes in causality judgments. Memory Cogn. 17, 27–34.

Shanks, D.R., Dickinson, A., 1987. Associative accounts of causality judgment. In: Bower, G.H. (Ed.), The Psychology of Learning and Motivation, 21. Academic Press, San Diego, CA, pp. 229–261.

Shanks, D.R., Dickinson, A., 1991. Instrumental judgment and performance under variations in action-outcome contingency and contiguity. Memory Cogn. 19, 353–360.

Shanks, D.R., Pearson, S.M., Dickinson, A., 1989. Temporal contiguity and the judgement of causality by human subjects. Quart. J. Exp. Psychol. 41B, 139–159.

Wasserman, E.A., Chatlosh, D.L., Neunaber, D.J., 1983. Perception of causal relations in humans: factors affecting judgments of response-outcome contingencies under free-operant procedures. Learn. Motiv. 14, 406–432.