Revista Mexicana de Análisis de la
Conducta
Sociedad Mexicana de Análisis de la Conducta
jburgos@ cucba.udg.mx
ISSN (Versión impresa): 0185-4534
MÉXICO

2007
Celia Lie / Brent Alsop
REINFORCEMENT AND PUNISHMENT IN BEHAVIORAL MODELS OF SIGNAL DETECTION
Revista Mexicana de Análisis de la Conducta, año/vol. 33, número Especial
Sociedad Mexicana de Análisis de la Conducta
Guadalajara, México
pp. 45-55

REVISTA MEXICANA DE ANÁLISIS DE LA CONDUCTA 2007 NÚMERO ESPECIAL
MEXICAN JOURNAL OF BEHAVIOR ANALYSIS **33**, 45-55 SPECIAL ISSUE

REINFORCEMENT AND PUNISHMENT IN BEHAVIORAL MODELS OF SIGNAL DETECTION

REFORZAMIENTO Y CASTIGO EN MODELOS CONDUCTUALES DE DETECCIÓN DE SEÑALES

CELIA LIE[1] AND BRENT ALSOP[2]
[1]VICTORIA UNIVERSITY OF WELLINGTON, NEW ZEALAND
[2]UNIVERSITY OF OTAGO, NEW ZEALAND

ABSTRACT

Behavioral models of signal detection have focused almost exclusively on the effects of reinforcement for correct choices. In contrast, the effects of punishment for errors have been largely ignored. Two competing models of punishment can be derived from research using simple concurrent-schedule procedures. Subtractive models predict that punishers directly subtract from the effects of reinforcers for the same response alternative, and additive models predict that the effects of punishers add onto the effects of reinforcers obtained for the other response alternative. These two models were incorporated into Davison and Tustin's (1978) model of signal-detection performance. Some preliminary research using human subjects in a signal-detection procedure provides support for an additive punishment version of the Davison and Tustin model.

Keywords: signal detection, punishment, response bias, human.

RESUMEN

Los modelos conductuales de detección de señales se han focalizado casi exclusivamente sobre los efectos del reforzamiento para elecciones correc-

1. Celia Lie, School of Psychology, Victoria University of Wellington, P.O. Box 600, Wellington, New Zealand. celia.lie@vuw.ac.nz

tas. En contraste, los efectos del castigo por errores han sido largamente ignorados. Se pueden derivar dos modelos competitivos del castigo a partir de la investigación con programas concurrentes simples. Modelos sustractivos predicen que los castigadores sustraen directamente de los efectos de los reforzadores para la misma alternativa de respuesta, y modelos aditivos predicen que los efectos de los castigadores se sumen a los de los reforzadores obtenidos por la otra alternativa de respuesta. Estos dos modelos fueron incorporados en el modelo de Davison y Tustin (1978) de la ejecución de detección de señales. Investigación preliminar con sujetos humanos en un procedimiento de detección de señales provee apoyo a una versión de castigo aditivo del modelo de Davison y Tustin.

Palabras clave: detección de señales, castigo, sesgo de respuesta, humano.

Many situations require organisms to make choices involving the detection or identification of stimuli. For example, a bird must decide whether a butterfly is toxic or safe to eat, a motorist must decide whether it is safe to pass the car ahead, or a pathologist must decide whether a sample of cells are cancerous or not. From a behavioral analysis perspective, detection is not only an issue of stimulus discriminability; the consequences of the choices are also important. The bird gets sick from eating the toxic butterfly, the motorist crashes, or the pathologist is dismissed or reprimanded. In these examples, both the negative consequences arising from errors and the positive consequences arising from correct choices should influence the choices that are made.

Research in behavioral detection has focused almost exclusively on the effects of positive outcomes for correct responses. Studies have investigated the effects of varying reinforcer rate, reinforcer duration, and reinforcer delay on signal-detection performance (see Davison & McCarthy, 1988; Davison & Nevin, 1999). The research has, however, ignored the effects of negative outcomes for errors. Davison and McCarthy (1980) even studied the effects of reinforcement for errors, but not the effects of punishment.

The matrix in Figure 1 shows the response outcomes in a behavioral detection procedure. In a two-alternative signal-detection task, there are two different stimuli (S_1 and S_2) and two corresponding response options (B_1 and B_2). Thus, there are four possible response outcomes. Correct responses are B_1 following the presentation of S_1 (B_{11}), and B_2 following S_2 presentations (B_{22}). Often, the subject receives reinforcement (either intermittent or continuous) for these responses (R_{11} and R_{22}, Figure 1). The responses, B_{12} and B_{21} are the incorrect responses following S_1 and S_2, respectively. Behavioral detection experiments typically arrange no consequences for errors, but they could be punished. In Figure 1, P_{12} and P_{21} denote the number of punishers obtained for B_{12} and B_{21} responses, respectively.

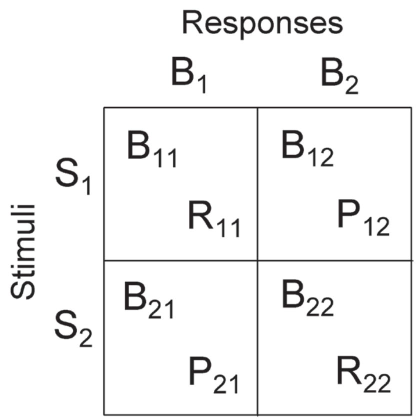

Figure 1. The four possible response outcomes and their corresponding reinforcement and punishment terms in a two-alternative signal-detection task. See text for details.

There are a number of reasons why little research has considered the effects of punishment for errors. A recent article by Critchfield, Paletz, MacAleese, and Newland (2003) provided a good summary of these issues, such as difficulties with equating reinforcer and punisher values, and analytical problems associated with extending current models of choice to incorporate punishment. There is a further practical consideration that complicates modelling punishment effects in choice procedures; punishment is typically arranged against a backdrop of reinforcement so that responding occurs. Any punishment model, therefore, must incorporate the combined effects of both reinforcement and punishment on choice behavior.

There are two main competing models of choice that incorporate punishment and reinforcement effects: an additive (also termed competitive-suppression, or indirect-suppression) model (e.g., Deluty, 1976), and a subtractive (or direct-suppression) model (e.g., de Villiers, 1980; Farley, 1980; Farley & Fantino, 1978). Additive models of punishment predict that the effects of punishers on one response alternative add to the effects of the reinforcers for the other alternative. Subtractive models predict that the effects of punishers for a response directly subtract from the effects of reinforcement for the same response.

Critchfield et al. (2003) reviewed the limited research on these two models. Research with non-human subjects has provided some support for both models. However, some studies made no direct comparisons between the differing predictions of the two models (e.g., Deluty, 1976; Deluty & Church, 1978; Farley & Fantino, 1978), or arranged conditions that did not test the two models effectively (e.g., Deluty, 1976). There are only two published studies that have made direct comparisons between additive and subtractive model predictions (de Villiers, 1980; Farley, 1980). In both these studies, rates of reinforcement were unequal while rates of punishment were equal across two response alternatives in concurrent schedules. With this arrangement, an additive model of punishment predicts reduced preference for the richer reinforcement schedule, while a subtractive model predicts an increased preference towards the richer reinforcement schedule. Both studies reported an increased preference for the alternative associated with the richer reinforcement schedule, that is, support for a subtractive model. Overall, research from non-humans provides better evidence for a subtractive instead of an additive model of punishment.

There is even less research on the effects of punishment on human choice behavior (e.g., Bradshaw, Szabadi, & Bevan, 1979; Gray, Stafford, & Tallman, 1991). Critchfield et al. (2003) published the only human study that directly compared the predictions of additive and subtractive models. They used a two-alternative concurrent schedule where computer-mouse clicking by undergraduate students was reinforced by monetary gain and punished by monetary loss. Critchfield et al. also found that a subtractive model of punishment described their individual subject data better than an additive model.

To date, there is no published research on signal-detection models of punishment. In order to derive signal-detection models of punishment, we must first look at behavioral models of detection performance. Davison and Tustin (1978) developed the most widely-used behavioral detection model. Their model was based on the generalised matching law (GML; Baum, 1974). The GML can be written

$$\log\left(\frac{B_1}{B_2}\right) = a\log\left(\frac{R_1}{R_2}\right) + \log c\,, \tag{1}$$

where B_1 and B_2 are as defined above, and R_1 and R_2 denote the number of obtained reinforcers for B_1 and B_2 responses respectively. The parameter *a* measures the sensitivity of behavior to the ratio of reinforcers; that is, the extent to which changes in reinforcer ratios (R_1/R_2) change the distribution of responses (B_1/B_2). The parameter *c* (or log *c*) measures inherent bias; that is, a constant preference the subject might display for one alternative, independent of changes to the reinforcer ratio. Davison and Tustin (1978) proposed that when two stimuli in a detection task are indiscriminable, then the distribution of responses across the two alternatives should depend only on the relative reinforcer ratio for the two alternatives (i.e., Equation 1). However, as S_1 and S_2 become more distinguishable, then the subject becomes more biased towards making correct responses (B_{11} and B_{22}, Figure 1). Following S_1 trials, this can be written as

$$\log\left(\frac{B_{11}}{B_{12}}\right) = a\log\left(\frac{R_{11}}{R_{22}}\right) + \log c + \log d\,, \tag{2}$$

and following S_2 trials,

$$\log\left(\frac{B_{21}}{B_{22}}\right) = a\log\left(\frac{R_{11}}{R_{22}}\right) + \log c - \log d\,, \tag{3}$$

where all notation is as above, and log *d* measures the discriminability between S_1 and S_2.

Measures of discriminability and bias can be derived algebraically from Equations 2 and 3. Algebraic subtraction of the two equations yields a formula for discriminability (log *d*), which is,

$$\log d = 0.5\log\left(\frac{B_{11}B_{22}}{B_{12}B_{21}}\right), \tag{4}$$

where all notation is as above. Algebraic addition of Equations 2 and 3 gives a formula for response bias (log *b*),

$$\log b = 0.5\log\left(\frac{B_{11}B_{12}}{B_{21}B_{22}}\right) = a\log\left(\frac{R_{11}}{R_{22}}\right) + \log c\,, \tag{5}$$

where all notation is as above. Equation 5 predicts that response bias, log *b*, should follow the generalised matching law (Equation 1).

From Equations 2 and 3, it is straightforward to incorporate the effects of punishment into Davison and Tustin’s (1978) GML-based model of signal-detection performance. An additive punishment model of signal-detection performance following S_1 trials can be written as

$$\log\left(\frac{B_{11}}{B_{12}}\right) = a\log\left(\frac{R_{11} + qP_{12}}{R_{22} + qP_{21}}\right) + \log c + \log d\,, \tag{6}$$

and following S_2 trials as

$$\log\left(\frac{B_{21}}{B_{22}}\right) = a\log\left(\frac{R_{11} + qP_{21}}{R_{22} + qP_{12}}\right) + \log c - \log d \,, \tag{7}$$

with response bias calculated as

$$\log b = 0.5\log\left(\frac{B_{11}B_{12}}{B_{21}B_{22}}\right) = a\log\left(\frac{R_{11} + qP_{12}}{R_{22} + qP_{21}}\right) + \log c \,, \tag{8}$$

where all notation is as above, and q is a scaling factor measuring the relative potency of the punisher compared to a reinforcer. For example, if q = .5, then a punisher would be half as effective as a reinforcer.

Applying a subtractive model of punishment to the Davison and Tustin (1978) model yields an equation following S_1 trials,

$$\log\left(\frac{B_{11}}{B_{12}}\right) = a\log\left(\frac{R_{11} - qP_{21}}{R_{22} - qP_{12}}\right) + \log c + \log d \,, \tag{9}$$

and following S_2 trials,

$$\log\left(\frac{B_{21}}{B_{22}}\right) = a\log\left(\frac{R_{11} - qP_{12}}{R_{22} - qP_{21}}\right) + \log c - \log d \,, \tag{10}$$

with response bias calculated as

$$\log b = 0.5\log\left(\frac{B_{11}B_{12}}{B_{21}B_{22}}\right) = a\log\left(\frac{R_{11} - qP_{21}}{R_{22} - qP_{12}}\right) + \log c \,, \tag{11}$$

where all notation is as above. Equations 9, 10, and 11 are undefined if $qP_{21} \geq R_{11}$ or $qP_{12} \geq R_{22}$.

Like the additive and subtractive free-operant models of punishment (de Villiers, 1980; Farley, 1980), the additive and subtractive signal-detection models also predict different effects of punishment. Figure 2 illustrates this difference. The solid straight line shows the predicted changes in response bias when no punishment is arranged, and the reinforcer ratios (B_{11}/B_{22}) were varied from 175 reinforcers/hour for B_{11} and 25 reinforcers/hour for B_{22} (7:1) to 25 reinforcers/hour and 175 reinforcers/hour (1:7). Sensitivity, *a,* was set at .9, and inherent bias, *c*, was set at 1. The dashed line and the dotted line show the predictions of the additive model (Equation 8) and subtractive model (Equation 11) when a constant low rate (20 punishers per hour) of punishment is superimposed on the reinforcer ratios. The subtractive model predicts a steeper function (i.e., response bias becomes more extreme), and the additive model predicts a shallower function (i.e., response bias becomes less extreme) than the no-punishment condition.

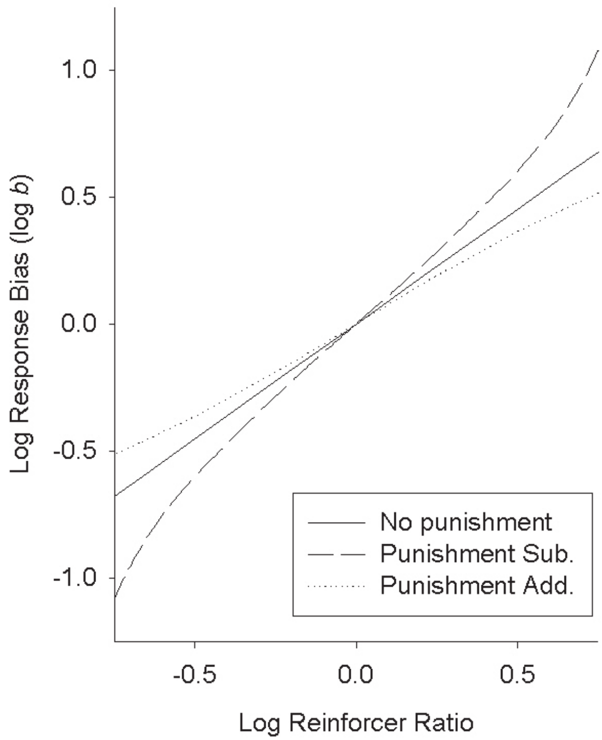

Figure 2. Predictions made by the Davison and Tustin (1978) model of signal detection behaviour. See text for details.

We have recently conducted some preliminary research using a discrete-trial signal-detection procedure with reinforcement for correct responses and punishment for errors. In this procedure, undergraduate students at the University of Otago judged whether there were more blue or yellow objects in various 10 x 10 arrays on a computer screen. We selected an unusual system of reinforcers and punishers. We had noticed that undergraduate students were eager to end our experimental sessions, therefore we made release from the session contingent upon performance. Participants were required to obtain a certain number of points (reinforcers) for correct responses before ending the experimental session. However, incorrect responses sometimes led to point deductions (punishers), thus increasing session time. A "thermometer bar" presented on the right-hand side of the computer screen displayed the net number of points obtained. When the thermometer bar reached the top, the experimental session ended.

The experiment used a between-groups design. For Group A, the reinforcer distribution for correct responses was varied in four different conditions across the two response alternatives (5:1, 2:1, 1:2, 5:1), and the arranged overall reinforcer rate was six reinforcers per minute. There was no punishment for errors. Group A determined whether the reinforcers affected response bias. For Group B, the distribution of the punishers was varied across the two alternatives (5:1, 2:1, 1:2, 5:1), and equal numbers of reinforcers were arranged on each alternative. Group B determined whether the punishers affected response bias. Group C received the same four reinforcer distributions used in Group A, but an equal number of punishers for errors was arranged across the response alternatives. Thus, the essential comparison was between Groups A and C, where the same reinforcer distributions were arranged, but Group C also included a low, equal rate of punishment.

Figure 3 plots the results from the three groups with response bias calculated using Equation 5. The positive regression slope for Group A (a = 0.36) shows that the participants were sensitive to the distribution of reinforcers across the two alternatives for the four conditions, and this change in bias was significant, $F(1,5) = 13.18$, $p < .05$. The negative regression slope for Group B (a = -0.20) shows that participants were sensitive to the distribution of punishers across the four conditions, and this effect was also significant, $F(1,5) = 4.387$, $p < .05$. Group C, like Group A, showed a positive regression slope, but it was shallower than that obtained in Group A (a = 0.15). However, a t-test on the slopes of the functions from the individual subject data across groups approached, but did not reach, significance, $t(10) = 2.115$, p =.061. The individual subject data showed that was largely attributable to one participant's performance in one condition.

These preliminary results provide support for an additive model rather than a subtractive model of punishment for signal-detection procedures (Figure 2). This finding was unexpected, because it differs from Critchfield et al.'s (2003) study which found support for a subtractive model using human participants, and differs from previous non-human research which also provides stronger evidence for a subtractive model than an additive model of punishment. There

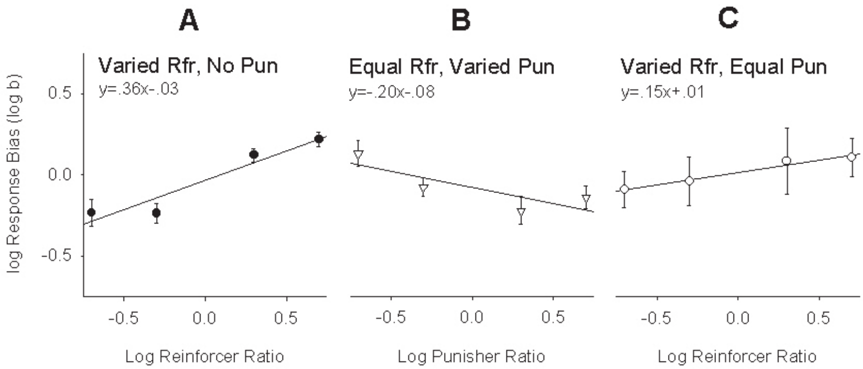

*Figure 3***.** *Fitted least-squares regression lines and error bars of subject data for response bias (log b) over changes in log Reinforcer Ratio for Groups A and C, and changes in log Punisher Ratio for Group B.*

are a number of reasons why the findings from this preliminary study might differ from those found with concurrent schedules. First, our preliminary study used a between-groups comparison with a limited number of participants. Obviously, this result needs replication, and it needs extension to within-subject comparisons. We are currently conducting such research.

Second, the type of reinforcers and punishers used in our study differ from those that might be used in non-human animal research (e.g., food and electric shock). Points obtained lead to escape from the experiment and perhaps can be viewed as conditioned negative reinforcers. Similarly, instances of point loss can be seen as conditioned negative punishers. It is possible that negative reinforcers and punishers differ in their effects when compared to positive reinforcers and punishers. We are currently conducting non-human animal research using more traditional reinforcers and punishers to address this issue.

Third, and more important, reinforcement (and, so too punishment) might not have identical properties on free-operant concurrent-schedule and discrete-trial detection procedures. For example, free-operant concurrent-schedule research (e.g, Davison & Baum, 2002; Davison & Baum, 2003; Landon, Davison, & Elliffe, 2003) has found that reinforcement creates “preference pulses” for the response alternative that was just reinforced. Preference pulses have also been found in signal-detection procedures (Alsop & Rowley, 1996), but they differed from those found in concurrent schedules; reinforcers were followed by a brief increase in response bias for the alternative with the higher rate of reinforcement, regardless of which alternative had just been reinforced. It may be the case that subtractive punishment models may better model performance in simple concurrent-schedule procedures, and additive punishment models better model performance in signal-detection procedures. Further research is required to establish whether there are underlying differences between these two procedure types.

Here we have examined only the predictions made by additive and subtractive punishment versions of the Davison and Tustin (1978) model. Other competing models of detection might accommodate the effects of punishment more readily. An obvious alternative is Alsop and Davison's (1991) behavioural model of signal detection. Interestingly, although simple extensions of Alsop and Davison's model accommodate additive punishment effects, the model does not extend to subtractive punishment effects so readily. In fact, the simplest subtractive version of the Alsop and Davison model becomes incalculable at quite modest rates of punishment. We are currently working on different mathematical versions of this model.

Regardless of whether or not there are differences in the effects of punishment between free-operant concurrent-schedule and discrete-trial signal-detection procedures, signal detection is a procedure well suited for studying punishment effects in non-human animals, and in particular, human participants. The arrangements of rewards for correct responses and punishers for errors parallel many everyday situations, so the procedures are easily explained to, and readily accepted by, human participants. As a method for studying general human choice behavior, signal-detection procedures may offer a worthwhile complement to free-operant tasks.

REFERENCES

Alsop, B., & Davison, M. (1991). Effects of varying stimulus disparity and the reinforcer ratio in concurrent-schedule and signal-detection procedures. *Journal of the Experimental Analysis of Behavior, 56*(1), 67-80.

Alsop, B., & Rowley, R. (1996). Types of responding in a signal-detection task. *Journal of the Experimental Analysis of Behavior, 65*(3), 561-574.

Baum, W.M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal fo the Experimental Analysis of Behavior*, 22(1), 231-242.

Bradshaw, C., Szabadi, E., & Bevan, P. (1979). The effect of punishment on free-operant choice behavior in humans. *Journal of the Experimental Analysis of Behavior, 31*(1), 71-81.

Critchfield, T. S., Paletz, E. M., MacAleese, K. R., & Newland, M. (2003). Punishment in human choice: Direct or competitive suppression? *Journal of the Experimental Analysis of Behavior, 80*(1), 1-27.

Davison, M., & Baum, W. M. (2002). Choice in a variable environment: Effects of blackout duration and extinction between components. *Journal of the Experimental Analysis of Behavior, 77*(1), 65-89.

Davison, M., & Baum, W. M. (2003). Every reinforcer counts: Reinforcer magnitude and local preference. *Journal of the Experimental Analysis of Behavior, 80*(1), 95-129.

Davison, M., & McCarthy, D. (1980). Reinforcement for errors in a signal-detection procedure. *Journal of the Experimental Analysis of Behavior, 34*(1), 35-47.

Davison, M., & McCarthy, D. (1988). *The matching law: A research review*: (1988).

Davison, M., & Nevin, J. A. (1999). Stimuli, reinforcers, and behavior: An integration. *Journal of the Experimental Analysis of Behavior, 71*(3), 439-482.

Davison, M., & Tustin, R. (1978). The relation between the generalized matching law and signal-detection theory. *Journal of the Experimental Analysis of Behavior, 29*(2), 331-336.

de Villiers, P. A. (1980). Toward a quantitative theory of punishment. *Journal of the Experimental Analysis of Behavior, 33*(1), 15-25.

Deluty, M. Z. (1976). Choice and the rate of punishment in concurrent schedules. *Journal of the Experimental Analysis of Behavior, 25*(1), 75-80.

Deluty, M. Z., & Church, R. M. (1978). Time-allocation matching between punishing situations. *Journal of the Experimental Analysis of Behavior, 29*(2), 191-198.

Farley, J. (1980). Reinforcement and punishment effects in concurrent schedules: A test of two models. *Journal of the Experimental Analysis of Behavior, 33*(3), 311-326.

Farley, J., & Fantino, E. (1978). The symmetrical law of effect and the matching relation in choice behavior. *Journal of the Experimental Analysis of Behavior, 29*(1), 37-60.

Gray, L. N., Stafford, M. C., & Tallman, I. (1991). Rewards and Punishments in Complex Human Choices. *Social Psychology Quarterly, 54*(4), 318-329.

Landon, J., Davison, M., & Elliffe, D. (2003). Concurrent schedules: Reinforcer magnitude effects. *Journal of the Experimental Analysis of Behavior, 79*(3), 351-365.