

## RESEARCH ARTICLE

# Is superstitious responding a matter of detectability? A replication of Killeen (1978)

Stephanie Gomes-Ng<sup>1,2</sup> | Sarah Cowie<sup>2</sup> | Douglas Elliffe<sup>2</sup> 

<sup>1</sup>Auckland University of Technology, New Zealand

<sup>2</sup>The University of Auckland, New Zealand

## Correspondence

Stephanie Gomes-Ng, Department of Psychology, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand.

Email: [stef.gomes-ng@aut.ac.nz](mailto:stef.gomes-ng@aut.ac.nz)

**Editor-in-Chief:** Mark Galizio

**Handling Editor:** Ray Pitts

## Abstract

Organisms may sometimes behave as if a contingency exists between behavior and consequences, even if this is not actually the case. Killeen (1978) suggested that such *superstition* occurs because of factors that bias subjects to behave “superstitiously” rather than because of failures of discrimination. We systematically replicated Killeen’s experiment and compared contingency discrimination between different consequences. Six pigeons responded in a matching-to-sample procedure in which a response-independent or response-dependent stimulus change, food delivery, or blackout occurred. The pigeons reported whether the consequence was response dependent or response independent by choosing between two side keys. Discrimination was strongest after stimulus changes, weaker after blackouts, and weakest after food deliveries. These differences persisted even after additional training, suggesting asymmetries that may reflect differences in the disruptive effects of different consequences on remembering and/or behavioral mnemonics. Importantly, the pigeons were not biased to report response-dependent consequences unless that response was consistent with locational biases; that is, they behaved “superstitiously” when there was a reason to be biased to do so. These findings corroborate Killeen’s and demonstrate that behavior may deviate from contingencies not necessarily because subjects cannot discriminate those contingencies but because they are biased to behave otherwise.

## KEYWORDS

consequence control, contingency discrimination, matching to sample, pigeon, superstition

Behavior typically comes under joint control of stimuli and consequences (Skinner, 1938; see also Davison & Nevin, 1999). In general, behaviors followed by appetitive consequences (*reinforcers*) are more likely to be emitted in the future, whereas behaviors followed by aversive consequences (*punishers*) are less likely to occur. For example, if a pigeon’s key pecks are followed by food in the presence of a blue but not a yellow stimulus, the pigeon will typically respond to the blue stimulus and not respond (or respond less) to the yellow stimulus. Likewise, a human driver accelerates at green traffic lights because this has, in the past, led to a desired consequence (e.g., arriving at one’s destination) but not at red lights because this is likely to lead to an undesired consequence (e.g., an accident). Prolonged exposure to such

contingencies allows organisms to learn about the structure of the environment and thus behave in accordance with what is likely to happen in the future (e.g., Baum, 2012; Cowie, 2018; Cowie & Davison, 2016, 2020; Killeen & Jacobs, 2017; Shahan, 2017).

Sometimes, subjects behave *as if* contingencies exist between stimuli, behavior, and consequences, even though this is not actually the case. In Skinner’s (1948) study of “superstition” in pigeons, response-independent food was delivered every 15 s and the pigeons came to emit idiosyncratic “pigeon-like” behaviors (e.g., head-bobbing, pecking) despite no programmed contingency between such behaviors and food (see also Catania & Cutts, 1963; Davis & Hubbard, 1972; Herrnstein, 1966; Herrnstein & Morse, 1957; Justice & Looney, 1990; Ono, 1987; Rey

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Journal of the Experimental Analysis of Behavior* published by Wiley Periodicals LLC on behalf of Society for the Experimental Analysis of Behavior.

et al., 2020; Sheehan et al., 2012; Wagner & Morris, 1987). Similarly, Morse and Skinner (1957) found that changing the color of a stimulus produced concomitant changes in response rates, even though there was no change in reinforcer rates. Likewise, humans often report having some control over producing or preventing response-independent outcomes (e.g., Aeschleman et al., 2003; Bloom et al., 2007; Hayashi & Modico, 2019; Matute, 1995; Mellon, 2009; Rudski, 2000, 2001; Rudski et al., 1999).

Such superstitious behavior may arise due to adventitious pairing of behavior and reinforcers or of a change in stimulus conditions and a brief change in local reinforcer rates (Morse & Skinner, 1957). If a behavior happened to occur before a food delivery or a stimulus change coincided with either food delivery or a period without food, food (or its absence) would have *appeared* contingent on the behavior or stimulus change. In support of this, delivering three response-dependent reinforcers appears sufficient to maintain responding even after the response–reinforcer dependency is removed, suggesting that superstitious behavior develops readily after only a few correlations between behavior and reinforcers (e.g., Fenner, 1980; Gleeson et al., 1989; Neuringer, 1970). This explanation implies a “gross breakdown in an organism’s ability to detect causality” (Killeen, 1978, p. 88)—that is, behavior fails to come under the control of arranged contingencies.

In contrast, other research suggests that adventitious reinforcers are not responsible for or play a smaller role than originally thought in superstitious responding (e.g., Herrnstein & Morse, 1957; Schwartz, 1973; Staddon & Simmelhag, 1971; Zeiler, 1968). For example, Staddon and Simmelhag (1971) replicated Skinner’s (1948) experiment and found that pigeons engaged in similar sequences of *interim* behaviors during the first two thirds of interfood intervals and *terminal* behaviors thereafter. Importantly, there was little evidence of accidental pairings between terminal behaviors and food deliveries. Staddon and Simmelhag suggested that superstitious response patterns do not reflect adventitious reinforcer processes but rather species-specific food-related behavior (e.g., pecking for a pigeon) *induced* by periodic reinforcer schedules (see also Fenner, 1980; Fernandez & Timberlake, 2020; Innis et al., 1983; Staddon, 1977; Timberlake & Lucas, 1985).

Furthermore, other research shows that response rates differ when consequences are arranged dependent on compared with independent of responding. Appel and Hiss (1962, see also Powell & Kelly, 1979, Rachlin & Baum, 1972) found that when response-independent and response-dependent food was delivered in separate components of a multiple schedule, pigeons’ response rates were low and declined across sessions in the response-independent component, whereas response rates remained high and stable in the response-dependent component. They concluded that subjects could discriminate<sup>1</sup> between components, but the sustained low response

rates in the response-independent component indicated that such discrimination was imperfect (but see Herrnstein, 1966). Similarly, when a response-dependent reinforcer schedule is changed to a response-independent schedule, response rates decrease (e.g., Lachter, 1971; Skinner, 1938; Zeiler, 1968; see Catania & Keller, 1981, for a brief overview). Catania and Keller (1981) further showed that even when response-dependent reinforcer schedules were modified such that they shared similarities with response-independent schedules (e.g., by inserting fixed or variable delays before reinforcer deliveries or by varying interreinforcer or response–reinforcer intervals), response rates were still higher than in response-independent schedules, suggesting some sensitivity to contingencies. Even stronger evidence for differential control by response-dependent and response-independent consequences comes from studies demonstrating that subjects can report relatively accurately which reinforcer schedule or response produced a reinforcer (Davison, 2018; Dodd, 1984; Jones & Davison, 1998; Lattal, 1975, 1979; Nussear & Lattal, 1983) and whether a stimulus change is response dependent or response independent (Killeen, 1978; Killeen & Smith, 1984).

For example, Killeen (1978) arranged a matching-to-sample (MTS) procedure in which pigeons reported whether or not they had caused a stimulus change. In Killeen’s procedure, pecks to an illuminated center key sometimes turned it off and illuminated two side keys, whereas at other times this change occurred independently of responding. Choice of the right key was defined as correct and reinforced after response-dependent stimulus changes, and choice of the left key was correct after response-independent changes. Thus, choosing the right key corresponded to reporting “I did it” and the left key to “I didn’t do it.” Killeen’s pigeons responded correctly in 80%–90% of trials, suggesting that they could discriminate between response-independent and response-dependent consequences. The likelihood of reporting a response-independent (or response-dependent) consequence depended on the reinforcer magnitude for making such a response; the pigeons were biased toward the response that produced a larger reinforcer. Additionally, responding depended on the time between the pigeon’s last peck and consequence, with shorter times increasing the likelihood of reporting “I did it.” Killeen and Smith (1984) further showed that pigeons could report whether they had caused either brief hopper-light presentations (the magazine aperture was illuminated, but no food delivered) or food deliveries. However, accuracy was worse after food deliveries and worsened as the duration of food deliveries increased.

Along similar lines, Chatlosh et al. (1985) asked college students to rate the degree to which their behavior (tapping a telegraph key) caused an appetitive (point

<sup>1</sup>Here, we use the word “discriminate” to mean that behavior “comes under stimulus control” (i.e., behavior is differential with respect to different stimuli).

gain), aversive (point loss), or neutral (no change in points) outcome. Across conditions, Chatlosh et al. varied the probability that a response produced the outcome. In all conditions, participants' ratings roughly matched this probability, suggesting that they were sensitive to behavior–consequence contingencies (see also Wasserman et al., 1983). Other studies have shown that people are more likely to report causal relations between their behavior and consequences when the delay between behavior and consequence is shorter (e.g., Rudski, 2000; Shanks et al., 1989; Wasserman & Neunaber, 1986), when the perceived or actual value of the consequence is higher (e.g., Benvenuti et al., 2018) and when prior knowledge or instructions bias them in favor of reporting a causal relation (e.g., Blanco et al., 2020; Collins & Shank, 2002, 2006; Matute, 1996). Collectively, the above findings suggest that behavior may sometimes appear to deviate from arranged behavior–consequence contingencies because of sources of *bias* (e.g., reinforcer magnitude, instructions delays between responses and reinforcers) rather than because subjects have failed to learn, or are insensitive to those contingencies. Killeen and Smith's (1984) findings also tentatively suggest asymmetries depending on the type of consequence (e.g., stimulus change vs. food delivery). In further support of such asymmetries, Davison (2018) showed that pigeons could more accurately report whether a left- or right-key response produced a stimulus change compared with a 3-s food delivery or blackout. Additionally, human participants sometimes appear more likely to report contingencies between their behavior and reinforcers than punishers (e.g., Alloy & Abramson, 1979; Chatlosh et al., 1985; Rudski et al., 1999). Rasmussen and Newland (2008) found that the effects of money loss on human choice were three times greater than the effects of money gain, suggesting differences between reinforcer and punisher control (see also Gonçalves & Silva, 2015; Kubanek et al., 2015; Kuroda et al., 2018; Pietras et al., 2010). Taken together, these findings suggest there is good reason to expect differences in discrimination between different types of response-independent and response-dependent consequences.

A relatively large body of research has examined the variables that influence superstitious behavior and judgments of contingency in humans. In contrast, besides Killeen (1978) and Killeen and Smith (1984), we know of no other published studies that have asked whether nonhuman subjects can report accurately whether consequences are dependent on or independent of responding. Further, only one of these published studies (Killeen & Smith, 1984) arranged different types of consequences. Although Davison (2018) asked a similar question, his subjects reported the location producing a consequence, not whether the consequence depended on behavior. Studies demonstrating differential response rates with response-dependent and response-independent consequences (e.g., Appel & Hiss, 1962) provide only tentative evidence for such discrimination because there are other

reasons why response rates may differ (e.g., local reinforcer distributions). Thus, at present, research comparing discrimination of different types of response-independent and response-dependent consequences is relatively limited.

Therefore, the present experiment investigated pigeons' discrimination of response-independent and response-dependent consequences in a systematic replication of Killeen (1978). Pigeons responded in a MTS procedure in which a response-dependent or response-independent consequence occurred, and they reported whether the consequence was dependent on or independent of responding by choosing a left or right key. Across conditions, the consequence was a stimulus turning off, food delivery, or blackout. Because preliminary analyses showed that accuracy was low after food deliveries, we ran additional conditions with forced-choice trials (i.e., only the correct choice was illuminated). This helped to clarify whether low accuracy after food deliveries was related to poor control by the matching contingencies or to the nature of the consequence.

## METHOD

### Subjects

Six pigeons, numbered 61 to 66, served as subjects. The pigeons were maintained at  $85\% \pm 15$  g of free-feeding body weight by supplementary feeding of mixed grain at around 9:30 a.m. daily. They were housed in a colony room in a time-shifted environment (lights on at 12 a.m., off at 4 p.m. daily). Water and grit were freely available. All pigeons had previous experience in one (currently unpublished) operant experiment, which consisted of differential reinforcement training followed by extinction and relapse tests. Thus, before the present experiment, the pigeons had never had experience with MTS (or similar) procedures.

### Apparatus

The pigeons were housed individually in home cages, which also served as experimental chambers. Each chamber measured 380 mm high, 380 mm wide, and 380 mm deep and contained two perpendicular wooden perches, one parallel to the front wall and the other parallel to the right wall. The perches were 90 mm away from their respective walls and were mounted 60 mm above the cage floor. An operant panel was mounted on the right wall of each chamber, 220 mm above the perches. The operant panel consisted of three 20-mm (diameter) response keys arranged horizontally and centered 85 mm apart. The center key could be lit blue, and the side keys could be lit red or orange. Responses exceeding 0.1 N to illuminated keys were recorded. A magazine aperture (50 mm high, 50 mm wide, 40 mm deep) was located 100 mm beneath the center key. A hopper filled with wheat was located behind the

**TABLE 1** Sequence of conditions

Condition	Label	Brief description	Sessions
1	Stimulus-off condition	Center-blue stimulus turned off	30
2	Food condition	3-s food delivery	30
3	Blackout condition	3-s blackout	30
4	Food-correction condition	Same as food condition, except incorrect responses were followed by correction trials	30
5	Food-forced condition	Same as food condition, except with forced-choice trials with only the correct comparison lit. The proportion of forced-choice trials was gradually reduced.	60
6	Blackout-forced condition	Same as blackout condition, except with forced-choice trials. The proportion of forced-choice trials was gradually reduced.	51

*Note.* Consequences occurred either dependent on or independent of responding. See text for further details.

magazine aperture. During reinforcer deliveries, the hopper was raised for 3 s, the keylights were darkened, and the magazine aperture was lit. In an adjacent room, a computer running MED-PC IV software ran the experiment and recorded all experimental events. No personnel entered the colony room during experimental sessions.

## Procedure

The pigeons were experimentally experienced, so no pre-training was necessary. Experimental sessions began at 3 a.m. daily for all pigeons and ended after 80 trials or 1 hr, whichever occurred first. In each session, the pigeons completed MTS trials in which the sample was a response-dependent or response-independent consequence, and the pigeons reported whether or not they had caused the consequence by choosing between left and right keys (the comparison stimuli). Trials were separated by a 10-s intertrial interval (ITI). Table 1 summarizes the order of conditions. Conditions lasted for 30 sessions, except for Conditions 5 and 6, which were longer due to additional training with forced-choice trials.

### Condition 1 (stimulus-off condition)

In Condition 1, each MTS trial began with the center key lit blue, and a variable-interval (VI) or variable-time (VT) 30-s schedule operated. This schedule arranged the consequence, which was the center key turning off. In each trial, the schedule was chosen randomly, with the only constraint being that no more than 40 trials of each schedule type occurred in a session. This constraint ensured that there were equal numbers of trials with each schedule in each session. In trials with a VI schedule, the first peck to the center key turned it off after the schedule arranged the consequence. In trials with a VT schedule, the center key turned off as soon as the schedule arranged the consequence. Hereafter, we term these trials *response-dependent* and *response-independent* trials, respectively. After the center key turned off, the left- and right-side keys were lit orange. A peck to the left key was defined

as correct in response-dependent trials and a peck to the right key was defined as correct in response-independent trials for Pigeons 61 to 63. The reverse was arranged for Pigeons 64 to 66. Correct responses were reinforced with a probability of .50. If no reinforcer was arranged, a correct response was followed immediately by the ITI. Incorrect responses were followed by a 3-s blackout and then the ITI (i.e., a total blackout duration of 13 s).

Reinforcers for correct responses were arranged dependently (Stubbs & Pliskoff, 1969). This meant that if a reinforcer was arranged for a particular trial type (e.g., a response-dependent trial), that reinforcer was held and no further reinforcers could be arranged until such a trial occurred and the pigeon responded correctly. This ensured that the overall obtained reinforcers for correct responding were similar in response-dependent and response-independent trials.

### Conditions 2 and 3 (food and blackout conditions)

Conditions 2 and 3 were the same as Condition 1, except that the consequence was either 3-s food delivery (Condition 2) or 3-s blackout (Condition 3). In these conditions, the consequence occurred after the center key turned off and before the presentation of the comparison stimuli.

### Condition 4 (food-correction condition)

Condition 4 replicated the food condition (Condition 2), except that correction trials occurred after incorrect responses. Thus, after an incorrect response, the same trial repeated, except that only the correct comparison was lit.

### Conditions 5 and 6 (food-forced and blackout-forced conditions)

Conditions 5 and 6 replicated Conditions 2 and 3, respectively, except that conditions began with training sessions in which only the correct comparison was lit

(forced-choice trials). After 10 sessions, the proportion of forced-choice trials was gradually reduced in increments of .25 every 10 sessions (Condition 5) or seven sessions (Condition 6). After this proportion reached 0 (i.e., all trials were free-choice trials in which both comparison keys were lit), the condition ran for 20 sessions.

## Data analysis

Visual inspection (Perone, 1991) of percentage correct in response-dependent and response-independent trials in each session indicated that performance was reasonably stable within the last 10 sessions of each condition. To illustrate, Appendix Figure A1 shows bias-free summary measures of performance, calculated using Equation 1 (see below) over the last 15 sessions of each condition. We separated the last 12 sessions into four blocks of three sessions each, and nonparametric trend tests (Elliffe & Elliffe, 2019; Kendall, 1955) on median  $\log d$  values across blocks were nonsignificant (all  $p > .10$ ). Thus, although some fluctuations in performance were evident for some pigeons, these were generally not monotonic.

Data from the last 10 sessions of each condition were aggregated and then separated according to the type of consequence (stimulus off, food delivery, or blackout) and whether the consequence was dependent on or independent of responding. Appendix Table B1 shows these data. We then calculated a bias-free measure of performance,  $\log d$  (Davison & Nevin, 1999; Davison & Tustin, 1978), for each consequence:

$$\log d_x = 0.5 \log \left( \frac{B_{\text{corr} | \text{dependent}}}{B_{\text{err} | \text{dependent}}} \cdot \frac{B_{\text{corr} | \text{independent}}}{B_{\text{err} | \text{independent}}} \right), \quad (1)$$

where  $x$  represents the type of consequence, and  $B_{\text{corr} | y}$  and  $B_{\text{err} | y}$  represent the total number of correct (corr) and incorrect (err) responses, respectively, in response-dependent and response-independent trials ( $y$ ), in the last 10 sessions of each condition. More positive values of Equation 1 indicate more correct than incorrect responses (i.e., more accurate responding). A value of zero indicates chance performance.

We also calculated a related measure,  $\log b$ , which quantifies bias toward one comparison over the other:

$$\log b_x = 0.5 \log \left( \frac{B_{\text{dependent} | \text{dependent}}}{B_{\text{independent} | \text{dependent}}} \cdot \frac{B_{\text{dependent} | \text{independent}}}{B_{\text{independent} | \text{independent}}} \right), \quad (2)$$

where  $x$  represents the type of consequence, and  $B_{x | y}$  represents the number of responses made to the comparison matching dependent or independent consequences ( $x$ ) in response-dependent and response-independent trials ( $y$ ). Thus, more positive values of Equation 2 indicate a bias to report response-dependent consequences (i.e., to report

“I did it”) and more negative values indicate a bias to report response-independent consequences (i.e., “I didn’t do it”). A  $\log b$  value of zero indicates no systematic bias.

We also assessed whether response rates to the center-blue key during the sample phase (hereafter, for brevity, we term these “sample response rates”) and the delay between a response and the delivery of a consequence influenced discrimination. For example, the pigeons may have been more likely to report that a consequence was dependent on responding if the delay between their most recent response and that consequence was shorter (see Killeen, 1978). Sample response rates were calculated by dividing the number of responses made to the center-blue key (aggregated across response-dependent and response-independent trials because the consequence had not yet been presented) by the total time spent in the sample phase in each condition. To assess the effects of response–consequence delays on discrimination, we separated data from response-independent trials into trials in which the delay was less than 0.25 s, trials in which the delay was between 0.25 s and 0.50 s inclusive, and trials in which the delay was greater than 0.50 s. We then calculated the proportion of correct responses for each group of trials. We used only data from response-independent trials for this analysis because there was never a delay between a response and consequence in response-dependent trials.

## RESULTS

In the present experiment, pigeons reported whether they had or had not caused a stimulus turning off, food delivery, or blackout. Figure 1 shows accuracy ( $\log d$ ; Equation 1) and bias ( $\log b$ ; Equation 2) for individual pigeons after each consequence. Mean  $\log d$  and  $\log b$  values are also plotted in Figure 1. Accuracy followed similar patterns across conditions for all pigeons, with only a few individual-pigeon exceptions. All pigeons displayed a bias toward the right key, as evidenced by the negative  $\log b$  values for Pigeons 61 to 63 and positive  $\log b$  values for Pigeons 64 to 66 (these opposite patterns occurred because the key locations were counterbalanced across pigeons). There appeared to be little systematic change in bias across conditions (Friedman test:  $\chi^2 = 6.86$ ,  $p = .232$ ), although bias did appear somewhat stronger in conditions with food deliveries. Because bias varied unsystematically across pigeons and conditions, we focus on accuracy henceforth.

Accuracy ( $\log d$ ) differed depending on the type of consequence (Figure 1); a Friedman test (a nonparametric alternative to the one-way repeated-measures analysis of variance) on  $\log d$  values was significant ( $\chi^2 = 17.05$ ,  $p = .004$ ). In general, accuracy was highest in the stimulus-off condition (black bars), suggesting that the pigeons could discriminate whether they had caused the stimulus to turn off. In comparison,

accuracy was lower after blackouts (blackout and blackout-forced conditions; gray bars) and even lower after food deliveries (food, food-correction, and food-forced conditions; white bars). Post hoc comparisons indicated that only the differences in accuracy between the stimulus-off and food or food-correction conditions were significant (both comparisons:  $p = .030$ ).

Correction trials (food-correction condition; dotted white bars) had little effect on accuracy, whereas forced-choice trials (food-forced and blackout-forced conditions; striped bars) increased accuracy (Figure 1). Nevertheless, these improvements were not significant (binomial sign tests: all  $p = .344$ ), and accuracy was still lower than in the stimulus-off condition. Thus, differences in accuracy after stimulus changes, blackouts, and food deliveries appeared to reflect the different types of consequences rather than poor control by the matching contingencies (i.e., which response was correct in response-dependent and response-independent trials).

Figure 2 shows response rates to the center-blue key during the sample phase, combined across response-independent and response-dependent trials. Overall, sample response rates were highest when the sample phase was followed by food delivery (food, food-correction, and food-forced conditions) and lowest when it was followed by blackout (blackout and blackout-forced conditions). There were few differences in sample response rates between conditions with and without correction or forced-choice trials, although response rates did appear to be higher in the food-correction condition. Sample response rates were negatively correlated with  $\log d$  values in the blackout condition (Spearman's rank correlation:  $r = -0.94$ ,  $p = .017$ ); pigeons with higher response rates had lower  $\log d$  values in this condition. Besides this, there were no other correlations between sample response rates and  $\log d$  or  $\log b$  values in any other conditions. Thus, overall there was little systematic relation between sample responding and accuracy or bias across conditions.

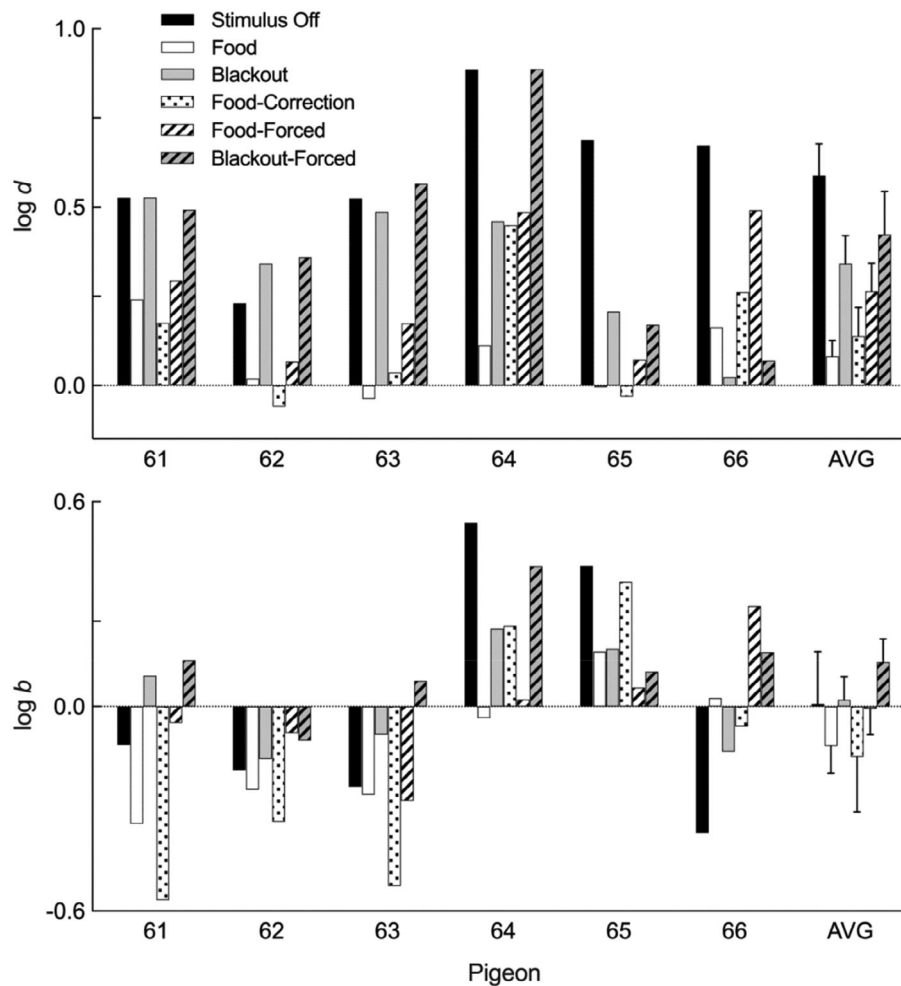
Figure 3 shows the proportion of correct responses in response-independent trials as a function of the time between the pigeon's most recent response and the delivery of the response-independent consequence. In the stimulus-off condition, accuracy increased as the time between a response and consequence increased. This also appeared to be the case in conditions with food deliveries and blackouts, although the effects of delay were much smaller and less systematic after food deliveries. These trends were generally not monotonic, and thus nonparametric trend tests (Elliffe & Elliffe, 2019; Kendall, 1955) on proportion correct were nonsignificant, except for the food-forced condition ( $\sum S = 14$ ,  $p = .002$ ). Nevertheless, overall the pigeons were generally more likely to report that they had *not* caused a consequence if there was a delay between their most recent response and a consequence, and this effect was more pronounced when the consequence was a stimulus turning off or a blackout, compared with a food delivery.

## DISCUSSION

Killeen (1978) asked whether failures of discrimination underlie superstitious responding and answered the question by showing that pigeons could accurately indicate whether they had caused a stimulus change. The likelihood of reporting a response-dependent stimulus change increased when the time between the pigeon's last peck and the stimulus change was shorter and when larger reinforcers were obtained for reporting response-dependent changes. Thus, Killeen suggested that superstitious responding arises not because of failures of discrimination but because subjects are sometimes biased to behave *as if* a behavior–consequence contingency exists (e.g., due to variables that produce such bias, such as reinforcer magnitude). The present study systematically replicated Killeen's; like Killeen's pigeons, our pigeons reported relatively accurately whether they caused a stimulus to turn off (Figure 1, top panel), and such discrimination depended on the time between a response and consequence (Figure 3). Discrimination was worse after blackouts and even worse—close to chance—after food deliveries, even when procedural variations were made to enhance discrimination learning. Bias varied unsystematically across pigeons and conditions (Figure 1, bottom panel). Thus, the type of consequence and response–consequence delay appeared to affect discrimination but not bias.

That accuracy was low—suggesting poor discrimination—in conditions with food deliveries appears inconsistent with Killeen's (1978) conclusion that superstition is not a matter of detectability. That is, our findings appear to suggest that failures of discrimination underlie superstition when the consequence is a reinforcer delivery. However, if superstitious responding occurs because subjects incorrectly attribute consequences to their behavior, they should be biased to report that consequences are response-dependent. No such bias was evident here; our pigeons were biased toward the right key, perhaps because the perch configuration made it easier to peck (we have observed similar right-key biases in other experiments in our lab; e.g., Krägeloh et al., 2005). The absence of systematic bias to report response-dependent consequences questions the view that superstitious behavior arises because subjects incorrectly attribute consequences to their behavior. Instead, our findings corroborate Killeen's by suggesting that subjects are not necessarily more likely to report response-dependent consequences (to behave superstitiously) unless there are reasons to be biased toward that response. This is consistent with a large body of research demonstrating that discrimination and bias are separable aspects of conditional-discrimination performance (e.g., Alsop & Davison, 1991; Davison & Nevin, 1999; McCarthy & Davison, 1979, 1980; Stubbs, 1976).

The present findings add to the literature highlighting the important role of behavior–consequence

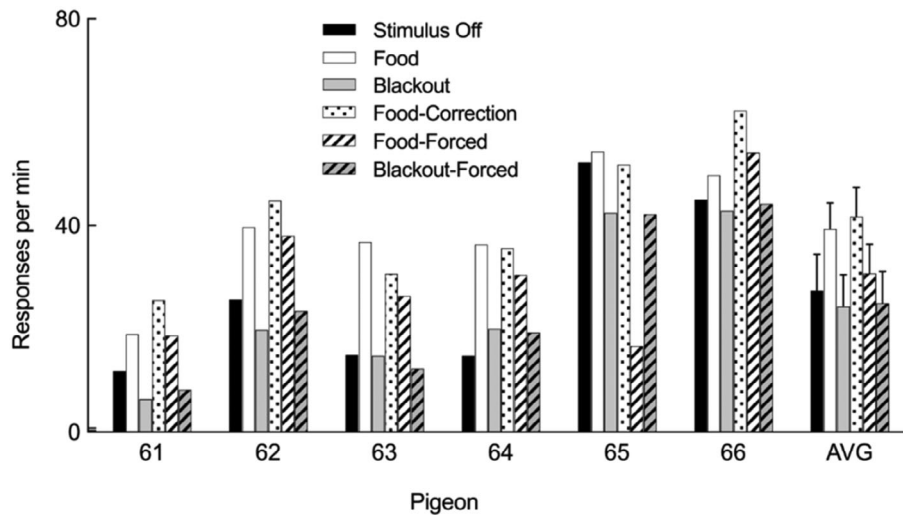


**FIGURE 1** Individual-pigeon and mean accuracy and bias across conditions. The right key was correct after response-independent consequences for Pigeons 61 to 63, whereas the right key was correct after response-dependent consequences for Pigeons 64 to 66. Accuracy was calculated using Equation 1 ( $\log d$ ), and bias was calculated using Equation 2 ( $\log b$ ). The colors of the bars represent the three types of consequences—stimulus turning off (black), food delivery (white), or blackout (gray), and patterned bars are from conditions with correction trials (dots) or forced-choice trials (diagonal lines). The error bars for averaged (AVG) data show the *SEM*.

intervals in discriminative control by response-dependent and response-independent consequences. Like Killeen (1978; Killeen & Smith, 1984), we found that pigeons were more likely to report that a consequence was independent of their responding when the time between pigeons' last response and the delivery of a consequence was longer (Figure 3). However, it is unlikely that behavior–consequence intervals *alone* determine differential control by response-dependent and response-independent reinforcer schedules (Catania & Keller, 1981). Catania and Keller suggested that such differential control depends on a combination of features—including learning history, baseline response rates, response–reinforcer intervals, and successive reinforcer deliveries without intervening reinforcers. Even so, subjects appear remarkably sensitive to behavior–consequence contingencies: In a series of experiments, Catania and Keller showed that pigeons' response rates were higher in response-dependent schedules than

response-independent schedules, even when delays occurred between responses and reinforcers and/or successive reinforcers occurred without intervening responses in the response-dependent schedules. In other words, response rates were differential, even when experimental manipulations made response-dependent and response-independent schedules more similar to each other. This adds further support to Killeen's (1978) and our conclusion that superstitious responding is not necessarily a failure of discrimination.

Recently, Jin et al. (2022) suggested that superstitious behavior may develop and be maintained by *generalization* processes. According to this view, subjects may behave as if contingencies exist between their behavior and consequences because of generalization from previous experience with response-dependent consequences. Indeed, many experimentally experienced subjects (like those in the current study) have prior experience with response-dependent consequences.



**FIGURE 2** Individual-pigeon and mean sample response rates. The colors of the bars represent the three types of consequences—stimulus turning off (black), food delivery (white), or blackout (gray), and patterned bars are from conditions with correction trials (dots) or forced-choice trials (diagonal lines). The error bars for averaged (AVG) data show the *SEM*.

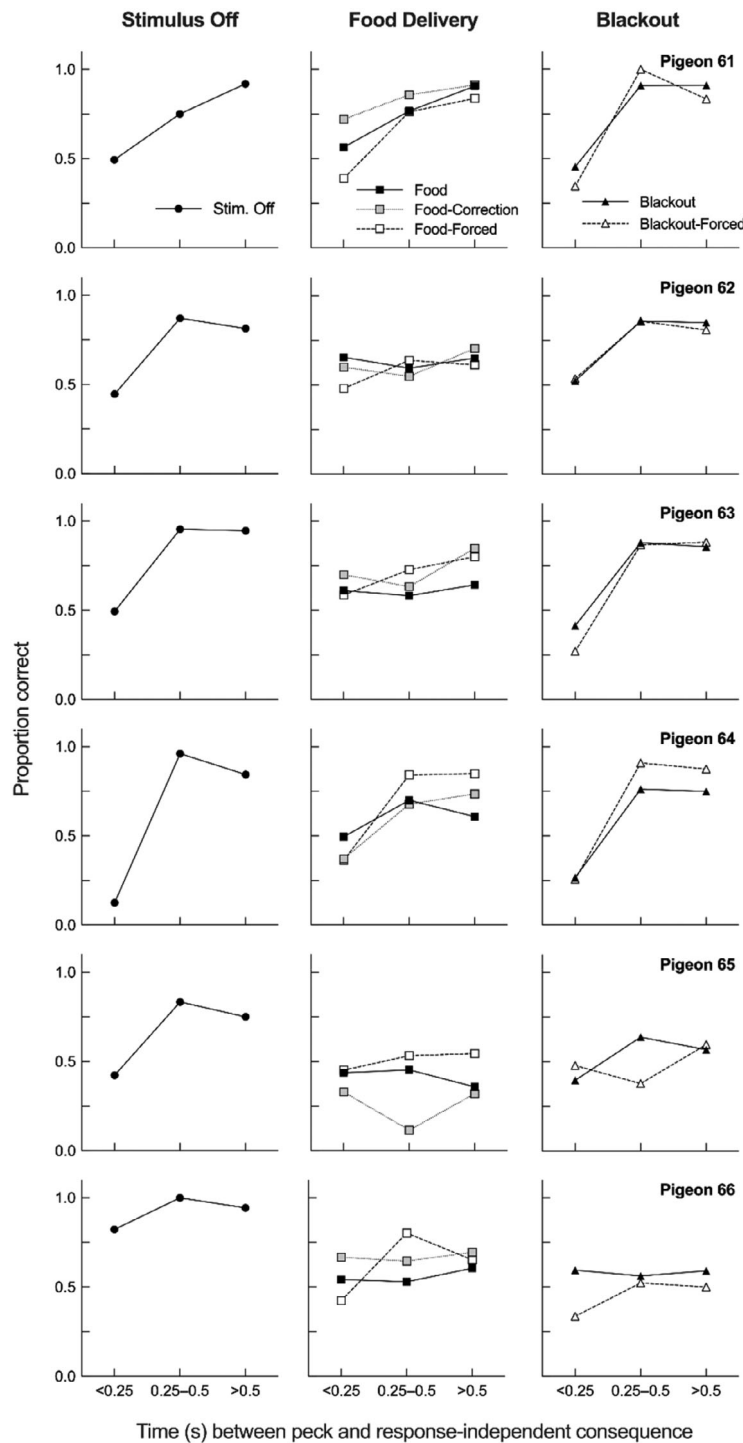
Relatedly, organisms may be predisposed to behave as if consequences depend on behavior because such a predisposition confers an evolutionary advantage in the wild; in other words, assuming causal relations between events may be adaptive (see, e.g., Beck & Forstmeier, 2007; Foster & Kokko, 2008, for related discussions). Although we did not set out to test such accounts of superstitious behavior, our findings—as well as Killeen's (1978; Killeen & Smith, 1984)—appear somewhat inconsistent with them because biases to report response-dependent consequences were only apparent when *other* factors (e.g., key location in our experiment, reinforcer magnitude in Killeen, 1978) produced such biases.

The present study is also relevant to research comparing the effects of different consequences on behavior. There is debate over whether reinforcers and punishers have symmetrical effects, with some evidence suggesting they have equivalent but opposite effects (e.g., Broughton & Lahey, 1978; Donaldson et al., 2014; Kaufman & O'Leary, 1972; Lie & Alsop, 2009) and other research suggesting asymmetries (e.g., Gonçalves & Silva, 2015; Kubanek et al., 2015; Kuroda et al., 2018; Lie, 2010; Pietras et al., 2010; Rasmussen & Newland, 2008). Our findings are consistent with the latter—discrimination differed between consequences, even with additional training (forced-choice trials). If discrimination is a critical element of learning (e.g., Cowie & Davison, 2016), these findings imply reinforcers may facilitate learning less than punishers. This is consistent with Rasmussen and Newland's (2008) suggestion that punishers have stronger effects than reinforcers and with applied studies demonstrating that punishment procedures are sometimes more effective than reinforcement-only procedures (e.g., see Holz & Azrin, 1963; Johnson et al., 1973; Rawson &

Leitenberg, 1973; for applied studies, see Barrett et al., 1981; Favell et al., 1982; Scotti et al., 1991).

Why might contingency discrimination differ depending on the consequence? One procedural variable that probably contributed to our results is consequence duration—the stimulus turned off immediately, whereas food deliveries and blackouts lasted for 3 s. Thus, comparison stimuli were presented immediately in the stimulus-off condition, whereas there was a 3-s delay in food and blackout conditions. This delay is probably partially responsible for lower accuracy in food and blackout conditions, as accuracy tends to be lower in conditional discriminations with longer delays (see Lind et al., 2015, for a review). However, this cannot explain differences between food and blackout conditions, in which consequence durations were equal and yet accuracy differed. Similarly, other studies have found lower accuracy after food deliveries than after equivalent-length hopper-light presentations (Killeen & Smith, 1984) or blackouts (Davison, 2018). Thus, differences in consequence duration notwithstanding, these findings point to fundamental differences in consequence control when the consequence is food delivery, compared with other consequences.

Killeen and Smith (1984) hypothesized that food deliveries disrupt memory of the stimulus or response that produced them; specifically, food deliveries increase the speed at which the memory trace decays, resulting in faster forgetting. In support of this, accuracy in delayed MTS is lower when food is delivered during the interval between sample and comparison stimuli (Jans & Catania, 1980) and control by local reinforcer ratios is weaker when such ratios depend on the last-food location than when they do not (Cowie et al., 2011, 2017). That accuracy was lowest in our food conditions is consistent with this account. This implies that other consequences such as hopper-light presentations (as in Killeen &



**FIGURE 3** Effects of delay between response and consequence on accuracy in response-independent trials

Smith, 1984) or blackouts (as in Davison, 2018, and this experiment) do not disrupt remembering to the same extent.

Alternatively, a more parsimonious explanation is that food deliveries affect behaviors implicated in performance. In MTS procedures, subjects may learn to orient toward the correct comparison (e.g., Davison, 2018; Gomes-Ng et al., 2019) or engage in other mediating behaviors that bridge the gap between sample and

comparison stimuli (e.g., Blough, 1959; Urcuioli, 1984, 1985; Urcuioli & Honig, 1980; Weaver et al., 1999). For example, Urcuioli (1985) showed that comparison choice was controlled by differential behavior during the sample phase rather than by the identity of the sample, so disrupting sample-specific behaviors reduced matching performance. In the present procedure, food deliveries may have disrupted any behavioral mnemonics (e.g., orienting toward the correct comparison)—thus reducing

performance—because subjects must turn away from the response keys to eat from the food hopper. In contrast, other consequences (e.g., blackouts) were probably less disruptive because they required no change in body position (see Davison, 2018, for a similar argument).

Relatedly, primary reinforcers, such as food deliveries, may have unique properties because they are *phylogenetically important events* (PIEs) which, under appropriate conditions (e.g., when the organism is hungry), induce PIE-related behaviors that may occur before, during, and after the presentation of the PIE (e.g., pecking, eating; Baum, 2012, 2018; Killeen & Jacobs, 2017; Segal, 1972; Staddon, 1977). In contrast, non-PIE stimuli do not induce such behaviors. Indeed, in the present experiment, sample response rates were higher in food conditions than in stimulus-off and blackout conditions (Figure 2), even though the same VI or VT schedule operated during the sample phase. The higher sample response rates in food conditions may reflect the fact that the center-blue stimulus reliably predicted food delivery and thus induced pecking upon its presentation in those conditions (see, e.g., Baum, 2012; Staddon, 1977, for discussion). To the extent that food deliveries induced food-related behaviors in the present experiment, these behaviors may have disrupted and/or competed with matching performance (e.g., by disrupting behavioral mnemonics). Therefore, differences in discrimination between the present food and blackout conditions may reflect the differential effects of food deliveries and blackouts on adjunctive (i.e., induced) behaviors (e.g., eating) and differential disruption of behavioral mnemonics (e.g., orienting).

This line of reasoning implies that (1) nonfood consequences may have similarly disruptive effects as food deliveries if they are PIEs, are paired with PIEs, or engender additional behaviors and (2) disruptive effects of food deliveries should be weaker when behavioral mnemonics are not possible. These implications can be tested in future research. For example, instead of blackouts, a future study could present a colored stimulus on a central key and require subjects to peck this stimulus to progress to the comparison phase. This would introduce an additional behavior during the consequence, much like eating happens during food deliveries, and disrupt behavioral mnemonics by forcing subjects to peck a key. If accuracy in such trials is as poor as after food deliveries, this would support the suggestion that food deliveries disrupt performance because they induce additional behaviors. Additionally, a replication of the present experiment could use stimuli previously paired with PIEs. These should induce PIE-related behavior (Baum, 2012), and thus accuracy should be similar to trials with PIEs. It would also be worth including conditions in which the consequence is unpredictable (e.g., either a stimulus change, blackout, or food delivery, chosen probabilistically). Under such conditions, PIE-related behaviors should be less likely to occur (see Staddon, 1977), and thus accuracy should be higher relative to food-only conditions.

Using nonlocation comparison stimuli (e.g., colors) may eliminate orienting behavior under stimulus control of the location of the correct comparison. Indeed, Davison (2018) showed that accuracy after food deliveries and blackouts was more similar with color than location comparisons (see also Wilkie et al., 1985). Thus, with nonlocation comparisons, accuracy should be similar across conditions, at least to the extent that differences in discrimination in the present experiment reflected differential disruption of orienting behavior. Arranging nonlocation comparison stimuli may also reduce locational biases, which may help to clarify systematic differences in bias across consequences.

More generally, what are the features that produce differential control by response-dependent and response-independent consequences? In addition to behavior-consequence intervals (see, e.g., our Figure 3), Catania and Keller (1981) identified several other features that may be important, such as the shape of the behavior-consequence interval distribution, successive consequence deliveries with or without intervening responses, baseline response rates, and learning histories. The present procedure is well equipped to test the effects of such features on discriminative control—for example, a subsequent experiment may vary the shape of the behavior-consequence interval distribution or interpolate fixed or variable delays between responses and consequences—to investigate how these influence the likelihood of reporting response-dependent versus response-independent consequences.

Finally, an interesting avenue for future research is to replicate the current study with humans. Studies investigating humans' judgments of behavior-consequence contingencies typically ask participants to respond to survey questions after a block of trials or at the end of the experimental session. This differs from the present experiment in which subjects reported whether they caused a consequence in each trial, and we calculated separate measures of discrimination ( $\log d$ ; Equation 1) and bias ( $\log b$ ; Equation 2). Thus, replicating the present experiment with humans would provide insight into the cross-species generality of our (and Killeen's, 1978) approach, as well as the extent to which behavioral measures of performance ( $\log d$ ,  $\log b$ ) corroborate self-report measures used in surveys. More importantly, given the large body of existing research investigating superstitious behavior and causal learning in humans, the current approach would help to elucidate further the mechanisms underlying causal learning in humans by allowing researchers to identify and examine closely the variables that affect discrimination versus those that affect bias in contingency-judgment tasks.

In summary, the present findings suggest that pigeons can report relatively accurately whether they caused a stimulus to turn off, whereas discrimination between response-dependent and response-independent consequences is worse after blackouts and even worse after food deliveries. These differences may reflect differential disruptive effects of consequences; for example, food

deliveries may interfere with memory for which response produced them (Cowie et al., 2011, 2017; Jans & Catania, 1980; Killeen & Smith, 1984) or induce behaviors that interfere with behavioral mnemonics and affect performance (Davison, 2018). Importantly, even when discrimination was poor, our pigeons were not more likely to report response-dependent consequences. Thus, behavior that appears superstitious may arise not necessarily because of failures of discrimination but because other sources of bias push subjects to behave *as if* consequences are response dependent (due to, e.g., reinforcer magnitude, response–consequence delays, locational biases; Killeen, 1978). This has practical relevance for behavior analysts, who use consequences—such as food for a pigeon, access to preferred toys for a human, or noncontingent reinforcement in applied interventions—to alter behavior. When subjects' or clients' behavior appears inconsistent with the arranged behavior–consequence contingencies, it is the behavior analyst's job to determine whether this is indeed a problem of discrimination or bias.

## ACKNOWLEDGMENTS

The authors thank the 2021 PSYCH203 and PSYCH309 tutors and students, who helped run the experiment, and Lydia Beetham, who took excellent care of the pigeons. Open access publishing facilitated by Auckland University of Technology, as part of the Wiley - Auckland University of Technology agreement via the Council of Australian University Librarians.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## ETHICS APPROVAL

This experiment was approved by the University of Auckland Animal Ethics Committee (Ref. 2657).

## ORCID

Douglas Elliffe  <https://orcid.org/0000-0001-9449-9540>

## REFERENCES

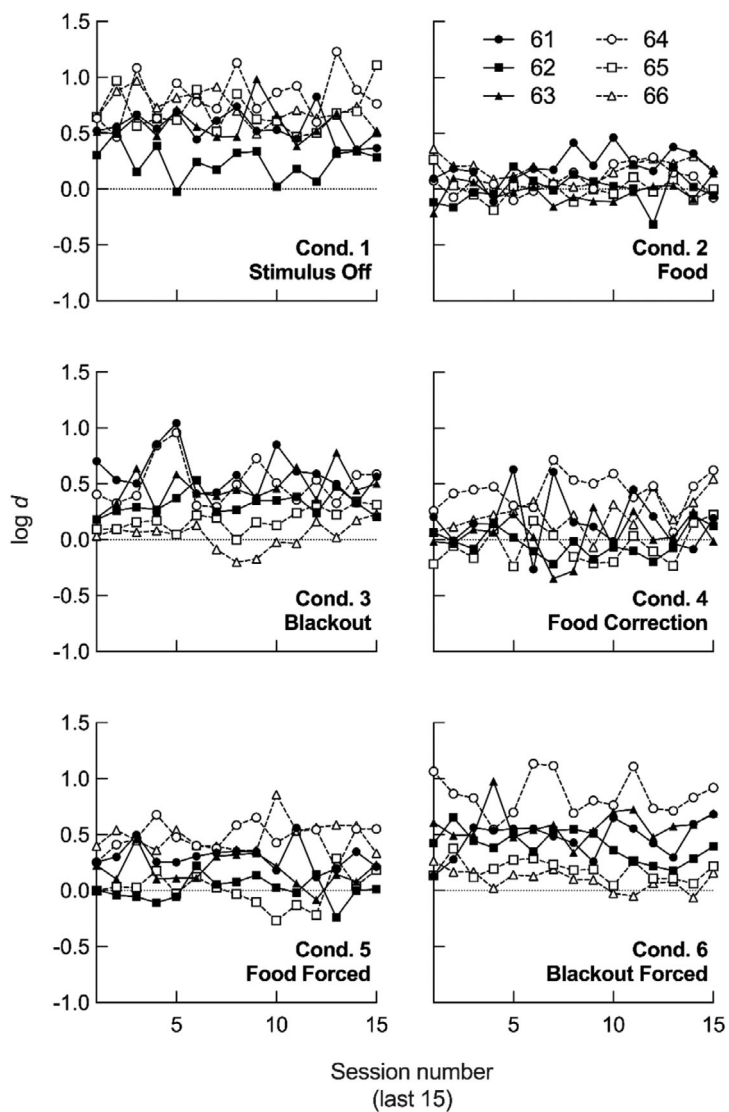
- Aeschleman, S. R., Rosen, C. C., & Williams, M. R. (2003). The effect of non-contingent negative and positive reinforcement operations on the acquisition of superstitious behaviors. *Behavioural Processes*, 61(1–2), 37–45. [10.1016/S0376-6357\(02\)00158-4](https://doi.org/10.1016/S0376-6357(02)00158-4)
- Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, 108(4), 441–485. [10.1037/0096-3445.108.4.441](https://doi.org/10.1037/0096-3445.108.4.441)
- 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. [10.1901/jeab.1991.56-67](https://doi.org/10.1901/jeab.1991.56-67)
- Appel, J. B., & Hiss, R. H. (1962). The discrimination of contingent from noncontingent reinforcement. *Journal of Comparative and Physiological Psychology*, 55(1), 37–39. [10.1037/h0044613](https://doi.org/10.1037/h0044613)
- Barrett, R. P., Matson, J. L., Shapiro, E. S., & Ollendick, T. H. (1981). A comparison of punishment and DRO procedures for treating stereotypic behavior of mentally retarded children. *Applied Research in Mental Retardation*, 2(3), 247–256. [10.1016/0270-3092\(81\)90018-7](https://doi.org/10.1016/0270-3092(81)90018-7)
- Baum, W. M. (2012). Rethinking reinforcement: Allocation, induction, and contingency. *Journal of the Experimental Analysis of Behavior*, 97(1), 101–124. [10.1901/jeab.2012.97-101](https://doi.org/10.1901/jeab.2012.97-101)
- Baum, W. M. (2018). Three laws of behavior: Allocation, induction, and covariance. *Behavior Analysis: Research and Practice*, 18(3), 239–251. [10.1037/bar0000104](https://doi.org/10.1037/bar0000104)
- Beck, J., & Forstmeier, W. (2007). Superstition and belief as inevitable by-products of an adaptive learning strategy. *Human Nature*, 18, 35–46. [10.1007/BF02820845](https://doi.org/10.1007/BF02820845)
- Benvenuti, M. F. L., de Toledo, T. F. N., Simões, R. A. G., & Bizarro, L. (2018). Comparing illusion of control and superstitious behavior: Rate of responding influences judgment of control in a free-operant procedure. *Learning and Motivation*, 64, 27–33. [10.1016/j.lmot.2017.10.002](https://doi.org/10.1016/j.lmot.2017.10.002)
- Blanco, F., Moreno-Fernández, M. M., & Matute, H. (2020). When success is not enough: The symptom base-rate can influence judgments of effectiveness of a successful treatment. *Frontiers in Psychology*, 11, Article 560273. [10.3389/fpsyg.2020.560273](https://doi.org/10.3389/fpsyg.2020.560273)
- Bloom, C. M., Venard, J., Harden, M., & Seetharaman, S. (2007). Non-contingent positive and negative reinforcement schedules of superstitious behaviors. *Behavioural Processes*, 75(1), 8–13. [10.1016/j.beproc.2007.02.010](https://doi.org/10.1016/j.beproc.2007.02.010)
- Blough, D. S. (1959). Delayed matching in the pigeon. *Journal of the Experimental Analysis of Behavior*, 2(2), 151–160. [10.1901/jeab.1959.2-151](https://doi.org/10.1901/jeab.1959.2-151)
- Broughton, S. F., & Lahey, B. B. (1978). Direct and collateral effects of positive reinforcement, response cost, and mixed contingencies for academic performance. *Journal of School Psychology*, 16(2), 126–136. [10.1016/0022-4405\(78\)90051-1](https://doi.org/10.1016/0022-4405(78)90051-1)
- Catania, A. C., & Cutts, D. (1963). Experimental control of superstitious responding in humans. *Journal of the Experimental Analysis of Behavior*, 6(2), 203–208. [10.1901/jeab.1963.6-203](https://doi.org/10.1901/jeab.1963.6-203)
- Catania, A. C., & Keller, K. J. (1981). Contingency, contiguity, correlation, and the concept of causation. In P. Harzem & M. D. Zeiler (Eds.), *Advances in analysis of behaviour Volume 2: Predictability, correlation, and contiguity* (pp. 125–167). John Wiley & Sons.
- Chatlosh, D. L., Nuenaber, D. J., & Wasserman, E. A. (1985). Response–outcome contingency: Behavioral and judgmental effects of appetitive and aversive outcomes with college students. *Learning and Motivation*, 16(1), 1–34. [10.1016/0023-9690\(85\)90002-5](https://doi.org/10.1016/0023-9690(85)90002-5)
- Collins, D. J., & Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition*, 30(7), 1138–1147. [10.3758/BF03194331](https://doi.org/10.3758/BF03194331)
- Collins, D. J., & Shanks, D. R. (2006). Conformity to the power PC theory of causal induction depends on the type of probe question. *Quarterly Journal of Experimental Psychology*, 59(2), 225–232. [10.1080/17470210500370457](https://doi.org/10.1080/17470210500370457)
- Cowie, S. (2018). Behavioral time travel: Control by past, present, and potential events. *Behavior Analysis: Research and Practice*, 18(2), 174–183. [10.1037/bar0000122](https://doi.org/10.1037/bar0000122)
- Cowie, S., & Davison, M. (2016). Control by reinforcers across time and space: A review of recent choice research. *Journal of the Experimental Analysis of Behavior*, 105(2), 246–269. [10.1002/jeab.200](https://doi.org/10.1002/jeab.200)
- Cowie, S., & Davison, M. (2020). Generalizing from the past, choosing the future. *Perspectives on Behavior Science*, 43(2), 245–258. [10.1007/s40614-020-00257-9](https://doi.org/10.1007/s40614-020-00257-9)
- Cowie, S., Davison, M., & Elliffe, D. (2011). Reinforcement: Food signals the time and location of future food. *Journal of the Experimental Analysis of Behavior*, 96(1), 63–86. [10.1901/jeab.2011.96-63](https://doi.org/10.1901/jeab.2011.96-63)
- Cowie, S., Davison, M., & Elliffe, D. (2017). Control by past and present stimuli depends on the discriminated reinforcer differential. *Journal of the Experimental Analysis of Behavior*, 108(2), 184–203. [10.1002/jeab.268](https://doi.org/10.1002/jeab.268)
- Davis, H., & Hubbard, J. (1972). An analysis of superstitious behavior in the rat. *Behaviour*, 43(1/4), 1–12.

- Davison, M. (2018). Divided stimulus control: Which key did you peck, or what color was it? *Journal of the Experimental Analysis of Behavior*, *109*(1), 107–124. [10.1002/jeab.295](https://doi.org/10.1002/jeab.295)
- Davison, M., & Nevin, J. A. (1999). Stimuli, reinforcers, and behavior: An integration. *Journal of the Experimental Analysis of Behavior*, *71*(3), 439–482. [10.1901/jeab.1999.71-439](https://doi.org/10.1901/jeab.1999.71-439)
- Davison, M. C., & Tustin, R. D. (1978). The relation between the generalized matching law and signal-detection theory. *Journal of the Experimental Analysis of Behavior*, *29*(2), 331–336. [10.1901/jeab.1978.29-331](https://doi.org/10.1901/jeab.1978.29-331)
- Dodd, P. W. D. (1984). Discrimination of response-reinforcer and response-stimulus contingencies in pigeons. *Journal of the Experimental Analysis of Behavior*, *41*(1), 7–15. [10.1901/jeab.1984.41-7](https://doi.org/10.1901/jeab.1984.41-7)
- Donaldson, J. M., DeLeon, I. G., Fisher, A. B., & Kahng, S. (2014). Effects of and preference for conditions of token earn versus token loss. *Journal of Applied Behavior Analysis*, *47*(3), 537–548. [10.1002/jaba.135](https://doi.org/10.1002/jaba.135)
- Elliffe, D., & Elliffe, M. (2019). Rank-permutation tests for behavior analysis, and a test for trend allowing unequal data numbers for each subject. *Journal of the Experimental Analysis of Behavior*, *111*(2), 342–358. [10.1002/jeab.502](https://doi.org/10.1002/jeab.502)
- Favell, J. E., Azrin, N. H., Baumeister, A. A., Carr, E. G., Dorsey, M. F., Forehand, R., Foxx, R. M., Lovaas, O. I., Rincover, A., Riskey, T. R., Romanczyk, R. G., Russo, D. C., Schroeder, S. R., & Solnick, J. V. (1982). The treatment of self-injurious behavior. *Behavior Therapy*, *13*(4), 529–554. [10.1016/S0005-7894\(82\)80015-4](https://doi.org/10.1016/S0005-7894(82)80015-4)
- Fenner, D. (1980). The role of contingencies and “principles of behavioral variation” in pigeons’ pecking. *Journal of the Experimental Analysis of Behavior*, *34*(1), 1–12. [10.1901/jeab.1980.34-1](https://doi.org/10.1901/jeab.1980.34-1)
- Fernandez, E. J., & Timberlake, W. (2020). Superstition revisited: Sex, species, and adventitious reinforcement. *Behavioural Processes*, *170*, Article 103979. [10.1016/j.beproc.2019.103979](https://doi.org/10.1016/j.beproc.2019.103979)
- Foster, K. R., & Kokko, H. (2008). The evolution of superstitious and superstition-like behaviour. *Proceedings of the Royal Society B*, *276*, 31–37. [10.1098/rspb.2008.0981](https://doi.org/10.1098/rspb.2008.0981)
- Gleeson, S., Lattal, K. A., & Williams, K. S. (1989). Superstitious conditioning: A replication and extension of Neuringer (1970). *The Psychological Record*, *39*(4), 563–571. [10.1007/BF03395084](https://doi.org/10.1007/BF03395084)
- Gomes-Ng, S., Elliffe, D., & Cowie, S. (2019). Relative reinforcer rates determine pigeons’ attention allocation when separately trained stimuli are presented together. *Learning & Behavior*, *47*(3), 245–257. [10.3758/s13420-019-00381-w](https://doi.org/10.3758/s13420-019-00381-w)
- Gonçalves, F. L., & Silva, M. T. A. (2015). Comparing individual delay discounting of gains and losses. *Psychology & Neuroscience*, *8*(2), 267–279. [10.1037/h0101057](https://doi.org/10.1037/h0101057)
- Hayashi, Y., & Modico, J. G. (2019). Effect of response-independent delivery of positive and negative reinforcers on the development of superstitious behavior and belief in humans. *Behavior Analysis: Research and Practice*, *19*(4), 327–342. [10.1037/bar0000147](https://doi.org/10.1037/bar0000147)
- Herrnstein, R. J. (1966). Superstition: A corollary of the principles of operant conditioning. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application* (pp. 33–51). Appleton-Century-Crofts.
- Herrnstein, R. J., & Morse, W. H. (1957). Some effects of response-independent positive reinforcement on maintained operant behavior. *Journal of Comparative and Physiological Psychology*, *50*(5), 461–467. [10.1037/h0041506](https://doi.org/10.1037/h0041506)
- Holz, W. C., & Azrin, N. H. (1963). A comparison of several procedures for eliminating behavior. *Journal of the Experimental Analysis of Behavior*, *6*(3), 399–406. [10.1901/jeab.1963.6-399](https://doi.org/10.1901/jeab.1963.6-399)
- Innis, N. K., Simmelhag-Grant, V. L., & Staddon, J. E. R. (1983). Behavior induced by periodic food delivery: The effects of inter-food interval. *Journal of the Experimental Analysis of Behavior*, *39*(2), 309–322. [10.1901/jeab.1983.39-309](https://doi.org/10.1901/jeab.1983.39-309)
- Jans, J. E., & Catania, A. C. (1980). Short-term remembering of discriminative stimuli in pigeons. *Journal of the Experimental Analysis of Behavior*, *34*(2), 177–183. [10.1901/jeab.1980.34-177](https://doi.org/10.1901/jeab.1980.34-177)
- Jin, Y., Jensen, G., Gottlieb, J., & Ferrera, V. (2022). Superstitious learning of abstract order from random reinforcement. *Proceedings of the National Academy of Sciences*, *119*(35), Article e2202789119. [10.1073/pnas.2202789119](https://doi.org/10.1073/pnas.2202789119)
- Johnson, D. L., McGlynn, F. D., & Topping, J. S. (1973). The relative efficiency of four response-elimination techniques following variable-ratio reinforcement training. *The Psychological Record*, *23*(2), 203–208. [10.1007/BF03394156](https://doi.org/10.1007/BF03394156)
- Jones, B. M., & Davison, M. (1998). Reporting contingencies of reinforcement in concurrent schedules. *Journal of the Experimental Analysis of Behavior*, *69*(2), 161–183. [10.1901/jeab.1998.69-161](https://doi.org/10.1901/jeab.1998.69-161)
- Justice, T. C., & Looney, T. A. (1990). Another look at “superstitions” in pigeons. *Bulletin of the Psychonomic Society*, *28*(1), 64–66. [10.3758/BF03337650](https://doi.org/10.3758/BF03337650)
- Kaufman, K. F., & O’Leary, K. D. (1972). Reward, cost, and self-evaluation procedures for disruptive adolescents in a psychiatric hospital school. *Journal of Applied Behavior Analysis*, *5*(3), 293–309. [10.1901/jaba.1972.5-293](https://doi.org/10.1901/jaba.1972.5-293)
- Kendall, M. G. (1955). *Rank correlation methods*. Charles Griffin.
- Killeen, P. R. (1978). Superstition: A matter of bias, not detectability. *Science*, *199*(4324), 88–90. [10.1126/science.199.4324.88](https://doi.org/10.1126/science.199.4324.88)
- Killeen, P. R., & Jacobs, K. W. (2017). Coal is not black, snow is not white, food is not a reinforcer: The roles of affordances and dispositions in the analysis of behavior. *The Behavior Analyst*, *40*(1), 17–38. [10.1007/s40614-016-0080-7](https://doi.org/10.1007/s40614-016-0080-7)
- Killeen, P. R., & Smith, J. P. (1984). Perception of contingency in conditioning: Scalar timing, response bias, and erasure of memory by reinforcement. *Journal of Experimental Psychology: Animal Behavior Processes*, *10*(3), 333–345. [10.1037/0097-7403.10.3.333](https://doi.org/10.1037/0097-7403.10.3.333)
- Krägeloh, C. U., Davison, M., & Elliffe, D. M. (2005). Local preference in concurrent schedules: The effects of reinforcer sequences. *Journal of the Experimental Analysis of Behavior*, *84*(1), 37–64. [10.1901/jeab.2005.114-04](https://doi.org/10.1901/jeab.2005.114-04)
- Kubaneck, J., Snyder, L. H., & Abrams, R. A. (2015). Reward and punishment act as distinct factors in guiding behavior. *Cognition*, *139*, 154–167. [10.1016/j.cognition.2015.03.005](https://doi.org/10.1016/j.cognition.2015.03.005)
- Kuroda, T., Cançado, C. R. X., & Podlesnik, C. A. (2018). Relative effects of reinforcement and punishment on human choice. *European Journal of Behavior Analysis*, *19*(1), 125–148. [10.1080/15021149.2018.1465754](https://doi.org/10.1080/15021149.2018.1465754)
- Lachter, G. D. (1971). Some temporal parameters of non-contingent reinforcement. *Journal of the Experimental Analysis of Behavior*, *16*(2), 207–217. [10.1901/jeab.1971.16-207](https://doi.org/10.1901/jeab.1971.16-207)
- Lattal, K. A. (1975). Reinforcement contingencies as discriminative stimuli. *Journal of the Experimental Analysis of Behavior*, *23*(2), 241–246. [10.1901/jeab.1975.23-241](https://doi.org/10.1901/jeab.1975.23-241)
- Lattal, K. A. (1979). Reinforcement contingencies as discriminative stimuli: Effects of changes in stimulus probability. *Journal of the Experimental Analysis of Behavior*, *31*(1), 15–22. [10.1901/jeab.1979.31-15](https://doi.org/10.1901/jeab.1979.31-15)
- Lie, C. (2010). Human conditional-discrimination performance: Local reinforcer and punisher effects. *Experimental Analysis of Human Behavior Bulletin*, *28*, 1–6.
- Lie, C., & Alsop, B. (2009). Effects of point-loss punishers on human signal-detection performance. *Journal of the Experimental Analysis of Behavior*, *92*(1), 17–39. [10.1901/jeab.2009.92-17](https://doi.org/10.1901/jeab.2009.92-17)
- Lind, J., Enquist, M., & Ghirlanda, S. (2015). Animal memory: A review of delayed matching-to-sample data. *Behavioural Processes*, *117*, 52–58. [10.1016/j.beproc.2014.11.019](https://doi.org/10.1016/j.beproc.2014.11.019)
- Matute, H. (1995). Human reactions to uncontrollable outcomes: Further evidence for superstitions rather than helplessness. *The Quarterly Journal of Experimental Psychology*, *48*(2), 142–157. [10.1080/14640749508401444](https://doi.org/10.1080/14640749508401444)
- Matute, H. (1996). Illusion of control: Detecting response-outcome independence in analytic but not naturalistic conditions. *Psychological Science*, *7*(5), 289–293. [10.1111/j.1467-9280.1996.tb00376.x](https://doi.org/10.1111/j.1467-9280.1996.tb00376.x)
- Mellon, R. C. (2009). Superstitious perception: Response-independent reinforcement and punishment as determinants of recurring

- eccentric interpretations. *Behaviour Research and Therapy*, 47(10), 868–875. [10.1016/j.brat.2009.06.016](https://doi.org/10.1016/j.brat.2009.06.016)
- McCarthy, D., & Davison, M. (1979). Signal probability, reinforcement and signal detection. *Journal of the Experimental Analysis of Behavior*, 32(3), 373–386. [10.1901/jeab.1979.32-373](https://doi.org/10.1901/jeab.1979.32-373)
- McCarthy, D., & Davison, M. (1980). Independence of sensitivity to relative reinforcement rate and discriminability in signal detection. *Journal of the Experimental Analysis of Behavior*, 34(3), 273–284. [10.1901/jeab.1980.34-273](https://doi.org/10.1901/jeab.1980.34-273)
- Morse, W. H., & Skinner, B. F. (1957). A second type of superstition in the pigeon. *The American Journal of Psychology*, 70(2), 308–311. [10.2307/1419345](https://doi.org/10.2307/1419345)
- Neuringer, A. J. (1970). Superstitious key pecking after three peck-produced reinforcements. *Journal of the Experimental Analysis of Behavior*, 13(2), 127–134. [10.1901/jeab.1970.13-127](https://doi.org/10.1901/jeab.1970.13-127)
- Nussey, V. P., & Lattal, K. A. (1983). Stimulus control of responding by response-reinforcer temporal contiguity. *Learning and Motivation*, 14(4), 472–486. [10.1016/0023-9690\(83\)90028-0](https://doi.org/10.1016/0023-9690(83)90028-0)
- Ono, K. (1987). Superstitious behavior in humans. *Journal of the Experimental Analysis of Behavior*, 47(3), 261–271. [10.1901/jeab.1987.47-261](https://doi.org/10.1901/jeab.1987.47-261)
- Perone, M. (1991). Experimental design in the analysis of operant behavior. In I. H. Iversen & K. A. Lattal (Eds.), *Experimental analysis of behavior* (pp. 135–168). Elsevier.
- Pietras, C. J., Brandt, A. E., & Searcy, G. D. (2010). Human responding on random-interval schedules of response-cost punishment: The role of reduced reinforcement density. *Journal of the Experimental Analysis of Behavior*, 93(1), 5–26. [10.1901/jeab.2010.93-5](https://doi.org/10.1901/jeab.2010.93-5)
- Powell, R. W., & Kelly, W. A. (1979). Crows learn not to respond under response-independent reinforcement. *Bulletin of the Psychonomic Society*, 13(6), 397–400. [10.3758/BF03336906](https://doi.org/10.3758/BF03336906)
- Rachlin, H., & Baum, W. M. (1972). Effects of alternative reinforcement: Does the source matter? *Journal of the Experimental Analysis of Behavior*, 18(2), 231–241. [10.1901/jeab.1972.18-231](https://doi.org/10.1901/jeab.1972.18-231)
- Rasmussen, E. B., & Newland, M. C. (2008). Asymmetry of reinforcement and punishment in human choice. *Journal of the Experimental Analysis of Behavior*, 89(2), 157–167. [10.1901/jeab.2008.89-157](https://doi.org/10.1901/jeab.2008.89-157)
- Rawson, R. A., & Leitenberg, G. (1973). Reinforced alternative behavior during punishment and extinction with rats. *Journal of Comparative and Physiological Psychology*, 85(3), 593–600. [10.1037/h0035312](https://doi.org/10.1037/h0035312)
- Rey, C. N., Betz, A. M., Sleiman, A. A., Kuroda, T., & Podlesnik, C. A. (2020). The role of adventitious reinforcement during differential reinforcement of other behavior: A systematic replication. *Journal of Applied Behavior Analysis*, 53(4), 2440–2449. [10.1002/jaba.678](https://doi.org/10.1002/jaba.678)
- Rudski, J. M. (2000). Effect of delay of reinforcement on superstitious inferences. *Perceptual and Motor Skills*, 90(3), 1047–1058. [10.2466/pms.2000.90.3.1047](https://doi.org/10.2466/pms.2000.90.3.1047)
- Rudski, J. M. (2001). Competition, superstition and the illusion of control. *Current Psychology*, 20, 68–84. [10.1007/s12144-001-1004-5](https://doi.org/10.1007/s12144-001-1004-5)
- Rudski, J. M., Lischner, M. I., & Albert, L. M. (1999). Superstitious rule generation is affected by probability and type of outcome. *The Psychological Record*, 49(2), 245–260. [10.1007/BF03395319](https://doi.org/10.1007/BF03395319)
- Schwartz, B. (1973). Maintenance of key pecking by response-independent food presentation: The role of the modality of the signal for food. *Journal of the Experimental Analysis of Behavior*, 20(1), 17–22. [10.1901/jeab.1973.20-17](https://doi.org/10.1901/jeab.1973.20-17)
- Scotti, J. R., Evans, I. M., Meyer, L. H., & Walker, P. (1991). A meta-analysis of intervention research with problem behavior: Treatment validity and standards of practice. *American Journal on Mental Retardation*, 96(3), 233–256.
- Segal, E. F. (1972). Induction and the provenance of operants. In R. M. Gilbert & J. R. Millenson (Eds.), *Reinforcement: Behavioral analysis* (pp. 1–34). Academic Press.
- Shahan, T. A. (2017). Moving beyond reinforcement and response strength. *The Behavior Analyst*, 40, 107–121. [10.1007/s40614-017-0092-y](https://doi.org/10.1007/s40614-017-0092-y)
- Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgement of causality by human subjects. *The Quarterly Journal of Experimental Psychology Section B*, 41(2b), 139–159. [10.1080/14640748908401189](https://doi.org/10.1080/14640748908401189)
- Sheehan, K. J., Van Reet, J., & Bloom, C. M. (2012). Measuring preschoolers' superstitious tendencies. *Behavioural Processes*, 91(2), 172–176. [10.1016/j.beproc.2012.07.005](https://doi.org/10.1016/j.beproc.2012.07.005)
- Skinner, B. F. (1938). *The behavior of organisms*. Appleton-Century.
- Skinner, B. F. (1948). 'Superstition' in the pigeon. *Journal of Experimental Psychology*, 38(2), 168–172. [10.1037/h0055873](https://doi.org/10.1037/h0055873)
- Staddon, J. E. R. (1977). Schedule-induced behavior. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 125–152). Routledge.
- Staddon, J. E., & Simmelhag, V. L. (1971). The "superstition" experiment: A reexamination of its implications for the principles of adaptive behavior. *Psychological Review*, 78(1), 3–43. [10.1037/h0030305](https://doi.org/10.1037/h0030305)
- Stubbs, D. A. (1976). Response bias and the discrimination of stimulus duration. *Journal of the Experimental Analysis of Behavior*, 25(2), 243–250. [10.1901/jeab.1976.25-243](https://doi.org/10.1901/jeab.1976.25-243)
- Stubbs, D. A., & Pliskoff, S. S. (1969). Concurrent responding with fixed relative rate of reinforcement. *Journal of the Experimental Analysis of Behavior*, 12(6), 887–895. [10.1901/jeab.1969.12-887](https://doi.org/10.1901/jeab.1969.12-887)
- Timberlake, W., & Lucas, G. A. (1985). The basis of superstitious behavior: Chance contingency, stimulus substitution, or appetitive behavior? *Journal of the Experimental Analysis of Behavior*, 44(3), 279–299. [10.1901/jeab.1985.44-279](https://doi.org/10.1901/jeab.1985.44-279)
- Urciuoli, P. J. (1984). Overshadowing in matching-to-sample: Reduction in sample-stimulus control by differential sample behaviors. *Animal Learning & Behavior*, 12, 256–264. [10.3758/BF03199966](https://doi.org/10.3758/BF03199966)
- Urciuoli, P. J. (1985). On the role of differential sample behaviors in matching-to-sample. *Journal of Experimental Psychology: Animal Behavior Processes*, 11(4), 502–519. [10.1037/0097-7403.11.4.502](https://doi.org/10.1037/0097-7403.11.4.502)
- Urciuoli, P. J., & Honig, W. K. (1980). Control of choice in conditional discriminations by sample-specific behaviors. *Journal of Experimental Psychology: Animal Behavior Processes*, 6(3), 251–277. [10.1037/0097-7403.6.3.251](https://doi.org/10.1037/0097-7403.6.3.251)
- Wagner, G. A., & Morris, E. K. (1987). "Superstitious" behavior in children. *The Psychological Record*, 37(4), 471–488. [10.1007/BF03394994](https://doi.org/10.1007/BF03394994)
- 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. *Learning and Motivation*, 14(4), 406–432. [10.1016/0023-9690\(83\)90025-5](https://doi.org/10.1016/0023-9690(83)90025-5)
- Wasserman, E. A., & Neunaber, D. J. (1986). College students' responding to and rating of contingency relations: The role of temporal contiguity. *Journal of the Experimental Analysis of Behavior*, 46(1), 15–35. [10.1901/jeab.1986.46-15](https://doi.org/10.1901/jeab.1986.46-15)
- Weaver, J. E., Dorrance, B. R., & Zentall, T. R. (1999). Present/absent sample matching in pigeons: Is comparison choice controlled by the sample stimulus or by differential sample responding? *Animal Learning & Behavior*, 27(3), 288–294. [10.3758/BF03199727](https://doi.org/10.3758/BF03199727)
- Wilkie, D. M., Jacobs, W. J., & Takai, R. (1985). Pigeons match location of sample more accurately than color of sample. *Bulletin of the Psychonomic Society*, 23(2), 156–159. [10.3758/BF03329811](https://doi.org/10.3758/BF03329811)
- Zeiler, M. D. (1968). Fixed and variable schedules of response-independent reinforcement. *Journal of the Experimental Analysis of Behavior*, 11(4), 405–414. [10.1901/jeab.1968.11-405](https://doi.org/10.1901/jeab.1968.11-405)

**How to cite this article:** Gomes-Ng, S., Cowie, S., & Elliffe, D. (2023). Is superstitious responding a matter of detectability? A replication of Killeen (1978). *Journal of the Experimental Analysis of Behavior*, 1–15. <https://doi.org/10.1002/jeab.855>

## APPENDIX A



**FIGURE A1** Performance in the last 15 sessions of each condition for individual pigeons. Performance was calculated using Equation 1 ( $\log d$ )

## APPENDIX B

TABLE B1 Correct and incorrect responses in each trial type in the last 10 sessions of each condition

Pigeon	Response-dependent		Response-independent		Response-dependent		Response-independent	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
	Condition 1 (stimulus off)				Condition 2 (food)			
61	259	100	288	66	153	194	273	71
62	206	186	282	108	141	237	242	132
63	249	128	328	57	127	251	233	140
64	372	14	269	121	209	174	221	158
65	365	29	260	138	222	155	141	221
66	262	131	365	33	240	156	231	168
	Condition 3 (blackout)				Condition 4 (food correction)			
61	203	49	186	68	62	153	188	34
62	217	141	284	91	74	185	154	81
63	264	104	285	77	64	198	194	53
64	293	60	228	134	243	50	178	109
65	273	115	200	183	164	76	69	172
66	164	211	226	158	182	114	190	91
	Condition 5 (food forced)				Condition 6 (blackout forced)			
61	202	115	228	204	250	59	207	91
62	150	154	178	128	242	133	285	99
63	163	207	279	99	302	69	272	88
64	272	85	268	92	358	18	281	94
65	119	89	112	108	245	131	203	174
66	334	55	240	153	233	138	164	202