

# Transfer of judgments of control to a target stimulus and to novel stimuli through derived relations

Charlotte Dack · Louise McHugh · Phil Reed

Published online: 17 February 2012  
© Psychonomic Society, Inc. 2012

**Abstract** Three experiments examined the effect of response–outcome contingencies on human ratings of causal efficacy and demonstrated that such ratings transfer to novel situations through derived stimulus relations. Efficacy ratings generally followed the delta probability rule when positive response–outcome contingencies were employed (**Experiment 1**) and when some outcomes were not contingent on participants’ responses (**Experiment 2**). **Experiment 3** employed a negative response–outcome contingency and manipulated performance expectancies in the task. All three groups overestimated their causal efficacy ratings. A learned helplessness effect was observed when the response–outcomes were uncontrollable and in the high-expectancy group when participants’ performance in the task was worse than they had expected. In all experiments, ratings transferred to a stimulus presented during the task and often generalized to novel stimuli through derived relations. These results corroborate the view that outcome probability is a determinant of causal efficacy ratings and that schedules can be employed as UCS in procedures that share characteristics of evaluative conditioning procedures.

**Keywords** Schedule of reinforcement · Equivalence classes · Human learning

The processes underlying how people learn to like or dislike certain foods (e.g., Stuart, Shimp, & Engle, 1987), people (Walther, 2002), and events (Dack, McHugh, & Reed, 2009) are important to understand, since preferences play a

significant role in all aspects of life. Some preferences may be genetically determined (Breen, Plomin, & Wardle, 2006; Ellis & Bonin, 2003; cf. Rozin & Millman, 1987); however, most research suggests that likes and dislikes are acquired through learning mechanisms (Barnes & Roche, 1997; Field, 2006; Hermans et al., 2005; Johnsrude, Owen, Zhao, & White, 1999). One approach to understanding the processes behind the development of preferences suggests that their acquisition emerges through associative learning, or conditioning, whereby a novel stimulus (equivalent to a conditioned stimulus [CS]), acquires a valence through its association with a stimulus with an existing preference (equivalent to the unconditioned stimulus [US]). This transfer of valence (i.e., the degree to which an individual likes or dislikes a stimulus) through an association is commonly referred to as *evaluative conditioning* (Levey & Martin, 1975; see De Houwer, Thomas, & Baeyens, 2001, for a review of evaluative conditioning).

Evaluative conditioning occurs when the valence of a novel “target” stimulus becomes altered by the valence of stimuli/events that are presented with or are contingent on the target stimulus. This phenomenon is often shown using the picture–picture paradigm. In a typical experiment, a picture that was previously rated as neutral is repeatedly paired with a picture that had been rated as pleasant (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992). As a consequence, the valence of the previously neutral picture changes in the direction of the picture with which it was paired (e.g., its ratings will change in the direction of pleasantness). Evaluative conditioning may also play a role in the development of several clinical disorders (Hermans, 1998). For example, depression and anxiety disorders are often characterized by a person’s negative evaluations about themselves, their environment, and their future (Abramson, Metalsky, & Alloy, 1989; Beck, 1967; Beck, Rush, Shaw, &

---

C. Dack · L. McHugh · P. Reed (✉)  
Department of Psychology, Swansea University,  
Singleton Park,  
Swansea SA2 8PP, UK  
e-mail: p.reed@swansea.ac.uk

Emery, 1979; Clark, Beck, & Alford, 1999; Dohr, Rush, & Bernstein, 1989; Loeb, Feshbach, Beck, & Wolf, 1964; MacLeod & Byrne, 1996; Roth & Rehm, 1980; Seligman, 1975). These evaluations, such as causal effectiveness and self efficacy, may be acquired through evaluative conditioning processes (see Fulcher, Mathews, Mackintosh, & Law, 2001; Walther, Nagengast, & Trasselli, 2005). Given the potentially important applications of this line of enquiry, it is essential that the mechanisms behind the production of negative and positive evaluations are investigated and understood.

Dack et al. (2009) demonstrated that participants' self-evaluations (i.e., good vs. bad) of their performance during two tasks transferred to stimuli that were presented during the performance of the tasks. Specifically, previously neutral target stimuli were paired with schedules of reinforcement—one schedule that produced high ratings of causal efficacy, and another that produced low ratings of causal efficacy—and the target stimuli were later rated as “good” or “bad,” respectively. That is, participants rated a neutral stimulus that had been paired with a task in which they judged themselves effective as good and a neutral stimulus that had been paired with a task in which they judged themselves ineffective as bad.

However, a feature of many clinical disorders (Morgan & Banerjee, 2006; Walther et al., 2005) is the extent to which such evaluations generalize across many aspects of a person's life—for example, depressive attributional style and generalized anxiety. Potentially, these negative expectations could lead to maladaptive behaviors, such as social withdrawal, avoiding particular activities, and functional impairments (Simon, 2003). It is unlikely that similar negative evaluative judgments would be directly learned across such a wide range of circumstances, and the question remains regarding how such evaluations are generalized across so many aspects of a person's life. Evidence for the indirect effect of learning was shown in a set of studies by Rozin, Markwith, and McCauley (1994). They demonstrated that an aversion toward a person with AIDS could be generalized to objects associated with this person (sweater, car, bed, etc.), even though participants were reassured about the low risk of contracting AIDS from these objects.

Similarly, Walther (2002) showed that an attitude (positive or negative) toward one stimulus can transfer to stimuli that had been preassociated with that stimulus. In the baseline phase of this study, participants were asked to rate a number of pictures of male faces for their likeability. In the preconditioning phase, two neutral faces were paired together. In the conditioning phase, one of the neutral faces (the target) was paired with a positively evaluated picture (Experiment 1) or a negatively evaluated picture (Experiment 2). In the test phase, participants were asked to rate all the visual stimuli again for their likeability. The results showed

not only that ratings of the target stimulus increased (Experiment 1) or decreased (Experiment 2) in likeability, but also that this attitude transferred to the neutral face that was preassociated with the target stimulus. The *spreading attitude* effect was also replicated using a second-order conditioning paradigm (Walther, 2002, Experiment 4), where the preconditioning and conditioning phases were reversed.

An additional explanation of the development of generalized evaluations and attitudes is produced through examining derived stimulus relations in equivalence classes, an account that suggests that stimuli can become associated with one another despite never having been directly paired and despite their lack of any shared physical properties (Sidman, 1971). The potential breadth of such classes could allow for widespread transfer of particular associations learned about one member of that class.

Equivalence classes can be established by training a minimal number of relations between individual stimuli in a group. For example, if a group of stimuli consisted of the letters A, B, and C, an equivalence class could be established by training two-term relations between AB and BC. If a class has been established, many new emergent relations that had not been taught directly are formed between the stimuli: reflexive (i.e.,  $A \rightarrow A$ ), symmetrical (i.e.,  $B \rightarrow A$ ), transitive (i.e.,  $A \rightarrow C$ ), and equivalence (i.e.  $C \rightarrow A$ ) (see Bush, Sidman, & de Rose, 1989). If all of the emergent relations control responding, the group of stimuli can be said to function as an equivalence class (Sidman, Kirk, & Wilson-Morris, 1985), and the stimuli are fully substitutable for one another (Sidman, 1990, 1994).

Importantly, similar to Walther's (2002) spreading attitude effect, after an equivalence class is established, training a novel function to one member of the equivalence class allows the same function to transfer to the other members of the equivalence class without further training (e.g., if A is paired with a sour taste, or negative attribution, the C [and B] stimulus will also acquire sour properties, or negative attributions). This phenomenon is referred to as the *transfer*, or *transformation*, of function and has been demonstrated with a number of derived stimulus relations and behavioral functions, such as avoidance responses, preferences, self discrimination, and moods (see Dymond & Rehfeldt, 2000, for a review).

Furthermore, Dack et al. (2009) showed that stimuli that had been associated with either high or low ratings of causal efficacy not only were rated as “good” or “useless” themselves, but also were categorized with stimuli participating in equivalence classes with the same evaluative function (“good” or “useless,” respectively). However, although Dack et al. demonstrated that new stimuli can join an existing equivalence class by virtue of sharing a common evaluative function with the members of that equivalence class, they did not show that directly training an evaluative function to a class member would lead to a transfer of that

evaluative function to other members of that class. Given the potential of equivalence classes to explain widespread transfer of functions (see Dymond & Rehfeldt, 2000), it would be useful to determine whether a target stimulus, used during an evaluative learning task, would produce transfer of that function to other members of an existing equivalence class that had not been directly conditioned to that evaluative function. In the present study, three experiments were designed to demonstrate that ratings of causal efficacy, produced by exposure to various judgment-of-contingency tasks, can transfer to stimuli present during these tasks and to unrelated stimuli through derived stimulus relations in existing equivalence classes.

The experiments reported here extend the earlier research reported by Dack et al. (2009) not only by exploring the critical issue of transfer of function through participation in equivalence classes, but also by using simpler probability judgment tasks to establish initial causal efficacy (e.g., Alloy & Abramson, 1979; Alloy & Tabachnik, 1984; Reed, 1999; Wasserman, Elek, Chatlosh, & Baker, 1993), instead of the previously employed schedules of reinforcement (Dack et al., 2009; Reed 2001a, b). Judgment-of-contingency tasks allow easier control of the many factors that can vary in schedules of reinforcement (e.g., rate of responses, rate of reinforcement, outcome probability) and that make production and interpretation of causal efficacy ratings difficult. Typically, in judgment-of-contingency tasks, participants are asked to evaluate the efficacy of their responses (e.g., keypressing) in producing a chosen outcome (e.g., gaining a point on the screen). The contingency between a response and an outcome can be defined probabilistically using metrics like the delta probability (dP) rule (Allan, 1980). This rule defines the contingency as the difference between the probability of an outcome occurring when a response is made and the probability of an outcome occurring when no response is made. Positive differences between these values indicate that outcomes are contingent upon responding, whereas negative differences indicate that outcomes are contingent upon not responding. A difference of zero indicates that responding is not contingently related to the outcome. In addition to promoting basic understanding of evaluative learning and its transfer across situations, these experiments will explore the effects in a way that is potentially important to applied contexts.

## Experiment 1

Many studies have found that participants accurately evaluate response–outcome contingencies (Chatlosh, Neunaber, & Wasserman, 1985; Neunaber & Wasserman, 1986; Wasserman, 1990), and individual participant’s ratings have been found to be highly correlated with the dP rule (e.g., Chatlosh et al., 1985; Wasserman, Chatlosh, & Neunaber, 1983). However, the majority of these studies have measured

such effects within subjects. An aim of the first experiment was to establish that accurate ratings of causal efficacy for different response–outcome contingencies could be produced using a between-subjects design and that these ratings could transfer to a stimulus (the target) that was presented during the judgment-of-contingency task.

Additionally, to investigate whether causal efficacy ratings can transfer to stimuli that have been previously related through both directly trained and derived relations to the target stimulus, but not directly exposed to the contingency, participants were taught 2 three-member equivalence classes (A1, B1, C1; and A2, B2, C2) before exposure to the judgment-of-contingency task. The class member C1 was then used as the target in the judgment-of-contingency task, and any transfer to the related stimuli (A1 and B1) was measured in a function acquisition test. In this function acquisition test, participants had to categorize their performance (“good,” “bad”) in the presence of each stimulus.

## Method

### Participants

Thirty-one adults participated in the current study (14 male, 17 female). The ages of the participants ranged from 20 to 55 years, with a mean of 26.0 ( $SD = 7.2$ ) years of age. All participants were recruited by advertisements in the psychology department and earned subject pool credit. Each participant had normal or corrected-to-normal eyesight. All were unaware about the purpose of the experiment.

### Apparatus and materials

The experiment was conducted in a quiet room free from distraction. The room contained only a desk, a chair, and a personal computer, with a 550-MHz processor, a 14-in. color monitor, and a standard computer mouse. All session presentations and response recordings were controlled by programs written in Visual Basic 6.0. All responses involved either mouse clicking or pressing the space bar key.

*Derived stimulus relations training* The stimuli used as members of the two equivalence classes were nonsense words and can be seen in Table 1, along with their

**Table 1** Nonsense words used as stimuli and their assignment to equivalence classes

	A	B	C
Class 1	Matser	Rigund	Boceem
Class 2	Lewoly	Gedeer	Casors

corresponding letter–number designation (A1, B1, C1 etc.). All stimuli were composed of Arial Bold characters in black, each of which occupied a certain proportion of the screen (screen width/200 twips). Each stimulus was surrounded by a box, its width equalling the screen width/2.5 twips, and its height equalling the screen height/6 twips, against a white background.

*Judgment-of-contingency task* Participants pressed the space bar of the computer freely during the course of the task. Each judgment-of-contingency session lasted 3 min. Participants received a point (the outcome) for pressing the space bar (response) with an outcome probability of .25, .50 or .75, depending on the dP rule assigned to them. Points accumulated were represented in a box in the corner of the computer screen. Every time a participant received a point, the box flashed red. At the end of this training, participants were asked to rate how effective they were at earning points by using the mouse to slide a pointer on an (onscreen) 10-point (1=*ineffective*, 10=*effective*) scale to the chosen number. The class member C1 (“BOCEEM”) was used as the target stimulus and was presented in the center of the computer screen for the duration of the task.

*Scales* The Beck’s Depression Inventory (BDI; Beck, 1978; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) was administered to assess the level of depression in participants. This is a self-report measure, composed of 21 statements concerning how participants felt over the last week. An overall BDI score over 10 indicates depressive responding. Only data from participants with BDI scores of less than 10 were employed in this study. This was to ensure that it was the manipulation of the response–outcome contingencies, and not any preexisting depressive condition, that was responsible for the effects seen in these experiments (see Reed & Antonova, 2007; Reed, Frasquillo, Colkin, Liemann, & Colbert, 2001).

## Procedure

Each participant was taken into a quiet room and given a consent form to read and sign before completing the study. The participants were allocated to one of three groups (dP = .25,  $n = 9$ ; .50,  $n = 12$ ; or .75,  $n = 10$ ) that differed in terms of the outcome probabilities they experienced during the judgment-of-contingency task. No outcomes were ever delivered in the absence of a response. All participants experienced the three computer tasks in the order described below.

*Derived stimulus relations training* Participants were exposed to a respondent training procedure followed by a match to sample (MTS) equivalence test (see Leader & Barnes-Holmes, 2001). At the start of the respondent

training, participants were presented with the following instructions:

Please just look at the nonsense words as they appear below. Do not click on the nonsense words with the mouse. Click Here To Continue.

During respondent training, six nonsense syllables (see Table 1) were presented to the participants in the form of four stimulus pairs (A1 → B1, B1 → C1, A2 → B2, B2 → C2). The first stimulus of each pair was presented for 1 s, and the computer screen was then cleared for 0.5 s (the within-pair delay). The second stimulus was then presented for 1 s, and the screen was then cleared for 3 s (the between-pair delay). Following the between-pair delay, the next stimulus pair was presented. Stimuli were always presented in the center of the computer screen. The stimulus pairs were presented 4 times each, in a quasirandom order, producing 16 trials.

When all 16 trials had been presented, the instructions to test for emergent relations appeared on the screen:

Look at the box above and then click on the box below that GOES WITH the one at the top. Try your best NOT to make any mistakes.

All tests for symmetry (B → A, C → B), one-node transitivity (A → C), and equivalence (C → A) were presented and tested in a single block, using the MTS procedure. Each type of relation (B1–A1, B2–A2, C1–B1, C2–B2, A1–C1, A2–C2, C1–A1, and C2–A2) was presented 4 times, in a quasirandom order, with 32 trials in total. For relations of symmetry, either a B (1 or 2) stimulus or a C (1 or 2) stimulus functioned as samples, and the A and B stimuli (1 or 2) functioned as comparison, respectively. For relations of transitivity, an A (1 or 2) stimulus functioned as the sample, and the C stimuli as comparisons. For equivalence relations, the C stimulus (1 or 2) functioned as the sample, and the A stimuli as comparisons. On each trial, the sample stimulus was always presented in the center of the top half of the screen. The two comparison stimuli were presented in each corner of the bottom half of the screen. When a comparison had been selected, the screen cleared and remained blank for 2 s, and then the next test trial was presented. No feedback was presented to the participants during testing. The mastery criterion for testing was at least 28 class-consistent selections for all derived relations (symmetry, transitivity, and equivalence) across the 32 test trials. Participants were recycled through the training and testing until they reached this criterion. If after six exposures to the equivalence training participants did not reach this criterion, they were excluded from the study.

*Judgment-of-contingency task* At the start of this phase, the participants were presented with the following instructions on the computer screen:

Please read the following instructions very carefully. Take as much time as you like. Your task in this experiment is to find out whether pressing the space bar has any effect on whether you receive a point. At any time you may choose whether or not to press the space bar. You can press it as often or as little as you like. Sometimes you will receive a point when you press the space bar. WE WILL START WITH A PRACTICE RUN TO GIVE YOU AN IDEA OF WHAT THE TASK IS LIKE. Click Here To Proceed.

These instructions were followed by a 20-s practice session, in which participants received points for pressing the space bar in accordance with the response–outcome contingency group to which they were assigned. At the end of the practice session, participants were presented with the following instructions.

THAT WAS A PRACTICE SESSION. NOW YOU WILL DO THE EXPERIMENTAL TASK ITSELF. CLICK TO PROCEED. Please read the following instructions very carefully. Take as much time as you like. Your task in this experiment is to find out whether pressing the space bar has any effect on whether you receive a point. At any time you may choose whether or not to press the space bar. You can press it as often or as little as you like. Sometimes you may receive a point when you press the space bar. You must rate how effective you think you are at getting points. You will be given several exposures to the problem, each lasting 3 minutes. Your task in this experiment is to earn as many points as you can. This may mean that you will need to press the space bar some of the time and not press it some of the time. At the end of the task you will be asked to give an estimate on a rating scale of how effective you thought you were at gaining points. YOU WILL SHORTLY SEE A WORD IN THE CENTER OF THE SCREEN, CLICK ON IT TO START THE TASK. THIS IS THE ACTUAL TEST NOT A PRACTICE. Click Here To Proceed.

The participants were then exposed to three sessions of the task, which each lasted for 3 min. The target stimulus (C1) was presented on the computer screen for the duration of the task. Participants received points during this session in accordance with the dP rule assigned for their response–outcome group.

The response–outcome probability was operationalized by a customized computer program. Every second, the program selected a number between 1 and 4, inclusive. If the response–outcome probability was 25%, the computer would generate an outcome (point) only if one particular number (between 1 and 4) was chosen, and if the

participants had responded (pressed the space bar) during the previous second. If the response–outcome probability was 50%, the program would generate an outcome only if one of two particular numbers were chosen and if the participants had responded during the previous second. If the response–outcome probability was 75%, the program would generate an outcome if one of three particular numbers was chosen and if the participants had responded during the previous second.

After the session had concluded, the participants were asked to make a judgment about the causal efficacy of their responses. The following instructions were presented on the screen directly above a 10-point Likert scale:

ON A SCALE OF ONE TO TEN, HOW EFFECTIVE DO YOU THINK YOUR SPACE BAR PRESSING WAS IN GAINING POINTS? Please slide the slider bar to make your choice and then press the CONFIRM CHOICE button underneath. CLICK TO PROCEED.

Participants were required to click on a pointer on the Likert scale with the mouse and slide the pointer to the number that represented how effective they thought they had been at getting points in the previous task (where 1 = *ineffective* and 10 = *effective*). When they had chosen a number, they had to click on a button underneath the Likert scale that had “confirm” written on it to confirm their choice and to proceed to the next task.

*Function acquisition test* The purpose of this phase was to determine the evaluation participants made for each of the stimuli in each equivalence class. To this end, the participants were presented with each of the three stimuli from the two classes generated in the first part of training (above). They were asked to evaluate how they thought that they would perform in the presence of these stimuli in another task by choosing one of the words (“good,” “average,” or “bad”) at the bottom of the screen. To this end, the following instructions on the computer screen:

Please click on the word that appears at the top of the screen, and drag it on top of one of the 3 words at the bottom of the screen. You should choose the word at the bottom of the screen which you feel best represents how you would perform on the previous space bar task if that word had appeared on the screen.

Each session began with the presentation of a stimulus from the equivalence classes, in the middle of the top of the screen, and the words GOOD, AVERAGE, and BAD at the bottom of the screen. Participants had to drag and drop the stimulus at the top of the screen onto whichever word (good, average, bad) that they thought reflected how they would perform when this stimulus was presented in the previous judgment-of-contingency task. Each stimulus was shown 4

times, in a random order, making 24 trials in total. This was to check whether participants' choices were stable over repeated trials. At no point during this test were participants given any feedback for their responses.

## Results and discussion

Two participants' data were excluded from the analysis since they did not reach the mastery criteria for the emergent relations testing (28/32) after six respondent training sessions. Five participants were excluded due to their showing some signs of preexisting depression (BDI scores > 10), following the procedure of Reed et al. (2001). The mean score for the remaining participants on the BDI was 3.8 ( $SD = 3.4$ ).

*Derived stimulus relations testing* All remaining participants passed the emergent relations test, with an overall mean of 98.96% ( $SD = 2.2$ ) class-consistent responding. Fifteen participants reached the criteria (28/32) on the first test session. Five participants needed a second training phase before reaching the test criterion, and 3 participants required a third. One participant required five respondent training phases before finally reaching the criteria.

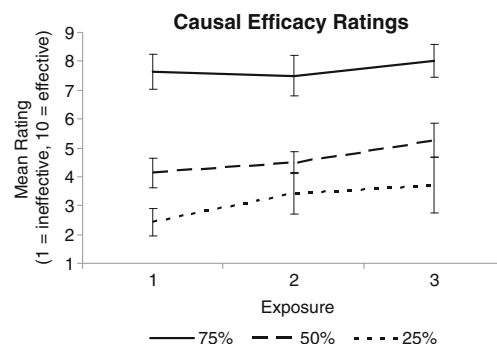
*Outcomes obtained* Participants earned a higher number of outcomes per minute in the 75% contingency group (mean number of outcomes per minute collapsed across sessions = 40.8,  $SD = 3.8$ ), followed by the 50% contingency group ( $21.6 \pm 5.3$ ) and then the 25% group (12.92.6). A mixed-model analysis of variance (ANOVA), with group and session as factors, was performed on these data. A rejection criterion of  $p < .05$  was adopted for these and all subsequent analyses. This ANOVA revealed significant main effects of group,  $F(2, 21) = 100.07$ ,  $\eta^2 = .91$ , and session,  $F(2, 42) = 4.10$ ,  $\eta^2 = .16$ , but no interaction between the two factors,  $F < 1$ . Tukey's honestly significant difference (HSD) tests were conducted on each group's number of outcomes averaged across all sessions. These showed the number of outcomes for the 75% contingency group were significantly higher than the number of outcomes obtained in the other two groups. The outcomes earned in the 50% contingency group were significantly higher than those for the 25% contingency group.

*Actual delta probabilities* To check that the computer program had delivered outcomes according to the response–outcome contingency that was assigned to each participant, the actual delta probabilities were calculated for each participant. These were computed for each participant by dividing the total number of outcomes per session by the total number of responses per session and averaging across the three sessions. The average response–outcome

probabilities for each group were .75 ( $SD = .01$ ) in the 75% contingency group, .48 ( $SD = .04$ ) in the 50% contingency group, and .24 ( $SD = .03$ ) in the 25% contingency group. This shows that the computer program was accurate in the 75% group but produced slightly lower response–outcome contingencies in the 50% and 25% groups.

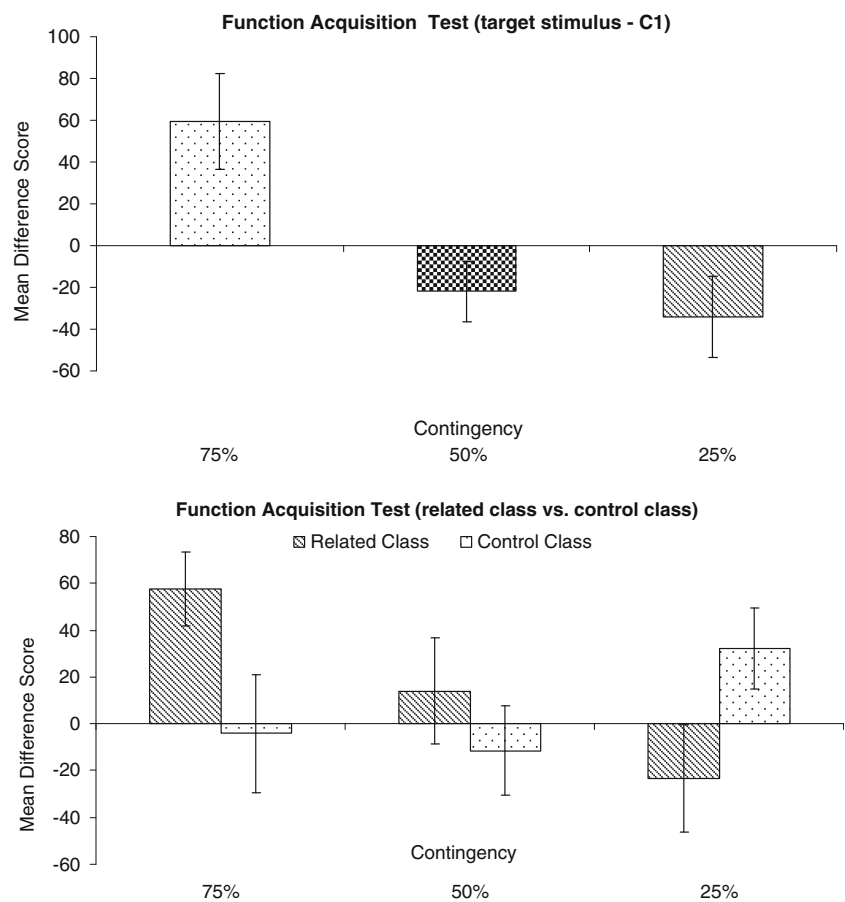
*Efficacy ratings* The mean ratings of causal efficacy emitted during each response–outcome contingency group across each exposure are shown in Fig. 1. Inspection of these data shows that ratings of causal efficacy were higher in the 75% group than in the 50% and 25% groups. A mixed-model ANOVA (group  $\times$  session) was performed on these data and revealed significant main effects of group,  $F(2, 21) = 18.83$ ,  $\eta^2 = .64$ , and session,  $F(2, 42) = 4.00$ ,  $\eta^2 = .16$ , but no interaction between the two factors,  $F < 1$ . Tukey's HSD tests conducted on the groups' ratings averaged across all sessions showed that efficacy ratings for the 75% contingency group were significantly higher than ratings obtained from the other two groups.

*Function acquisition test* To analyze the transfer of evaluation from the contingency to the target stimulus (C1), a difference score for the target was calculated for each contingency group: The percentage of “bad” categorization was subtracted from the percentage of “good” categorization. Positive scores reflect positive evaluations of the stimulus (more good than bad), and negative scores reflect a negative evaluation (more bad than good). An inspection of the top panel of Fig. 2 shows that, for participants in the 75% group, there was a positive (“good”) transfer to the target stimulus; that is, participants thought they would be “good” at earning points in the presence of that stimulus. In contrast, both the 50% and 25% groups displayed negative transfer (a tendency to choose “bad” more than “good”). An ANOVA revealed a significant main effect of group,  $F(2, 21) = 6.92$ ,  $\eta^2 = .63$ . Tukey's HSD tests showed the differences between the 75% contingency group and both other groups



**Fig. 1** Results from Experiment 1: Mean rating of causal efficacy for each response–outcome contingency group (75%, 50%, and 25%) across each session. Error bars represent mean standard error

**Fig. 2** Results from Experiment 1. Top panel: Mean difference score (i.e., the percentage “good”–the percentage “bad”) for the target stimulus in each contingency group (75%, 50%, 25%). Bottom panel: Mean difference scores for the related class and the control class. Error bars represent mean standard error



to be statistically significant. In addition, the difference scores for the stimulus for each group were tested against zero (reflecting a neutral valence). Only the 75% contingency group had a difference score that was significantly higher than zero,  $t(7) = 2.57$ ,  $\eta^2 = .70$  [50% group,  $t(7) = -1.51$ ,  $p > .05$ ,  $\eta^2 = .25$ ; 25% group,  $t(7) = -1.77$ ,  $p > .05$ ,  $\eta^2 = .56$ ], although the effect sizes for the other two groups were moderate.

The bottom panel of Fig. 2 shows the difference scores (the percentage good minus percentage bad) for the stimuli in the class related to the target stimulus (i.e., the scores for A1 and B1 were averaged to produce one score), and the stimuli in the control class (i.e., the scores of A2, B2, and C2 were averaged to produce one score). As the response–outcome contingency decreased, the percentage that stimuli related to the target were chosen with “good” also decreased, and the percentage that they were chosen with “bad” increased. This finding was not noted in the control class.

A mixed-model ANOVA (with class [related or unrelated] as the within-subjects factor and contingency group [75%, 50%, and 25%] as the between-subjects factor) was conducted on these data and revealed no significant main effects of group or class (both  $ps > .30$ ), but the interaction between group and class was significant,  $F(2, 21) = 3.32$ ,

$\eta^2 = .24$ . Simple effects analyses revealed a significant effect of group for the related class,  $F(2, 21) = 3.85$ ,  $\eta^2 = .52$ , but not for the control class,  $F(2, 21) = 1.26$ ,  $\eta^2 = .33$ . Tukey’s HSD comparisons between the difference scores of all the contingency groups for the related class revealed a significant difference only between the 75% and 25% contingency groups.

The difference scores for the related and control classes in each contingency group were tested against zero (reflecting a neutral valence). In the 75% group, the difference scores for the related class were significantly higher than zero,  $t(7) = 3.65$ ,  $\eta^2 = .81$ . For the control class, the difference scores did not differ significantly from zero,  $t < 1$ . In the 50% and 25% groups, there were no significant differences between the difference scores and zero for either of the equivalence classes, all  $ps > .1$ .

In addition to the analyses above, an ANOVA was performed on the difference score data for just the A stimuli, rather than the difference score data of the A and B stimuli combined and averaged. This was to check that a true *transfer of function* had occurred, since any evaluation that the A stimuli has gained is likely to be purely derived (whereas it could be argued that the B stimuli acquire a function through backward conditioning). For the A stimulus in an equivalence relation with the target C stimulus,

there was a significant effect of contingency group,  $F(2, 23) = 5.88$ ,  $\eta^2 = .60$ . This was not the case for the A stimulus from the unrelated class,  $F(2, 23) = 1.73$ ,  $p > .05$ ,  $\eta^2 = .38$ .

The findings of **Experiment 1** showed that differences in causal efficacy ratings can be produced from exposure to different response–outcome contingencies, using a between-subjects design. In the function acquisition test, participants in the 75% group had more positive difference ratings for the target stimuli than did the 50% and 25% groups; participants in the 75% group categorized their performance in the presence of the target as “good” more than did the other groups. The 25% response–outcome group had the largest negative difference scores for the target stimulus, suggesting that participants categorized their performance as “bad” in the presence of the target. However, there was no significant difference between the difference score and zero (neutral valence), which supports previous research (e.g., Hammerl & Grabitz, 1996) that has reported difficulties in showing a transfer of affect in aversive conditioning.

The function acquisition test also showed that the difference scores (“positive” vs. “negative”) for stimuli in the related class (A1 and B1) followed the same trends as the causal efficacy ratings associated with each response–outcome contingency and the difference scores for the directly conditioned target (C1). As the response–outcome contingency decreased, the difference scores for the related class also decreased. The 75% group had a significantly higher (more positive) difference score than the 25% group and from zero, suggesting that participants also predicted their performance at the task in the presence of the related class as “good” without direct experience of those tasks. However, in the 50% and 25% groups, the difference scores for stimuli in the related class were not significantly different from zero, replicating again the difficulties found in previous research (Hammerl & Grabitz, 1996) to produce a significant negative transfer.

Overall, this experiment provides evidence not only that self-evaluations about performance on a task can transfer to a stimulus present during the task, but also that they transfer to stimuli previously associated with the target through derived relational networks of equivalence. This effect is more prominent in the transfer of positive affect, as compared with negative affect.

## Experiment 2

Experiment 2 was designed to determine the impact of outcomes not contingent on participants responses on ratings of causal efficacy and whether these ratings would also transfer to stimuli that were not directly related to the task, through their participation in derived stimulus networks. The contingencies that were used in the judgment-of-control task

were 75/0 ( $dP = +75$ ), where outcomes were dependent on responding; 75/75 ( $dP = .00$ ), where outcomes were dependent on both responding and not responding; and 0/75 ( $dP = -75$ ), where outcomes were dependent on not responding.

These contingencies have been studied in the context of learned helplessness, which is often used as a model of depression. The theory of human learned helplessness (Seligman, 1975) grew out of research (Overmier & Seligman, 1967; Seligman & Maier, 1967) on nonhumans’ responses to uncontrollable electric shocks and showed that these participants often developed behaviors that resembled depressive symptoms when later exposed to controllable shocks (e.g., cowering passively and whimpering) and failed to learn how to escape. From these early findings, it was proposed that depression arises from a perception that important environmental events cannot be controlled.

Given the potential importance of the present research to clinical contexts, extension to a learned helplessness effect would seem particularly relevant. Previous research (e.g., Maldonado, Martos, & Ramirez, 1991; Reed et al., 2001) has noted that exposure to an uncontrollable relationship between a response and its outcome would interfere with subsequent learning. Thus, a learned helplessness effect was expected in the case of participants exposed to the 75/75 contingency group, where the relationship between responses and outcomes may be perceived as uncontrollable. This finding, in itself, would be important for our understanding of how evaluations about the self in contexts that are uncontrollable might affect subsequent behavior.

## Method

### Participants

Thirty-seven adults participated in the present study (13 male, 24 female). The ages of the participants ranged from 19 to 59 years, with a mean of 27.1 ( $SD = 9.6$ ) years of age. All participants were recruited by advertisements in the psychology department, and they earned participant pool credit or £6. Each participant had normal or corrected-to-normal eyesight. All were naïve about the purpose of the experiment.

### Apparatus and materials

The apparatus and materials were identical to those employed in **Experiment 1**, except for the introduction of a maze-learning task. Participants were required to complete a pencil-and-paper maze task straight after completing the judgment-of-contingency task. They were given a photocopy of the maze displayed in Fig. 3. The maze had a width of 23 cm and a height of 9.5 cm. They were asked to draw a





**Fig. 3** Schematic representation of the maze used in Experiment 1 and 3

path from the entrance of the maze to its center. The time taken to complete the maze was recorded by a hand-held stopwatch.

#### Procedure

Each participant was taken into a quiet room and were given a consent form to read and sign and the Beck's Depression Inventory to answer, before completing the study. The experimenter randomly allocated the participant to one of three contingency groups:  $dP = .75$  ( $n = 13$ ),  $dP = .00$  ( $n = 12$ ), and  $dP = -.75$  ( $n = 12$ ). Each participant received each task in the order that is described below.

*Derived stimulus relations training* Derived stimulus relations training was identical to that in Experiment 1.

*Judgment-of-contingency task* The judgment-of-contingency task was identical to that in Experiment 1, except for the contingency groups (see above). The outcome probability for pressing the space bar (response) was programmed as described in Experiment 1. The outcome probability for not pressing the space bar was manipulated as follows. If the outcome probability was 75%, the computer would generate an outcome if one of three particular numbers (between 1 and 4) was chosen and if the participant had not responded during the previous second.

Group 75 ( $dP = .75$ ) had a high response–outcome contingency (75%), and the no response–outcome contingency was set to zero (0%). Group 0 ( $dP = .00$ ) had a high response–outcome contingency (75%), and a high no response–outcome contingency (75%). Group  $-75$ 's ( $dP = -.75$ ) response–outcome contingency was set to zero (0%), and the no response–outcome contingency was high (75%).

After each 2-min trial had concluded, the participants were asked to make a judgment about the causal efficacy of their responses. The scale that participants used to rate their effectiveness during the task ranged from  $-10$  to  $+10$ . This was to allow participants in the  $-75$  group to accurately rate their response–outcome contingency.

*Maze task* After completion of the three sessions, participants were immediately asked to complete the paper-and-

pencil maze-learning task. They were given the following instructions:

START FROM THE X MARKED AT THE TOP OF THE PAGE AND FIND YOUR WAY TO THE MIDDLE OF THE MAZE. SHOUT 'DONE' WHEN YOU GET THERE. THANK YOU ☺

Participants were timed (using a hand-held stop watch) in how long it took them to complete the maze. After this was finished, they were exposed correctly to the function acquisition test.

*Function acquisition test* The function acquisition task was identical to that in Experiment 1.

#### Results and discussion

Four participants' data were excluded from the analysis since they did not reach criteria for the emergent relations testing (28/32) after four training sessions. Seven participants were excluded due to their showing some signs of preexisting depression (BDI scores  $> 10$ ; see Reed et al., 2001). The mean score for the remaining participants on the BDI was 4.4 ( $SD = 3.0$ ). This left 10 participants in Group 75, 8 participants in Group 0, and 8 participants in Group  $-75$ .

*Derived stimulus relations testing* All remaining participants passed the emergent relations test, with an overall mean of 98.3% ( $SD = 3.3$ ) class-consistent responding. Thirteen participants reached the criteria (28/32) on the first test session. Seven participants needed a second training phase before reaching the test criterion, and 4 participants required a third. Two participants needed four training phases before satisfying the criteria.

*Total outcomes obtained* A higher number of overall outcomes were gained in Group 0 (mean number of outcomes per minute collapsed across sessions = 44.0,  $SD = 1.4$ ), followed by Group  $-75$  (mean number of outcomes per minute collapsed across sessions = 41.0,  $SD = 2.7$ ) and Group 75 (mean number of outcomes per minute collapsed across sessions = 28.8,  $SD = 12.5$ ). A mixed-model ANOVA (group  $\times$  session) revealed a significant main effect of group,  $F(2, 23) = 9.32$ ,  $\eta^2 = .45$ , but no main effect of session and no interaction,  $ps > .20$ . Tukey's HSD tests showed that there were significant pairwise differences between Group 75 and the other two groups (0 and  $-75$ ).

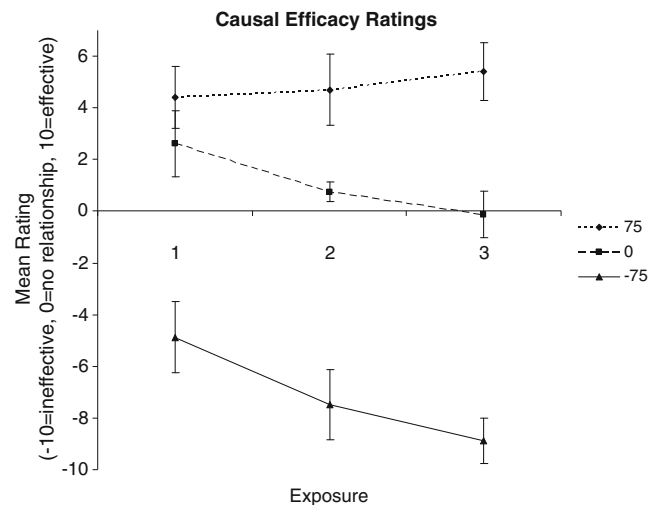
*Earned outcomes* The number of earned outcomes (contingent on participants' responses) was analyzed, which

revealed that Group 75 earned more points per minute (28.8,  $SD = 12.5$ ) than Group 0 (12.8,  $SD = 6.8$ ) and Group -75 (0). A mixed-model ANOVA (group $\times$ session) revealed a significant main effect of group,  $F(2, 23) = 24.88$ ,  $\eta^2 = .68$ , but no main effect of session and no interaction between the two factors,  $ps > .40$ . Tukey's HSD tests showed significant pairwise differences between all three groups.

**Actual delta probabilities** The contingent response–outcome probabilities were calculated as described for [Experiment 1](#). The noncontingent response–outcome probabilities were computed by calculating the number of “no” responses during each session. This was done by subtracting the number of responses from 120 (the duration [120 s] of each session). The numbers of noncontingent response–outcomes were then divided by this number. To calculate the dP value for each participant, the noncontingent outcome probability was subtracted from the contingent response–outcome probability for each session. As programmed, Group 75 had a dP of .73 ( $SD = .03$ ) averaged across trials, Group 0 had a dP of .02 ( $SD = .03$ ), and Group -75 had a dP of -.74 ( $SD = .03$ ).

**Efficacy ratings** As can be seen from inspection of the mean ratings of causal efficacy shown in [Fig. 4](#), judgments of causal efficacy were made in line with the dP contingencies. Ratings of causal efficacy were all positive for Group 75, as compared with negative ratings for Group -75. Group 0 started with positive ratings, but these ratings decreased over training. A mixed-model ANOVA (group $\times$ session) revealed significant main effects of group,  $F(2, 23) = 39.81$ ,  $\eta^2 = .78$ , and session,  $F(2, 46) = 4.68$ ,  $\eta^2 = .17$ , and a significant interaction between the two factors,  $F(4, 46) = 2.90$ ,  $\eta^2 = .20$ . Simple effect analyses were performed to investigate the effect of group at each session. For session 1,  $F(2, 23) = 14.42$ ,  $\eta^2 = .75$ , and session 2,  $F(2, 23) = 27.51$ ,  $\eta^2 = .84$ , there was a significant effect of group. Pairwise comparisons for these two sessions (with Bonferroni corrections) revealed that the efficacy ratings were significantly lower in Group -75 than in the other two groups (75, 0). For, session 3, there was also an effect of group,  $F(2, 23) = 53.22$ ,  $\eta^2 = .91$ . All pairwise differences between groups for session 3 were found to be significant.

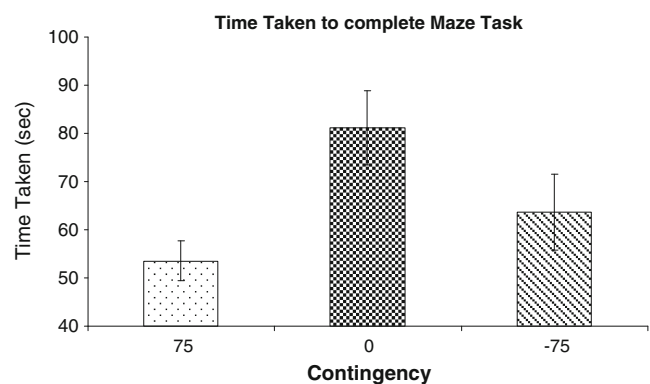
**Maze learning task** The results demonstrated a learned helplessness effect in the 75/75 contingency group, where there was no clear relationship between responding/not-responding and an outcome. Participants in this group took longer to complete the maze learning task than did the other two groups. The time taken for participants in each contingency group to complete the maze task can be seen in [Fig. 5](#). An ANOVA performed on these data revealed a significant main effect of group,  $F(2, 22) = 4.66$ ,  $\eta^2 = .55$ . Tukey's HSD tests showed significant differences between the 0%



**Fig. 4** Results from [Experiment 2](#): Mean rating of causal efficacy for each contingency group (dP=.75, .00, -.75) across each session. Error bars represent mean standard error

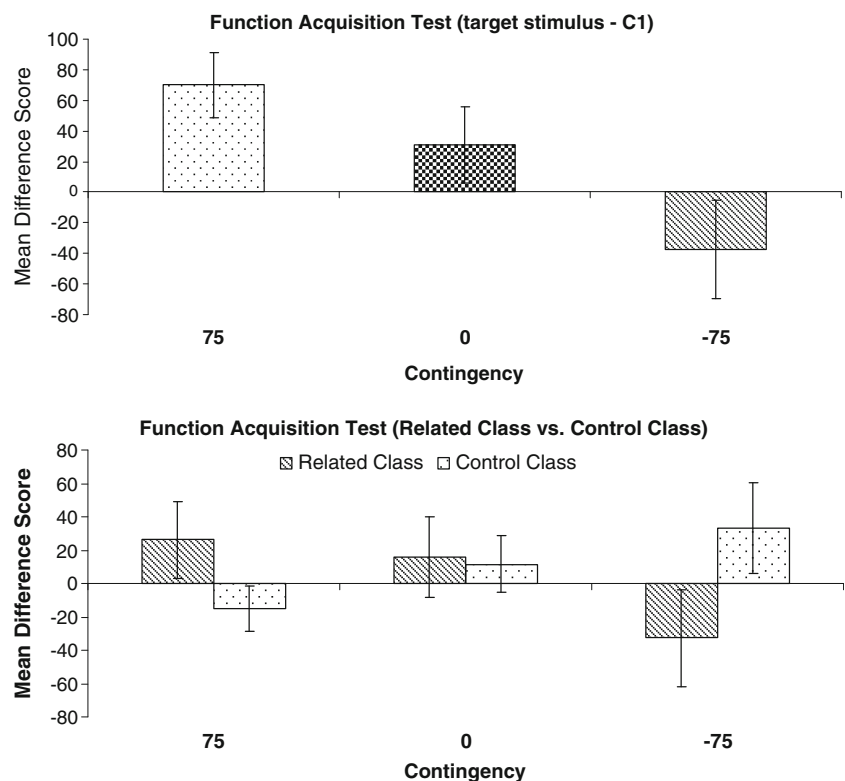
group, and the 75% group. This effect replicated previous research (e.g., Reed et al., 2001) that found exposure to uncontrollable contingencies interfered with subsequent learning.

**Function acquisition test** A difference score was calculated for the target stimulus (C1) as described in [Experiment 1](#). Inspection of the top panel of [Fig. 6](#) shows that in the 75 and 0 contingency groups, participants were more likely to rate themselves as “good” when the stimulus was presented. In contrast, in the -75 group, participants showed a negative transfer to the stimulus and categorized themselves as “bad.” An ANOVA revealed a significant main effect of group,  $F(2, 23) = 4.43$ . Tukey's HSD tests revealed differences between the 75 and -75 groups, with significantly higher scores (positive) in the 75 group than in the -75 group (negative). In addition, the difference scores for the



**Fig. 5** Results from [Experiment 2](#). Top panel: Mean difference score (i.e., the percentage “good”–the percentage “bad”) for the target stimulus in each contingency group (.75, .00, -.75). Bottom panel: Mean difference scores for the related class and the control class, in each contingency group. Error bars represent mean standard error

**Fig. 6** Results from Experiment 2: Time taken (in seconds) to complete the maze task for each response–outcome contingency group. Error bars show mean standard error



target stimulus in each contingency group were tested against zero (reflecting a neutral valence). In the 75 group, the difference scores of the target stimulus were significantly higher than zero,  $t(9) = 3.28, p < .01$ . In the 0 and -75 groups, there were no significant differences between the participant's difference scores for the CS and zero, all  $ps > .2$ . A similar transfer of causal efficacy to the stimuli (A1 and B1 percentages were combined and averaged) that were in the same equivalence class as the target (C1) was noted numerically and is shown in the bottom panel of Fig. 6. Participants in the 75 and 0 groups showed a positive transfer; and the -75 group showed a negative transfer. This trend occurred despite the fact that neither stimulus had been directly paired with the judgment-of-contingency task. The difference score for the control class (A2, B2, and C2, combined and averaged) showed the opposite trend; as the response–outcome contingency decreased, the difference scores increased.

A mixed-model ANOVA (group  $\times$  class) revealed no statistically significant main effects of group or class, both  $F$ 's  $< 1$ , but the interaction between the two factors approached statistical significance,  $F(2, 23) = 2.49, p < .09, \eta^2 = .18$ . Simple effects analyses were carried out (as recommended by Howell, 1997) to investigate the effect of group for each class. For the difference scores for the related class, a significant simple effect of group was found,  $F(2, 23) = 3.92, p < .05, \eta^2 = .34$ . Tukey's HSD test revealed higher scores in both the 75 and 0 groups, as compared with the -75 group,  $ps < .05$ . No significant simple effect of group was found for the scores for the control class,  $p > .05$ . In addition, the

difference scores for the related stimuli in each contingency group were tested against zero (reflecting a neutral valence). No significant differences were found between the participant's difference scores for the related stimuli and zero in any of the contingency groups, all  $ps > .2$ .

The results of the two preceding experiments have shown evidence that an evaluative function can transfer to a stimulus associated with the task that produces the evaluation and that this function will transfer through a derived relational network. In both experiments, efficacy judgments could be described by the dP rule. In Experiment 1, higher ratings of efficacy and positive transfer were associated with the outcome schedules with higher probabilities of an outcome given a response. The number of earned outcomes may have been another factor that impacted on the evaluative conditioning and transfer effects seen. The outcome schedules in which participants gained a high number of earned outcomes also produced high ratings of efficacy and more positive transfer, as compared with the outcome schedules in which participants gained a low number of earned outcomes. This was the case when all possible outcomes were earned by responding in Experiment 1 (e.g., points were received only when the participant gave a response) and also when it was possible to receive an outcome for not responding (Experiment 2). However, it should be noted that the transfer of causal efficacy ratings to the indirectly related stimuli (A1, B1) has been somewhat weaker than the direct transfer to the target stimulus particularly in Experiment 2, where outcomes were earned for responding and for

not responding. This is also especially pronounced in the case of the negative evaluations, which have shown less evidence of evaluative learning and transfer in the present experiments.

### Experiment 3

Dack et al. (2009) did show transfer of causal efficacy, both good and bad, to a CS and also generalization to stimuli that shared the same evaluative function (good, bad). The contrast between the methods used in the present set of experiments and the experiments by Dack et al. provides a potential explanation of the relative weaker transfer effects noted to the related stimuli in the present experiments. Participants in the study reported by Dack et al. were each exposed to two contingencies of reinforcement, so it was possible for them to compare their performance across the two contingencies. This was not possible in the present Experiment 2, since it used a between-subject design in which they were exposed to only one contingency. Participants had no other contingency to compare their performance with, which, in turn, may have produced the present weaker transfer results.

In order to address the explanation above, Experiment 3 gave participants a comparison against which to rate the contingency to which they were actually exposed. To this end, the present experiment manipulated the participants' 'expectancies about their performance. Cella, Taylor, and Reed (2007) manipulated participants' expectancies of an outcome by introducing pretask information and found behavior altered in line with these expectancies. In the present experiment, these instructions were adapted so that participants were given expectancies of efficacy that were (1) higher than the contingency that they would experience, (2) accurate relative to the contingency that they would experience, or (3) lower than their likely performance on the task. If evaluations of self-efficacy require a comparison in order to form strongly and transfer, this expectancy manipulation may impact on causal efficacy judgments, as well as the subsequent test for transfer.

In addition to this attempt to produce a stronger effect using a between-subjects design, the present experiment investigated whether a learned helplessness effect would be produced when participants were given expectations about their performance in the task that they would not be able to match in actuality. After the differing expectancy instructions were given, all participants received a 0/75 contingency ( $dP = -.75$ ). It was hoped that by employing a negative contingency for all groups, stronger aversive conditioning and transfer would be observed than in the previous experiments. Focus was given to the potential negative evaluations in this final experiment, since these have been shown to be most resistant to transfer across stimuli.

## Method

### Participants and apparatus

Forty-six adults participated in the present experiment (13 male, 33 female). The ages of the participants ranged from 18 to 25 years, with a mean of 20.1 ( $\pm 1.7$ ) years of age. All participants were recruited as described in Experiment 2. Each participant had normal or corrected-to-normal eyesight and was not aware of the purpose of the experiment. The apparatus and materials were identical to those employed in Experiment 2.

### Procedure

Each participant was taken into a quiet room and was given a consent form to read and sign and the Beck's Depression Inventory to answer, before completing the experiment. The participants were assigned randomly to one of three groups: high expectancy ( $n = 15$ ), accurate expectancy ( $n = 16$ ), and low expectancy ( $n = 15$ ). The study consisted of three computer tasks, as described below.

*Derived stimulus relations training* Derived stimulus relations training was identical to that in Experiment 1 and 2.

*Judgment-of-contingency task* The judgment-of-contingency task was identical to that in Experiment 2, except that all participants were given the 0/75 ( $dP = -.75$ ) contingency schedule. In the instructions for each group, the participants' expectancies of performance were manipulated, so that it was higher than estimated actual performance, accurate, or lower than estimated performance. All of the instructions were exactly the same as those presented in Experiment 2, except that the participants were told about the number of points (higher, accurate, or lower than actual performance) that they were expected to get on each session of the task. In Group High (high expectations about performance), the following was added to the end of the instructions on the computer screen: "NOTE: The average number of points earned by participants on each 2 minute phase is 120 (range = 110–130)." In Group Accurate, the following was added to the end of the instructions on the computer screen: "NOTE: The average number of points earned by participants on each 2 minute phase is 80 (range = 70–90)." In Group Low, the following instruction was added: "NOTE: The average number of points earned by participants on each 2 minute phase is 40 (range = 30–50)."

All other aspects of the procedure were the same as in Experiment 2, including the completion of the maze paper-and-pencil task.

*Function acquisition test* The function acquisition test was identical to those in Experiment 1 and 2.

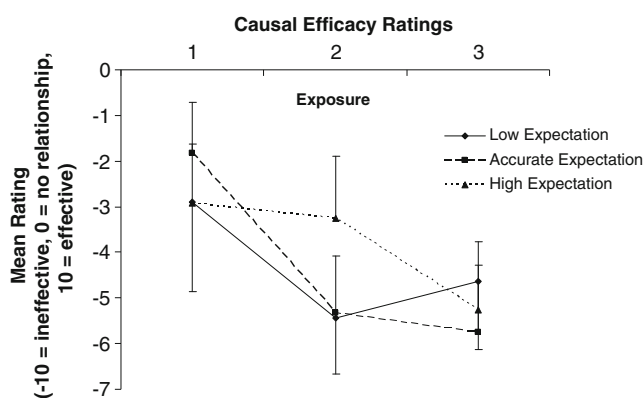
## Results and discussion

Three participant's data were excluded from the analysis since they did not reach the mastery criteria for the emergent relations testing (28/32) after six training sessions. Eight participants were excluded due to showing signs of preexisting depression (see Reed et al., 2001). The mean score for the remaining participants on the BDI was 4.7 ( $SD = 2.7$ ). This left 12 participants in the high-expectancy group, 12 participants in the accurate-expectancy group, and 11 participants in the low-expectancy group.

**Derived stimulus relations testing** All remaining participants passed the emergent relations test, with an overall mean of 98.1% ( $SD = 2.7$ ) class-consistent responding. Nineteen participants reached the criteria (28/32) on the first test session. Eleven participants needed a second training phase before reaching the test criterion needed to continue, and 3 participants required a third. Two participants required four respondent training phases before finally reaching the criterion.

**Total outcomes obtained** Since the same response–outcome contingency schedule ( $dP = -75$ ) was used for all three groups, the number of outcomes obtained per minute were very similar for each group: low-expectancy group mean from all three exposures = 38.6 ( $SD = 4.7$ ); accurate-expectancy group mean from all three exposures = 38.3 ( $SD = 5.0$ ); and high-expectancy group mean from all three exposures = 37.6 ( $SD = 4.2$ ). A mixed-model ANOVA (session  $\times$  group) revealed no significant main effects of session or group and no significant interaction, all  $ps > .20$ .

**Efficacy ratings** The mean ratings of causal efficacy for each group, after each exposure to the task, are shown in Fig. 7. An inspection of these data shows that ratings decreased across exposures in all groups. A mixed-model

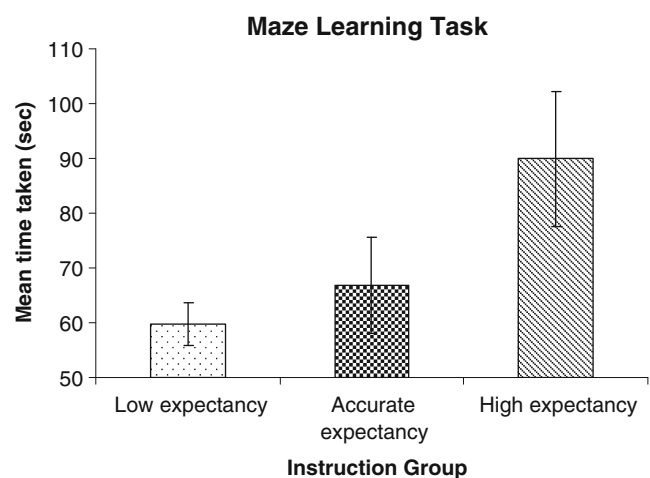


**Fig. 7** Results from Experiment 3: Mean rating of causal efficacy for each instruction group (high, accurate and low expectancies). Error bars represent mean standard error

ANOVA (group  $\times$  session) performed on these data revealed a significant main effect of session,  $F(2, 64) = 5.62$ ,  $\eta^2 = .15$ , but no significant main effect of group and no interaction between group and session,  $F_s < 1$ .

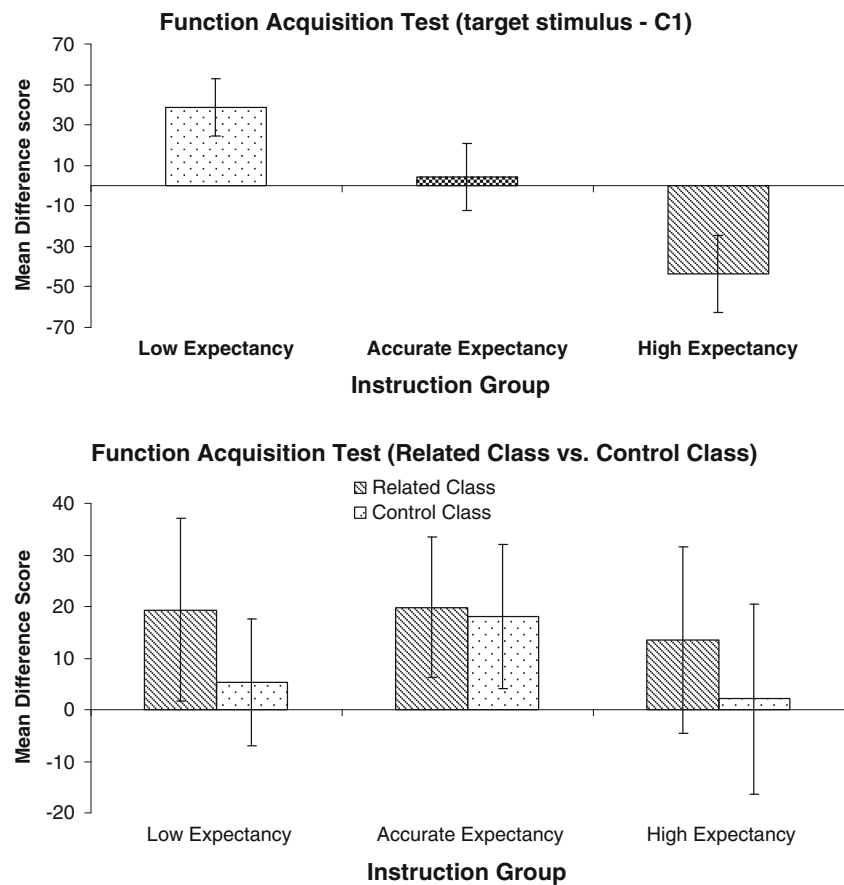
**The maze-learning task** The time taken for participants in each group to complete the maze task can be seen in Fig. 8. On average, participants in the high-expectancy group took longer to complete the task, as compared with the other two groups. An ANOVA performed on these data revealed a main effect of group in the time taken to complete the maze that was approaching significance,  $F(2, 32) = 2.88$ ,  $p = .07$ ,  $\eta^2 = .39$ . Tukey's HSD tests showed that the difference between the high- and low-expectancy group approached significance,  $p = .07$ . This suggests that having high expectations about a task that are impossible to match might induce a learned helplessness effect, where participants are much slower at completing a subsequent task. Having low expectations about a task, which are easily surpassed, may have the opposite effect and increase the speed of performance in a subsequent task.

**Function acquisition test** As with Experiment 1 and 2, a difference score (good – bad) was calculated for the target stimulus (C1). Inspection of the top panel of Fig. 9 shows that participants in the low-expectancy group had positive scores; they felt that they would be good at the task in the presence of the stimulus. In the high-expectancy instruction group, participants had negative scores, suggesting that they felt that they would be bad at a task if the stimulus appeared on the screen. An ANOVA revealed a significant effect of group,  $F(2, 32) = 6.00$ ,  $\eta^2 = .52$ . Tukey's HSD tests revealed significantly higher scores in the low-expectancy group than in the high-expectancy group. The difference scores for the



**Fig. 8** Results from Experiment 3: Time taken (in seconds) to complete the maze task for each instruction group. Error bars show mean standard error

**Fig. 9** Results from Experiment 3. Top Panel: Mean difference score (i.e., the percentage “good”–the percentage “bad”) for the target stimulus in each instruction group (low expectation, accurate expectation, high expectation). Bottom panel: Mean difference scores for the related class and the control class. Error bars represent mean standard error



target stimulus in each contingency group were tested against zero (reflecting a neutral valence). In the low-expectancy group, the difference scores of the target stimulus were significantly higher than zero,  $t(10) = 2.75$ ,  $\eta^2 = .66$ . In the high-expectancy group, the difference scores were significantly lower than zero,  $t(11) = 2.30$ ,  $\eta^2 = .60$ . In the accurate-expectancy group, there was no significant difference between the participant’s difference scores and zero, all  $t$ s < 1.

The bottom panel of Fig. 9 shows the difference scores for the stimuli in the related class (A1 and B1, combined and averaged) and for the stimuli in the control class (A2, B2, and C2, combined and averaged) for each instruction group. Participants tended to have positive scores for the related class. The control class was also given positive scores, but these were all closer to zero (neutral valence). A mixed-model ANOVA (group  $\times$  class) revealed no significant main effect of group or class and no significant interaction between the two factors,  $F$ s < 1.

In summary, Experiment 3 showed that a discrepancy between participants’ expectations about how well they should perform on a task and how well they actually performed had some impact on their behavior. If there was a negative discrepancy between the participants’ expectancy and their performance, helplessness was noted, at least in terms of the time taken to learn the maze (however, the

differences between the low- and high-expectancy groups were found only to be approaching significance, so the results must be interpreted with caution). These negative effects impacted upon the categorization of the target stimulus paired with the task, in that the stimulus was given a negative rating that was significantly lower than zero (neutral in affect).

This experiment has shown that false perceptions of control can produce negative evaluations about a stimulus that is paired with these perceptions. This suggests that to get a strong transfer of negative evaluations, it may be important that participants have some sort of comparison to evaluate their performance against. However, despite this transfer, the present experiment did not find any evidence of further transfer of these ratings to stimuli in the same equivalence class as the target stimulus.

## General discussion

The aim of the present research was to demonstrate that exposure to different response–outcome contingencies would produce different ratings of causal efficacy in a between-subjects design. This was thought to be important, since most previous work has used a within-subjects design

to produce these effects (e.g., Dack et al., 2009; Reed 2001a, b; Wasserman et al., 1983). Most of the results obtained in the present experiments demonstrated that participants were able to evaluate accurately different response–outcome contingencies and that these ratings could generally be described by the dP rule (Wasserman et al., 1983), although it should be noted that other rules may also accommodate the present data.

However, there were a few exceptions to this generalization that may be worth brief comment. In **Experiment 2**, participants in the group with a dP of 0 (75/75 group) greatly overestimated the efficacy of their responses in the first session. This finding could have been due to a significantly higher number of outcomes obtained in this group, as compared with the other two groups (Wasserman et al., 1993). Another possible explanation could be that one exposure to the contingency was not enough for participants to accurately rate their causal efficacy when outcomes were contingent on responding, as well as being noncontingent (Lopez & Shanks, cited in Shanks, 1995). This suggestion is supported, since the participants in the 75/75 (dP=0) group in **Experiment 2** were accurate at rating their causal efficacy by session 3.

It should be noted that the complexities of the instructions for the judgment-of-contingency task may have also impacted the participants' causal efficacy ratings. Participants were asked to find out whether pressing the space bar had any effect on whether or not they received a point, as well as being told that their task was to earn as many points as possible. This may have meant that participants were either rating their ability at gaining points in the task or rating the contingency between responding and outcomes. Research investigating the impact of instructions upon efficacy ratings could address this point in the future.

**Experiment 2** and **3** also investigated whether a learned helplessness effect could be produced by exposure to contingencies in which outcomes may be perceived as uncontrollable (**Experiment 2**) and by the inclusion of instructions that gave participants high expectations about how they would perform on the task that they were not able to match (**Experiment 3**). A learned helplessness effect was reported in **Experiment 2**, where participants in the 75/75 (dP = 0) group took significantly longer to complete the maze learning task than did participants in the 75/0 (dP = .75) group. This replicates previous research (Reed et al., 2001) that showed that exposure to uncontrollable contingencies (i.e., where outcomes are just as likely to follow a response as they are no response) interfered with subsequent learning. The effect was not significant in **Experiment 3**, although the differences between the low-expectancy and high-expectancy group were approaching significance. This may have been because the expectancy manipulations employed did not lead to significant differences in the

ratings of causal efficacy. This suggests that false perceptions of control may impair subsequent learning, but only if they also impact perceptions of causal efficacy. Further research would be needed to investigate this hypothesis.

Additionally, the experiments examined whether these ratings of causal efficacy would transfer to a stimulus presented during the judgment-of-control task, as well as transferring to stimuli that had previously been related to that target stimulus through derived stimulus relations. All three experiments showed a general trend of evaluative learning in the judgment-of-contingency task, replicating previous research (Dack et al., 2009). The stimulus paired with each contingency, typically, was categorized in accordance with participants' causal efficacy ratings/response–outcome contingencies or performance expectancy. It should be noted that it is unclear whether this target stimulus was serving as a standard conditioned stimulus, as in an evaluative conditioning experiment, or, since it was present continually during the task, whether it functioned as a contextual stimulus.

The present experiments also extend the work reported by Dack et al. (2009) by showing, under some conditions (**Experiment 1** and **2**), the transfer of function from the target to stimuli not present during the task, through derived stimulus relations in existing equivalence classes. This was thought to be an important aspect to examine, since the transfer of valence from the target stimulus to new stimuli, through derived stimulus networks, could be a possible mechanism behind the acquisition of evaluative judgments in the absence of direct experience of the event. The latter phenomenon is thought to be often the case in clinical contexts (Hermans, 1998).

In **Experiment 3**, the evaluations given to the target stimulus in the low-expectancy and high-expectancy groups were not found to transfer to the stimuli linked to the target stimulus through equivalence relations. The difference between the function acquisition effects in **Experiment 1** and **2** and **Experiment 3** may have occurred for a number of reasons. The three groups in **Experiment 3** were all exposed to the same response–outcome contingency (dP = -.75) but had different expectancies about their performance, whereas the three groups in **Experiment 1** and **2** were exposed to three different response–outcome contingencies. This led to significant group differences in causal efficacy ratings in the first two experiments, but not in the third experiment. It would seem that evaluations based on expectancies do not show the same differentiation across groups or the same degree of transfer via derived relations as do evaluations based on exposure to differing response–outcome contingencies. Further research would be needed to resolve this issue.

In the present experiments, the transfer of affect to the target stimulus and to the related class members was

stronger when there was a strong positive contingency between outcomes and responding. Participants with such contingencies consistently produced positive transfer scores. This differential effect of transfer (or spread of affect) mirrors the findings from previous research (e.g., Hammerl & Grabitz, 1996; Walther et al., 2005) that also found a transfer only on the basis on appetitive conditioning, and not aversive conditioning. This could reflect that people generally desire to view themselves positively, and, therefore, stimuli related to the self are also viewed positively (Feys, 1995; Walther & Trasselli, 2003). This view has been supported by research, showing that people often overestimate judgments of control (Alloy & Abramson, 1979) and that people tend to make self-serving attributions (Zuckerman, 1979) as well as self-serving evaluations of ability (Dunning, Meyerowitz, & Holzberg, 1989). It is not entirely clear why such a differential transfer effect emerges, and this could be the focus of future research.

In conclusion, this series of experiments used outcome probabilities and their associated judgments of efficacy (i.e., “I feel effective during this task,” “I feel ineffective during this task”) in an evaluative learning paradigm and showed that such properties could transfer to stimuli present during a schedules task. Importantly, the judgments made about participants’ performances were shown to transfer to novel stimuli related to the target stimulus through derived stimulus relations, despite the fact that these stimuli were never directly related to the schedules task. When participants were exposed to outcome probabilities where there was no relationship between responding, or not responding, and an outcome, a learned helplessness effect was seen in a subsequent task. These are important findings to help in the understanding of clinical disorders such as anxiety and depression, by investigating the contingencies that may produce negative and positive evaluations about the self and suggesting a possible mechanism by which these evaluations can spread to other stimuli that have not been in direct contact with these contingencies.

**Author note** Charlotte Dack, Louise McHugh, and Phil Reed, Department of Psychology, Swansea University, Swansea, Wales.

We thank the anonymous reviewers and the Action Editor Randolph Grace for their detailed and helpful comments. The research described here was supported by a research grant from the United Kingdom Economic and Social Research Council (ESRC).

Portions of these data were presented at the Symposium on Derived Stimulus Relations at the fourth conference of the European Association of Behaviour Analysis, Madrid, 2008 and at the Symposium on Conditional Discriminations at the Annual Conference of the Association for Behavior Analysis International, Phoenix, 2009.

Correspondence concerning this article should be addressed to Charlotte Dack, Department of Psychology, Swansea University, Singleton Park, Swansea, SA2 8PP, Wales, U.K. Electronic mail may be sent to charlottedack@hotmail.com.

## References

- Abramson, L. Y., Metalsky, G. I., & Alloy, L. B. (1989). Hopelessness depression: A theory-based subtype of depression. *Psychological Review*, *96*, 358–372.
- Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*, 147–149.
- Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, *108*, 441–485. doi:10.1037/0096-3445.108.4.441
- Alloy, L. B., & Tabachnik, M. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, *91*, 112–149.
- Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behavior Research and Therapy*, *30*, 133–142.
- Barnes, D., & Roche, B. (1997). Relational frame theory and the experimental analysis of human sexuality. *Applied and Preventative Psychology*, *6*, 117–135.
- Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York: Hoeber.
- Beck, A. T. (1978). *The depression inventory*. Philadelphia: Center for Cognitive Therapy.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.
- Breen, F. M., Plomin, R., & Wardle, J. (2006). Heritability of food preferences in young children. *Physiology and Behavior*, *88*, 443–447.
- Bush, K. M., Sidman, M., & de Rose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior*, *51*, 29–46. doi:10.1901/jeab.1989.51-29
- Cella, M., Taylor, K., & Reed, P. (2007). Violation of expectancies produces more false positive reports in a word detection task in people scoring high in unusual experiences scale. *Personality and Individual Differences*, *43*, 59–70.
- Chatlosh, D. L., Neunaber, D. J., & Wasserman, E. A. (1985). Response–outcome contingency: Behavioral and judgmental effect of appetitive and aversive outcomes with college students. *Learning and Motivation*, *16*, 1–34.
- Clark, D. A., Beck, A. T., & Alford, B. A. (1999). *Scientific foundations of cognitive theory and therapy of depression*. New York: Wiley.
- Dack, C., McHugh, L., & Reed, P. (2009). Generalization of causal efficacy judgments after evaluative learning. *Learning & Behavior*, *37*, 336–348.
- De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*, 853–869.
- Dohr, K. B., Rush, A. J., & Bernstein, I. H. (1989). Cognitive biases and depression. *Journal of Abnormal Psychology*, *98*, 263–267.
- Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology* (American Psychological Association) *57*, 1082–1090.
- Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *Behavior Analyst*, *23*, 239–254.
- Ellis, L., & Bonin, S. L. (2003). Genetics and occupation related preferences: Evidence from adoptive and non-adoptive families. *Personality and Individual Differences*, *35*, 929–937.



- Feys, J. (1995). Mere ownership: Affective self-bias or evaluative conditioning? *European Journal of Social Psychology*, *25*, 559–575.
- Field, A. (2006). I don't like it because it eats sprouts: Conditioning preferences in children. *Behavior Research and Therapy*, *44*, 439–455.
- Fulcher, E. P., Mathews, A., Mackintosh, B., & Law, S. (2001). Evaluative learning and the allocation of attention to emotional stimuli. *Cognitive Therapy and Research*, *25*, 261–280. doi:10.1023/A:1010732328104
- Hammerl, M., & Grabitz, H.-J. (1996). Human evaluative conditioning without experiencing a valued event. *Learning and Motivation*, *27*, 278–293.
- Hermans, D. (1998). Evaluative conditioning. *Gedragstherapie*, *31*, 3–6.
- Hermans, D., Baeyens, F., Lamote, S., Spruyt, A., & Eelen, P. (2005). Affective priming as an indirect measure of food preferences acquired through odor conditioning. *Experimental Psychology*, *52*, 180–186.
- Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). London: Wadsworth.
- Johnsrude, I. S., Owen, A. M., Zhao, W. V., & White, N. M. (1999). Conditioned preferences in humans: A novel experimental approach. *Learning and Motivation*, *30*, 250–264.
- Leader, G., & Barnes-Holmes, D. (2001). Matching-to-sample and respondent-type training as methods for producing equivalence relations: Isolating the critical variable. *The Psychological Record*, *51*, 429–444.
- Levey, A. B., & Martin, I. (1975). Classical conditioning of human "evaluative" responses. *Behavior Research and Therapy*, *13*, 221–226. doi:10.1016/0005-7967(75)90026-1 DOI:dx.doi.org
- Loeb, A., Feshbach, S., Beck, A. T., & Wolf, A. (1964). Some effects of reward upon the social perception and motivation of psychiatric patients varying in depression. *Journal of Abnormal and Social Psychology*, *68*, 609–616.
- MacLeod, A. K., & Byrne, A. (1996). Anxiety, depression and the anticipation of future positive and future negative experiences. *Journal of Abnormal Psychology*, *105*, 286–289.
- Maldonado, A., Martos, R., & Ramirez, E. (1991). Human judgements of control: The interaction of the current contingency and previous controllability. *Quarterly Journal of Experimental Psychology*, *43B*, 347–360. doi:10.1080/14640749108401274
- Morgan, J., & Banerjee, R. (2006). Social anxiety and self-evaluation of social performance in a nonclinical sample of children. *Journal of Clinical Child and Adolescent Psychology*, *35*, 292–301.
- Neunaber, D. J., & Wasserman, E. A. (1986). The effects of unidirectional versus bidirectional rating procedures on college students' judgments of response outcome contingency. *Learning and Motivation*, *17*, 162–179.
- Overmier, J. B., & Seligman, M. E. P. (1967). Effects of inescapable shock upon subsequent escape and avoidance learning. *Journal of Comparative and Physiological Psychology*, *63*, 23–33.
- Reed, P. (1999). Effect of perceived cost on judgments regarding the efficacy of investment. *Journal of Economic Psychology*, *20*, 657–676. doi:10.1016/S0167-4870(99)00029-X
- Reed, P. (2001a). Human response rates and causality judgments on schedules of reinforcement. *Learning and Motivation*, *32*, 332–348. doi:10.1006/lmot.2001.1085
- Reed, P. (2001b). Schedules of reinforcement as determinants of human causality judgments and response rates. *Journal of Experimental Psychology. Animal Behavior Processes*, *27*, 187–195.
- Reed, P., & Antonova, M. (2007). Interference with judgments of control and attentional shift as a result of prior exposure to controllable and uncontrollable feedback. *Learning and Motivation*, *38*, 229–241.
- Reed, P., Frasquillo, F., Colkin, C., Liemann, V., & Colbert, S. (2001). Interference with judgements of control and learning as a result of prior exposure to controllable and uncontrollable feedback during concept learning tasks. *Quarterly Journal of Experimental Psychology*, *54B*, 167–183. doi:10.1080/02724990042000155
- Roth, D., & Rehm, L. P. (1980). Relationships among self-monitoring processes, memory and depression. *Cognitive Therapy and Research*, *4*, 149–157.
- Rozin, P., Markwith, M., & McCauley, C. (1994). Sensitivity to indirect contacts with other person: AIDS aversion as a composite of aversion to strangers, infection, moral taint, and misfortune. *Journal of Abnormal Psychology*, *103*, 495–504. doi:10.1037/0021-843X.103.3.495
- Rozin, P., & Millman, L. (1987). Family environment, not heredity, accounts for family resemblances in food preferences and attitudes: A twin study. *Appetite*, *8*, 125–134.
- Seligman, M. E. P. (1975). *Helplessness: On depression, development, and death*. San Francisco: Freeman.
- Seligman, M. E. P., & Maier, S. F. (1967). Failure to escape traumatic shock. *Journal of Experimental Psychology*, *74*, 1–9.
- Shanks, D. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology*, *48A*, 257–279. doi:10.1080/14640749508401390
- Sidman, M. (1971). Reading and auditory–visual equivalences. *Journal of Speech and Hearing Research*, *14*, 5–13.
- Sidman, M. (1990). Equivalence relations: Where do they come from? In D. E. Blackman & H. Lejeune (Eds.), *Behavior analysis in theory and practice: Contributions and controversies* (pp. 93–114). Hillsdale: Erlbaum.
- Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Authors Cooperative.
- Sidman, M., Kirk, B., & Wilson-Morris, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior*, *43*, 21–42. doi:10.1901/jeab.1985.43-21
- Simon, G. E. (2003). Social and economic burden of mood disorders. *Biological Psychiatry*, *54*, 216–226.
- Stuart, E. W., Shimp, T. A., & Engle, R. W. (1987). Classical conditioning of consumer attitudes: Four experiments in an advertising context. *Journal of Consumer Research*, *14*, 334–351.
- Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, *82*, 919–934.
- Walther, E., & Trasselli, C. (2003). I like him, because I like me. Self-evaluation as a source of interpersonal attitudes. *Experimental psychology*, *50*, 239–246.
- Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition and Emotion*, *19*, 175–196. doi:10.1080/02699930441000274
- Wasserman, E. A. (1990). Detecting response–outcome relations: Toward an understanding of the causal texture of the environment. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 27–82). London: Academic Press.
- 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*, 406–432.
- Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: Role of probability in judgments of response–outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 174–188.
- Zuckerman, M. (1979). Attribution of success and failure revisited or: The motivational basis is alive and well in attribution theory. *Journal of Personality*, *47*, 245–287.