Learning & Behavior
2009, 37 (4), 336-348
doi:10.3758/LB.37.4.336

Generalization of causal efficacy judgments after evaluative learning

Charlotte Dack, Louise McHugh, and Phil Reed
Swansea University, Swansea, Wales

In three experiments, we examined the effect of response–outcome relations on human ratings of causal efficacy and demonstrated that such efficacy ratings transfer to novel situations through derived stimulus relations. Causal efficacy ratings were higher, and probability of an outcome given a response was lower, for a differential reinforcement of high rate schedule than for either a differential reinforcement of low rate schedule (Experiment 1) or a variable interval schedule (Experiment 2). In Experiment 3, we employed schedules that were equated for outcome probability and noted that ratings of causal efficacy and the rate of response were higher on a variable ratio than on a variable interval schedule. For participants in all three experiments, causal efficacy ratings transferred to the stimulus present during each schedule and generalized to novel stimuli through derived relations. The results corroborate the view that schedules are a determinant of both response rates and causal efficacy ratings. In addition, the novel demonstration of a mechanism of generalization of these ratings via derived relations has clinical implications.

Likes and dislikes play a vital role in many aspects of people's lives, including their social interactions (e.g., Zajonc, 1980), emotions (e.g., Lazarus, 1991; Scherer, 1993), and consumer behaviors (e.g., Stuart, Shimp, & Engle, 1987). Preferences may well result from how stimuli within the environment are evaluated, and one approach to understanding the processes behind the development of preferences suggests that their acquisition emerges through associative learning, or *conditioning*, whereby a neutral stimulus and a preference become associated with one another. This transfer of valence through an association (i.e., the degree to which an individual likes or dislikes a stimulus) 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 (the conditioned stimulus [CS]) becomes altered by the valence of stimuli/events (the unconditioned stimulus [US]) that are presented with, or are dependent on, the CS. For example, Zellner, Rozin, Aron, and Kulish (1983) and Baeyens, Eelen, Van den Bergh, and Crombez (1990) showed that when a novel flavor (the CS) was paired with a liked or disliked flavor (the US), the valence of the novel flavor changed in the direction of the already liked or disliked flavor (i.e., the participants began to either like or dislike the previously neutral novel flavor).

Evaluative conditioning may also play a role in determining a wide range of human behaviors, including the development of several clinical disorders (Hermans, 1998). For example, in anxiety, depression, phobias, and so on, certain types of evaluations, such as self-efficacy, causal effectiveness, self-esteem, and self-evaluations about areas of a person's life, may be acquired through evaluative conditioning processes (see Fulcher, Mathews, Mackintosh, & Law, 2001). In these types of disorders, the negative evaluations that patients often have about themselves can be conceptualized as the USs. Any aspects of the patient's life, such as other individuals, objects, or events that they have regular contact with, can be thought of as the CSs. Through evaluative conditioning, these negative evaluations of the self may transfer to any stimuli that the patients come in contact with. For example, in depression, if negative self-evaluations are a symptom, and these evaluations are able to generalize to people or objects, especially those closely related to the self through evaluative conditioning, a negative cascade could be started that may exacerbate the condition (Walther, Nagengast, & Trasselli, 2005).

Evaluative conditioning may explain how negative self-evaluations could potentially lead to and maintain certain clinical disorders, such as depression and anxiety (Fulcher et al., 2001). Operant conditioning procedures may be useful models for investigating the processes underlying evaluative conditioning, since similar mechanisms may be involved in the two types of conditioning. Experimentally, differential evaluations of self-efficacy and causal effectiveness have been produced by exposing experimental participants to various schedules of reinforcement in which the relationship between an action and an outcome is manipulated. Those schedules of reinforcement with a strong action–outcome relationship (e.g., differential reinforcement of high rate [DRH] or variable ratio [VR]

C. Dack, charlottedack@hotmail.com

schedules) have been found to produce high ratings of causal efficacy, often along with high rates of response. This finding contrasts with those from conditions with weak action–outcome relationships (such as differential reinforcement of low rate [DRL] and variable interval [VI] schedules), which tend to produce low ratings of causal efficacy and low rates of response (Reed, 1994, 1999, 2001a, 2001b, 2003). Such effects have particularly been studied in the context of learned helplessness, which is often used as a model of depression (Maldonado, Martos, & Ramirez, 1991; Reed, Frasquillo, Colkin, Liemann, & Colbert, 2001). Thus, exposure to conditions that generate only a weak perception of the occurrence of action-dependent outcomes will produce negative evaluations about the person's efficacy. Our first aim in the present study is to ascertain whether these negative evaluations transfer to the context in which those evaluations are learned, as was suggested by Walther et al. (2005).

A feature of many clinical disorders is the extent to which such evaluations (including self-evaluations of efficacy and effectiveness) generalize across many circumstances—for example, depressive attributional style (Walther et al., 2005) and generalized anxiety (Morgan & Banerjee, 2006). It is unlikely that similar negative evaluative judgments would be directly conditioned across such a wide range of circumstances, and thus, the question is how such evaluations are generalized across so many aspects of a person's life. Evidence for the indirect effect of learning was shown by Rozin, Markwith, and McCauley (1994), who demonstrated that an aversion toward a person with AIDS could be generalized to objects associated with this person (sweater, car, bed, etc.) even though it was perfectly clear that the objects could not be affected. One possible approach to explaining the development of such generalized evaluations without direct experience of each circumstance is by appealing to the concept of derived stimulus relations, such as those seen in the formation of equivalence classes. Stimulus equivalence and derived stimulus relations more generally provide an example of a process in which stimuli can become associated with one another, despite having never been directly paired and despite their lack of any shared physical properties (Sidman, 1971).

Equivalence classes can be established by training a minimal number of relations between individual stimuli in a group. For example, if the 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 using a conditional discrimination paradigm. If a class has been established, many new emergent relations are formed between the stimuli that had not been taught directly. There are four types of emergent relations, and examples of each are given here for the group (ABC) described above: (1) reflexive relations (A→A, B→B, C→C), (2) symmetrical relations (B→A, C→B), (3) transitive relations (A→C), and (4) equivalence relations (C→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).

After an equivalence class is established and a function is established for one member of the class, that function may transfer to other members of that class in the absence of explicit training. For example, if A, B, and C are members of an equivalence class and A acquires anxiety eliciting functions through pairing with shock, B and C may acquire a similar function without being similarly associated with shock. 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 avoidant responses, preferences, self-discrimination, moods, and so on (see Dymond & Rehfeldt, 2000, for a review). One of our aims in the present study was to demonstrate how this phenomenon might also be implicated in the acquisition of evaluative judgments, in that the transfer of the latter to previously neutral stimuli might occur through a process that includes transfer of function through derived relations.

In the present experiments, we sought to establish whether evaluative ratings of causal efficacy (high vs. low causal efficacy), produced by exposure to particular schedules of reinforcement, will become associated with discriminative stimuli presented during the different schedules. If the causal efficacy ratings are found to transfer to the discriminative stimuli, we will also explore whether the discriminative stimuli (low vs. high causal efficacy) would later join established equivalence classes with the same evaluative function (good or bad) on the basis of a process that included transfer of function through derived relations.

A further aim of the present research was that we investigate which factors cause the causal efficacy ratings produced by different schedules of reinforcement (the rate of response, the rate of reinforcement, or the probability of an outcome given a response associated with each different schedule) and also the evaluations of the discriminative stimuli associated with each schedule (causal efficacy ratings or another factor determining these evaluations). It may be the case that, in different contexts (different schedules of reinforcement), different factors have a greater influence over causal efficacy ratings and the evaluations of the discriminative stimuli. In addition to promoting a basic understanding of evaluative learning, we will explore the effects in a way that is potentially important to understanding in applied contexts, such as psychotherapy. More specifically, the results of this study may help explain the processes behind the development of problems experienced by individuals that can occur without direct learning experience (e.g., a person reading that a particular food is unhealthy and subsequently disliking it without actually having tasted it), such as are observed in some anxiety disorders and phobias.

EXPERIMENT 1

In Experiment 1, we sought to establish basic schedule control over the causal efficacy ratings, replicating previous studies of humans (Reed, 1994, 1999, 2001a, 2001b) in which a schedule that is known to produce high response rates (DRH) has also been found to produce high ratings of causal efficacy relative to a schedule that produces low response rates (DRL). In Experiment 1, we also sought to show that these different ratings of causal efficacy (high vs. low) produced by the schedules (DRH vs. DRL) could transfer to discriminative stimuli (colored circles) present during the two different schedule tasks and then, in a later task, generalize further to other stimuli on the basis of a process that includes transfer of function.

The participants performed three simple tasks: They acquired two 3-member equivalence classes (in Class 1, the word *useless* was trained as a class member with Stimuli A1 and B1; in Class 2, the word *good* was trained in an equivalence class with Stimuli A2 and B2). In a separate task, the participants were asked to press the space bar on a keyboard and to rate how effective they felt that they were at gaining points as a result of their actions on two different schedules of reinforcement that were predicted to produce high (DRH) and low (DRL) causal efficacy ratings. A discriminative stimulus was present during each schedule. With the third task, we tested whether the causal efficacy ratings generated during the schedule task transferred to the discriminative stimuli presented on the computer screen during each schedule (e.g., the stimulus present during the DRH schedule, with higher causal efficacy ratings, will be rated as *good* and the stimulus present during the DRL schedule, with lower causal efficacy ratings, will be rated as *useless*). With the final task, we tested for conditioned transfer of evaluative functions.

Method

Participants

Fifteen adults participated in Experiment 1 (2 male, 13 female). The ages of the participants ranged from 18 to 33 years, with a mean of 21.1 years ($SD = 3.8$). All of the participants were students at Swansea University and were recruited through advertisements in the psychology department. In return for their participation, they earned subject pool credit. Each participant had normal or corrected-to-normal eyesight and did not suffer from color blindness. All were naive to the purpose of the experiment.

Apparatus and Materials

Experiment 1 was conducted in a quiet room free from distraction. The experimental room contained only a desk; a chair; and a computer with a 550-MHz processor, a 14-in. color monitor, and a standard computer mouse. The experiment was conducted entirely by means of the computer. All of the trial presentations and response recordings were controlled by programs written in Visual Basic 6.0. All responses involved either mouse clicking or space bar keypresses.

Derived stimulus relations training. The stimuli used as members of the two equivalence classes were nonsense words and affective words, as are shown in Table 1, along with their corresponding letter–number designation. All of the stimuli were composed of Arial Bold characters in black, each of which occupied a certain proportion of the screen (screen width/4 in.). Each stimulus was surrounded by a box, its width equaling 4 in. and its height equaling 1 in., against a white background. The positioning of the comparison stimuli was as follows: The horizontal position of the left stimulus started 1.25 in. from the left edge, and the horizontal position of the right stimulus started 1.25 in. from the right edge of the screen. Their vertical positions were 7.75 in. down the screen. All feedback choices were in red surrounded by a box that was 6 in. wide and 1.5 in. in height, presented in the middle of the screen.

Table 1
Nonsense and Affective Words Used As Stimuli and Their Assignment to Equivalence Classes

	Stimulus		
	A	B	C
Class 1	lewoly	gedeer	useless
Class 2	matser	rigund	good

Evaluative conditioning task. Two differently colored circles were used to signal the two schedule conditions (DRH and DRL) and acted as the discriminative stimuli. Six different color combinations were used (black/white, black/blue, black/yellow, yellow/blue, blue/white, yellow/white), and these were counterbalanced across the schedules so that each participant had a different color permutation for the stimuli used for the DRH and DRL schedules of reinforcement. This was done to ensure that the postevaluative ratings of the discriminative stimuli were due to the conditioning task and not to any preexisting preferences toward particular colors. The number of points earned throughout each schedule session was shown in the corner of the computer screen. At the end of each session, the participants were asked to rate how effective they were at earning points on a 10-point scale (1, *ineffective*; 10, *effective*).

Scales. Beck's Depression Inventory (BDI; Beck, 1978; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) was administered to assess the level of depression in the participants. Any participant scoring highly (>10) on this measure was excluded from the data analysis, as in previous studies (Reed et al., 2001). This was to ensure that the different schedules of reinforcement, and not any preexisting depressive condition, were responsible for the effects reported in this experiment. A short questionnaire was designed to allow the participants to rate the evaluative function of a collection of stimuli, including the nonsense syllables used as members of the two equivalence classes and the discriminative stimuli present during the conditioning task. For each stimulus, there was a 7-point scale (1, *useless*; 4, *neutral*; 7, *good*). From this scale, the participants' evaluations for each item could be measured and converted into a number score to test for any transfer of function. A consent form was constructed to inform the participants of the aim of the experiment and to assure them of the confidentiality of the results.

Procedure

Each participant was taken into a quiet room and given a consent form to read and sign and the BDI test before they completed the study. An overview of the experimental procedure is shown in the schematic diagram in Figure 1 and is also described below.

Derived stimulus relations training. All training and testing was conducted using a 2 × 3 matching-to-sample conditional discrimination paradigm using stimuli that consisted of two nonsense words and an affective word, as is shown in Table 1. Two three-member equivalence classes were established, by training AB and BC relations in a linear series structure. Each relation (A1–B1, A2–B2, B1–C1, and B2–C2) was presented at least three times during training. The criterion to proceed to the testing phase was 12 consecutively correct trials across all stages. There was no time limit for responding to individual trials.

Each trial started with the presentation of a sample (Sa) and two comparison stimuli. The positive comparison (Co+) stimulus was chosen from the same equivalence class as Sa, and the negative comparison (Co−) stimulus was chosen from the other class. The stimuli were displayed in an isosceles triangle display on the monitor, with Sa at the vertex of the triangle and Co+ and Co− at the corners of the base. The following instructions were presented across the middle of the screen on the first trial only: "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."

The participants chose a comparison by clicking on the left- or righthand box. When feedback was provided during training, the

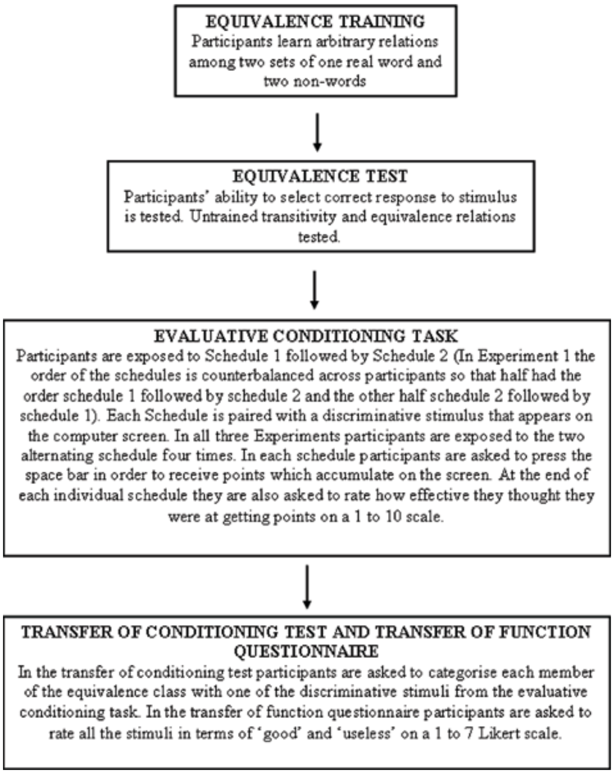

Figure 1. A schematic diagram showing each part of the overall procedure.

choice of the Co+ produced a 1-sec display of the word "Correct," whereas the choice of the Co− produced a 1-sec display of the word "Wrong." Both were displayed in red across the middle of the computer screen.

Testing emergent relations. Once the criterion for the training session had been met, the test phase commenced, and the corrective feedback was terminated. On the first test trial, the following instructions were shown across the middle of the computer 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. DURING THESE TRIALS THE COMPUTER WILL NOT GIVE YOU ANY FEEDBACK."

All tests for one-node transitivity (A→C) and equivalence (C→A) were presented in a single block. Each type of relation (A1–C1, A2–C2, C1–A1, and C2–A2) was presented nine times, with 36 trials in total. The mastery criterion for testing was at least 30 correct class-consistent selections across the block of 36 test trials.

Evaluative conditioning task. The participants were presented with the following instructions on the computer screen:

> You will shortly see a circle in the centre of the screen. The circle will be colored either color 1 or color 2.[1] When you see this circle you must press the space bar to earn points. Try to earn as many points as possible. You may press either quickly or slowly to earn points. You must work out what rate of pressing makes you the most points. Click here to proceed.

After presentation of these instructions, the participants were exposed to eight schedule sessions, which consisted of two alternating schedules (DRH and DRL). Each schedule was therefore presented four times to each participant, with each schedule session lasting 4 min in total. Each schedule was signaled by the presence of a different colored circle. The colors used as the discriminative stimuli in each schedule were counterbalanced so that each participant had a different combination of colors signaling the DRH and DRL schedules. In one schedule (DRH), the participants responded on a DRH 5/2 schedule: The participants were required to emit five responses in 2 sec to produce an outcome (receive a point). In a second schedule (DRL), the participants responded on a DRL 5-sec schedule. In this schedule, an outcome (point) was received immediately after a response on the space bar, provided that at least 5 sec had elapsed since the last response was received. A response before 5 sec had elapsed would reset the DRL time requirement. After each schedule session had concluded, the participants were asked to make a judgment of the causal efficacy of their responses. The following instructions were presented:

> 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.

After the participants had chosen a number, the next schedule was presented. For half of the participants, the DRL schedule was presented first, followed by the DRH schedule. For the other half of the participants, the order of presentation of the DRL and DRH conditions was reversed.

Transfer of conditioning test. The participants were presented with the following instructions on the computer screen: "Look at the Image Above and then Click on the Image Below that GOES WITH the one at the Top."

Each trial began with the presentation of one of the six stimuli, from either equivalence class, in the middle at the top of the screen, as an Sa, and the two discriminative stimuli from the evaluative conditioning task, in each corner at the bottom of the screen, as Co+ and Co−. The participants chose one of the discriminative stimuli by clicking on it with the mouse before moving on to the next trial. Each word from the derived stimulus relations training was shown four times, giving a total of 24 trials. At no point during this test were the participants given any feedback for their responses.

Six participants received the derived stimulus relations task before the evaluative conditioning task, and the other 7 received the tasks in the opposite order. All of the participants completed the transfer of conditioning test last. At the end of the experiment, each participant completed the transfer of function questionnaire, was thanked for participating, and was debriefed.

Results and Discussion

One participant's data were excluded from the analysis, because their causal efficacy ratings were extreme outliers (greater than 3 *SD* from the grand mean). Four participants were excluded because they showed 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 5.4 ($SD = 2.41$).

Evaluative Conditioning Task

The mean ratings of causal efficacy emitted during each exposure to each schedule are shown in Figure 2. These data show that ratings of causal efficacy were higher on the DRH than on the DRL schedule. In both schedules, the ratings increased with exposure. A two-factor ANOVA was performed on these data, with schedule (DRH vs. DRL) and session as within-subjects factors. A rejection criterion of $p < .05$ was adopted for this and all subsequent analyses. This analysis revealed a significant main effect of session [$F(3,27) = 6.63$] and a significant main effect of schedule [$F(1,9) = 32.11$]. No significant interaction was found between session and schedule ($p > .05$). Unfortunately, because of a programming malfunction,

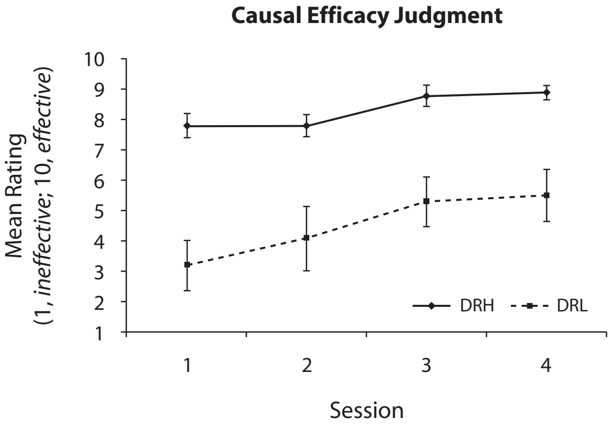

Figure 2. Results from Experiment 1. Mean ratings of causal efficacy over the four sessions to each schedule. Error bars represent the mean standard error. DRH, differential reinforcement of high rate; DRL, differential reinforcement of low rate.

the number of responses for each schedule across the four sessions was not recorded and so could not be analyzed.

Derived Stimulus Relations Testing

All of the participants passed the emergent relations test, with an overall mean of 96.39% (*SD* = 5.72) class-consistent responding. There were significantly more correct responses than would be predicted by chance [$t(9) = 25.66$]. All of the participants reached the criterion (>83%) on the first test session.

Transfer of Function Questionnaire

The mean ratings (*useless–neutral–good*) for each nonsense stimulus in each class and for the two discriminative stimuli that signaled the different schedules can be seen in Figure 3. The data show that lower ratings were given for stimuli associated with the word *useless* (Class 1, mean self-report rating = 1.55, *SD* = 1.09) than for stimuli associated with the word *good* (Class 2, mean self-report rating = 6.75, *SD* = 0.35). These differences were statistically significant for each cross-class stimulus pair [A1–A2, $t(9) = 12.53$; A1–B2, $t(9) = 20.25$; B1–A2, $t(9) = 7.15$; B1–B2, $t(9) = 10.11$]. The discriminative stimuli that signaled the DRH schedule were given higher valence ratings than the discriminative stimuli that signaled the DRL schedule [$t(9) = 2.11$]. This result suggests that the previously neutral stimuli, when paired with the different schedules of reinforcement, were rated in the same direction as were the causal efficacy judgments associated with them.

Transfer of Conditioning Test

Figure 4 reveals that the participants chose the discriminative stimuli associated with the high causal efficacy schedule (DRH) in the presence of the stimuli in the *good* equivalence class significantly more often than chance and significantly more than they chose the discriminative stimuli associated with the DRL schedule with the *good* equivalence class [$t(9) = 2.55$].

The present procedure allowed the assessment of the effect of schedules of reinforcement on human ratings of causal efficacy. Schedules reinforcing high rates of behavior (DRH) were shown to produce high ratings of causal efficacy relative to schedules producing low rates of behavior (see also Reed, 2001a, 2001b). We also demonstrated in this experiment that evaluative learning took place. The negative or positive evaluations associated with each schedule of reinforcement were shown to transfer to the context (the discriminative stimulus presented during training) in which those evaluations were learned. A further novel aspect of Experiment 1 is that the participants derived a relation between the stimuli (the colored circles) presented during each schedule and other stimuli in equivalence classes with the same evaluative function.

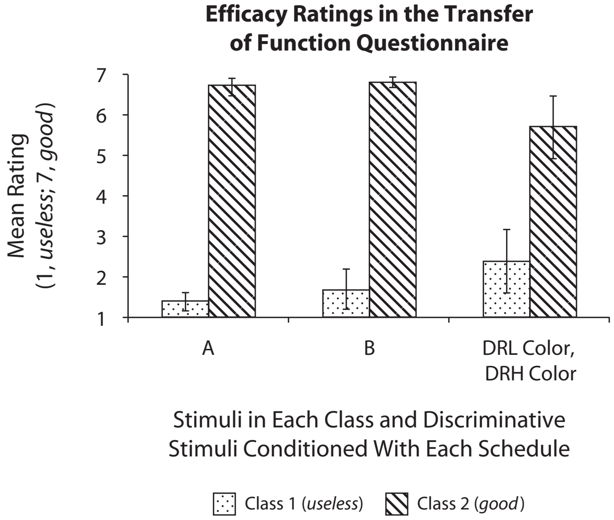

Figure 3. Results from Experiment 1. Mean rating for each nonsense stimuli (Class 1, *useless*; Class 2, *good*) and for the discriminative stimuli that signaled the two different schedules. Error bars show the mean standard error. DRH, differential reinforcement of high rate; DRL, differential reinforcement of low rate.

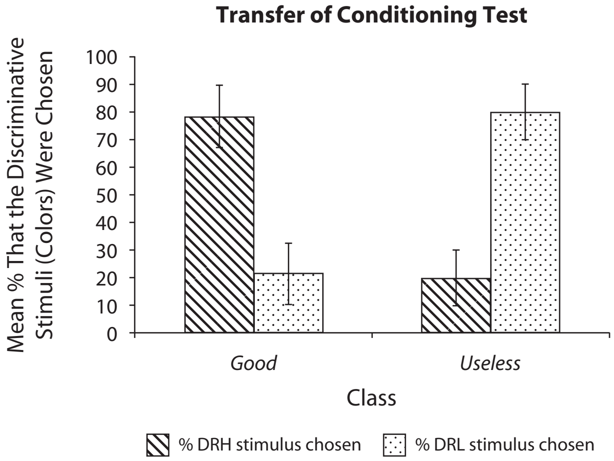

Figure 4. Results from Experiment 1. Mean percentages that the stimuli (colored circles) conditioned with each schedule were chosen with stimuli from each equivalence class (Class 1, *useless*; Class 2, *good*). Error bars represent the mean standard error. DRH, differential reinforcement of high rate; DRL, differential reinforcement of low rate.

This shows that the stimuli became functionally equivalent to one another, despite there being no direct contact between the discriminative stimuli and the members of each equivalence class. The discriminative stimuli associated with low ratings of causal efficacy tended to be classed with the *useless* equivalence class members, and the discriminative stimuli associated with high ratings of causal efficacy tended to be classed with the *good* equivalence class members. Thus, ratings of causal efficacy were shown to transfer to stimuli present during training and then to other stimuli associated with similar ratings.

However, although this experiment is the first to show such transfer-of-schedule-induced efficacy ratings through derived stimulus relations, it is not clear which determinants of the schedules influenced the causal efficacy ratings produced, or which factors caused the different evaluations of the discriminative stimuli presented during each schedule. For example, the difference in the mean number of outcomes was significantly higher in the DRH condition than in the DRL condition. Therefore, the relative rate of reinforcement associated with each schedule may have been an important factor influencing the participants' causal efficacy ratings. In Experiment 2, we investigated whether this is the case by equating the rates of reinforcement across each schedule in order to see whether there are still differences in causal efficacy ratings. Many factors may have caused the evaluations of the discriminative stimuli associated with each schedule: It may have been the differences in causal efficacy, the rate of reinforcement, or another determinant of the schedules.

EXPERIMENT 2

In Experiment 2, we explored whether schedules of reinforcement affect evaluations of causal efficacy when the number of outcomes obtained per minute is equated. To this end, participants first performed according to a master DRH schedule, with the number of outcomes per minute becoming the interval of a yoked VI schedule. Such a yoking procedure has been found to result in higher rates of response and higher causal efficacy judgments on a DRH schedule than on a VI schedule (Reed, 2003). In Experiment 2, we also sought to replicate the generalization effect seen in Experiment 1; that is, the causal efficacy ratings (high vs. low) could transfer to the stimuli (colored circles) present during each schedule of reinforcement (DRH and DRL) through evaluative conditioning and then later generalize to other stimuli through derived relations to each stimulus class when the rate of reinforcement for each schedule was equal.

Method

Participants

Thirty-one adults participated in the present experiment (7 male, 24 female). The ages of the participants ranged from 18 to 25 years, with a mean of 20 years ($SD = 1.86$). All of the participants were recruited by advertisements in the psychology department of Swansea University, and in return for their participation, they earned subject pool credit. Each participant had normal or corrected-to-normal eyesight and did not suffer from color blindness. All of the participants were naive to the purpose of the experiment.

Apparatus and Materials

The apparatus and materials were identical to those in Experiment 1, with the exception of the evaluative conditioning task. The only differences in this task were that different combinations of colored circles were employed to signal the two scheduling conditions and that these were DRH and VI schedules in this experiment. This was done to ensure that the results could be replicated using a wide range of colors, to ensure that the findings were not due to any preexisting preference for a particular color. Three different combinations were used (yellow/white, yellow/blue, blue/white), and these were counterbalanced across participants so that a different color was used to signal each of the schedules.

Procedure

Each participant was taken into a quiet room and given a consent form to read and sign and the BDI test before they completed the experiment.

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

Evaluative conditioning task. The participants were presented with the following instructions on the computer screen:

> You will shortly see a circle in the centre of the screen. The circle will be colored either color 1 or color 2. [See note 1.] When you see this circle you must press the space bar to earn points. Try to earn as many points as possible. You may press either quickly or slowly in order to earn points. You must work out what rate of pressing makes you the most points. Click here to proceed.

After presentation of these instructions, the participants were exposed to eight schedule sessions, which consisted of two alternating schedules (DRH and VI). Each schedule was, therefore, presented four times to each participant. The DRH schedule was always presented first and lasted for 2 min. Each VI schedule lasted until the participants had received the same amount of points as they had on the previous DRH schedule. Each schedule was signaled by the presence of a different-colored circle. These colors were counterbalanced so that all possible color combinations were used as the DRH and VI discriminative stimuli across participants.

In the first schedule, the participants responded on a DRH 5/2 schedule, as was described in Experiment 1. In the second schedule, the participants responded on a VI schedule yoked to the temporal distribution of outcomes obtained in the DRH schedule. This was done by dividing the total number of outcomes (points) per session on the DRH schedule by the total time of each session (120 sec). This number was then used as the overall mean time interval between outcomes (points) on the yoked VI schedule. Points were received only after a response was made after a time interval that varied around this mean value. After each schedule session, the participants were asked to make a judgment of their causal efficacy. The following instructions were presented:

> 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.

After the participants had chosen a number, the next schedule was presented. This continued until the participants had completed all eight schedule sessions.

Transfer of conditioning test. The transfer of conditioning test was identical to that in Experiment 1.

Fifteen of the participants received the derived stimulus relations task before the evaluative conditioning task, and 16 participants received the tasks in the opposite order. All of the participants completed the transfer of conditioning test last. At the end of the

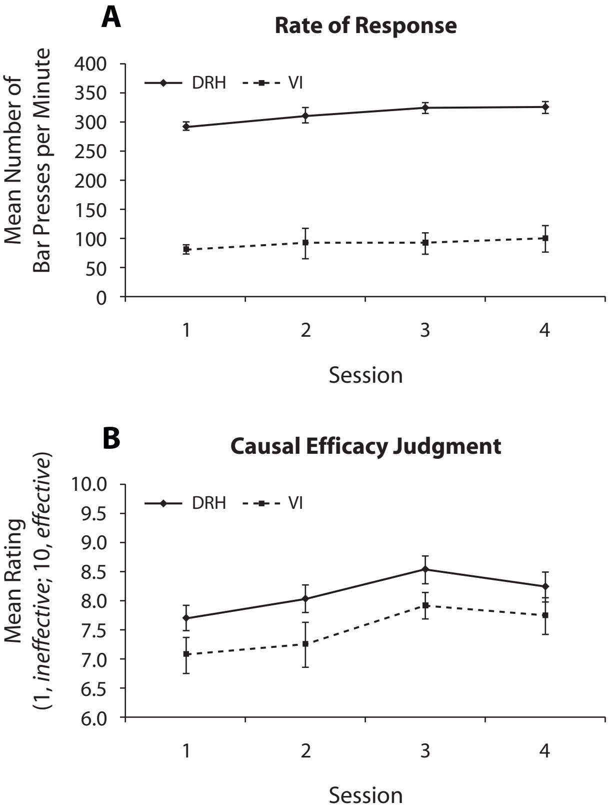

Figure 5. Results from Experiment 2. (A) Mean responses per minute over the four sessions for each of the schedules. (B) Mean ratings of causal efficacy over the four sessions to each schedule. Error bars show the mean standard error. DRH, differential reinforcement of high rate; VI, variable interval.

experiment, each participant completed the transfer of function questionnaire, was thanked for participating, and was debriefed.

Results and Discussion

On the basis of the BDI scores (see Reed et al., 2001), 3 participants were excluded from the analysis (BDI scores >10). The mean score for the remaining participants on the BDI was 3.61 ($SD = 2.75$).

Evaluative Conditioning Task

The mean rates of responses to the two schedules are shown in Figure 5A. These data show that the number of responses was higher on the DRH than on the VI schedule. A two-factor ANOVA was performed on these data, with schedule (DRH vs. VI) and session as within-subjects factors. This revealed a significant main effect of schedule [$F(1,27) = 303.21$]. There was no significant effect of session [$F(3,81) = 1.68$], nor was there a significant interaction between the two factors ($F < 1$).

The mean ratings of causal efficacy during each exposure to the schedules are shown in Figure 5B. These data show that ratings of causal efficacy were higher on the DRH than on the VI schedule. In both conditions, the ratings increased with exposure. A two-way, repeated measures ANOVA was performed on the data, with schedule (DRH vs. VI) and session as within-subjects factors. This revealed significant main effects of schedule [$F(1,27) = 12.00$] and session [$F(3,81) = 4.57$]. No significant interaction was found between session and schedule ($F < 1$).

Because the VI schedule was yoked to the temporal distribution of outcomes obtained on the DRH schedule, the number of outcomes was the same for the two schedules on each session. The mean number of outcomes were 95.82 ($SD = 14.36$) in Session 1, 104.89 ($SD = 25.92$) in Session 2, 110.18 ($SD = 20.98$) in Session 3, and 109.25 ($SD = 20.60$) in Session 4.

The mean probabilities of an outcome given a response for each schedule (DRH and VI) were higher on the VI (.65) than on the DRH (.17) schedule. A two-factor, repeated measures ANOVA was performed on these data, with schedule (DRH vs. VI) and session as within-subjects factors. This revealed a significant main effect of schedule [$F(1,27) = 183.69$]. There was no significant effect of session [$F(3,81) = 1.47$] and no significant interaction between session and schedule [$F(3,81) = 1.21$].

Derived Stimulus Relations Training and Testing

All of the participants passed the emergent relations test with an overall mean of 98.11% ($SD = 3.12$) class-consistent responding. This number is significantly more correct responses than would be expected by chance [$t(27) = 81.55$]. Twenty-five participants reached the passing criterion on the first test session. Three participants needed a second training phase before reaching the test criterion (i.e., 83% correct).

Transfer of Function Questionnaire

The mean ratings (*useless–neutral–good*) for each nonsense stimulus in each class and for the two discriminative stimuli that signaled the different schedules can be seen in Figure 6. The data show lower ratings (more *useless*) for the members of Class 1 than for those of Class 2. These differences were found to be significant for each cross-class pair [A1–A2, $t(27) = 11.15$; A1–B2, $t(27) = 15.20$;

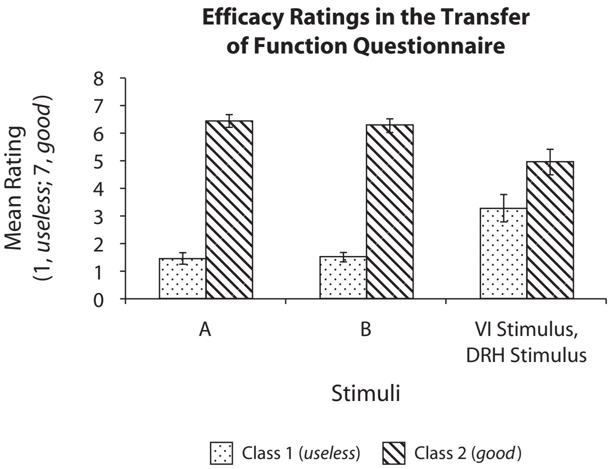

Figure 6. Results from Experiment 2. Mean rating for each nonsense stimulus (Class 1, *useless*; Class 2, *good*) and for the discriminative stimuli that signaled the two different schedules. Error bars represent the mean standard error. DRH, differential reinforcement of high rate; VI, variable interval.

B1–A2, $t(27) = 16.45$; B1–B2, $t(27) = 12.27$]. The discriminative stimuli that signaled lower response rates and lower ratings of causal efficacy (but high probability of an outcome given a response) were rated lower (more *useless*) by the participants than the discriminative stimuli that signaled higher response rates and higher ratings of causal efficacy. This shows that the valence of the efficacy ratings had transferred to the stimulus presented during the evaluative conditioning task. This difference was found to approach significance [$t(27) = 1.77$, $p = .08$].

Transfer of Conditioning Test

Figure 7 reveals that the participants chose the discriminative stimuli associated with the higher causal efficacy schedule (DRH) in the presence of the stimuli in the *good* equivalence class significantly more than chance and significantly more than they chose the discriminative stimuli associated with the VI schedule with the *good* equivalence class members [$t(27) = 2.08$].

In summary, the response rates were higher on a DRH schedule than on a VI schedule with the same frequency of reinforcement, mirroring previous findings (Reed, 2003). The ratings of causal efficacy also mirror this pattern, with higher ratings on the DRH schedule than on the VI schedule, despite equating the number of outcomes per minute. The valence of the efficacy ratings transferred to the stimulus (color) present during each schedule of reinforcement, as is shown in the results of transfer of function questionnaire and which replicates the findings of Experiment 1. The transfer of conditioning test results also show that the participants tended to class the discriminative stimuli with the stimuli in the equivalence classes with the same evaluative function.

Although the findings of Experiment 2 replicated those of Experiment 1, the differences in causal efficacy ratings were much smaller between each schedule of reinforcement, and the difference between the ratings of the discriminative stimuli associated with each schedule was only found to approach significance. This result suggests that the reinforcement rate may predict causal efficacy ratings, as well as the ratings of discriminative stimuli associated with each schedule. Also, by the fourth session of each schedule, there appears to be very little difference in the causal efficacy ratings. This result might suggest that as participants become more experienced with each schedule, the relationship between responding and reinforcement may become an important factor in determining causal efficacy differences.

Another factor that may have caused smaller differences in the efficacy rating may have been the probabilities of an outcome for each response on the DRH and VI schedules. These probabilities were found to be higher on the VI schedule than on the DRH schedule; this may have increased the causal efficacy ratings of the VI schedule and discriminative stimulus and decreased the causal efficacy ratings of the DRH schedule and discriminative stimulus, making the difference smaller than that in Experiment 1. In Experiment 3, we investigated whether this is the case by equating the outcome probabilities of each schedule to see whether there are still differences in causal efficacy ratings.

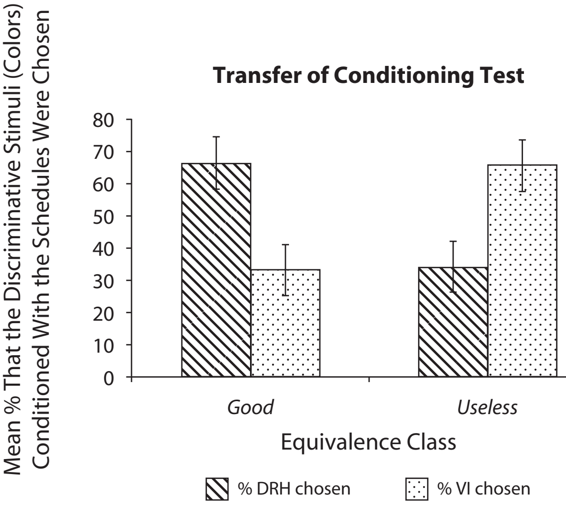

Figure 7. Results from Experiment 2. Mean percentages that the stimuli (colored circles) conditioned with each schedule were chosen with stimuli from each equivalence class (Class 1, *useless*; Class 2, *good*). Error bars represent the mean standard error. DRH, differential reinforcement of high rate; VI, variable interval.

EXPERIMENT 3

In Experiments 1 and 2, causal efficacy ratings followed the same pattern as response rates (high response rates predicted high causal efficacy ratings). In both of these experiments, the effect of the probability of producing an outcome given a response was overshadowed by other aspects of the schedule, even though, in previous research, this factor has been found to be important in controlling causal efficacy judgments; the greater the outcome probability is given a response, the greater the causal efficacy rating is (Allan, 1980; Reed, 2001a). Nevertheless, this potentially important factor was not directly controlled for in the preceding two experiments, and so in Experiment 3, we examined the effect of two schedules when the two probabilities of an outcome given a response was equated. Under these conditions, it might be expected that ratings of causal efficacy and their transfer of function will follow the rate of response.

Method

Participants

Twenty-nine adults participated in the present experiment (6 male, 23 female). The ages of the participants ranged from 18 to 51 years, with a mean of 24.79 years ($SD = 7.61$). All of the participants were either undergraduate or postgraduate students at Swansea University. All of the participants were recruited by advertisements in the psychology department, and they earned subject pool credit for their participation. Each participant had normal or corrected-to-normal eyesight and did not suffer from color blindness. All were naive to the purpose of the experiment.

Apparatus and Materials

The apparatus and materials were the same as those in Experiment 1, apart from in the evaluative conditioning task. The only differences in this task were again that different combinations of colored circles were used to signal the two scheduling conditions and that these were VR and VI schedules in this experiment. Four different combinations were used (yellow/red, yellow/blue, blue/green, and green/yellow) and were counterbalanced across participants to make sure any differences in ratings in the transfer test were not due to a preexisting preference for a particular color.

Procedure

Each participant was taken into a quiet room and given a consent form to read and sign and the BDI test before completing the study.

Derived stimulus relations training and testing. The derived stimulus relations training and testing were identical to those of Experiments 1 and 2.

Evaluative conditioning task. The participants were presented with the following instructions on the computer screen:

> You will shortly see a circle in the centre of the screen. The circle will be colored either color 1 or color 2. [See note 1.] When you see this circle you must press the space bar to earn points. Try to earn as many points as possible. You may press either quickly or slowly in order to earn points. You must work out what rate of pressing makes you the most points. Click here to proceed.

After presentation of these instructions, the participants were exposed to eight schedule sessions, which consisted of two alternating schedules (VI and VR). Each schedule was therefore presented four times to each participant. The VI schedule was always presented first and lasted for 3 min. Each VR schedule lasted until the participants had received approximately the same amount of points as they had in the previous VI schedule. Each schedule was signaled by the presence of a different-colored circle (discriminative stimulus). These colors were counterbalanced across participants, with four different combinations (yellow/blue, yellow/red, blue/green, and green/yellow).

In the first schedule, the participants responded on a VI 20-sec schedule. In this schedule, an outcome (point) was received after a response on a space bar, using a formula in which the responses were distributed around an average of 20 sec, with a range of 1 to 39 sec. In a second schedule, the participants responded on a VR schedule, yoked to the probability of an outcome given a response obtained on the VI schedule. This was done by dividing the total number of outcomes (points) per session on the VI schedule by the total number of responses. This number was then used as the overall mean probability of receiving an outcome (point) given a response on the yoked VR schedule. Points were received with a probability that varied around this mean value. After each schedule, the participants were asked to make a judgment of their causal efficacy in that session. The following instructions were presented:

> 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.

After the participant had chosen a number, the next schedule was presented, until all eight sessions were completed.

Transfer of conditioning test. The transfer of conditioning test was identical to that in Experiment 1.

Overall, 15 of the participants received the derived stimulus relations task before the evaluative conditioning task, and the other 14 participants received the tasks in the opposite order. All of the participants completed the transfer of conditioning test last. At the end of the experiment, each participant completed the transfer of function questionnaire. They were then thanked for participating and were debriefed.

Results and Discussion

Before the analyses were performed, 7 participants were excluded for high scores on the BDI (>10) (see Reed et al., 2001). The mean score for the remaining participants on the BDI was 4.73 ($SD = 3.24$).

Evaluative Conditioning Task

The mean rate of response to the schedules is shown in Figure 8A. These data show that the rate of response was lower on the VI than on the VR schedule. A two-factor ANOVA was performed on the data with schedule (VI vs. VR) and session as within-subjects factors. This revealed a significant main effect of session [$F(3,63) = 3.18$] and an effect of schedule [$F(1,21) = 28.28$]. No significant interaction was found between session and schedule [$F(3,63) = 1.89$].

The mean judgments of causal efficacy during each exposure to the two schedules are shown in Figure 8B. These data show that ratings of causal efficacy were higher on the VR than on the VI schedule. A two-factor, repeated measures ANOVA was performed on these data, with schedule (VI vs. VR) and session as within-subjects factors. This analysis revealed a significant effect of schedule [$F(1,21) = 27.82$] but no significant main effect of session [$F(3,63) = 1.81$]. No significant interaction was found between session and schedule ($F < 1$).

The number of outcomes obtained per minute across all four sessions was greater on the VR (2.75 points) than on the VI (2.73 points) schedule. This difference, however, was not significant [$t(21) = 1.00$]. The probability of an

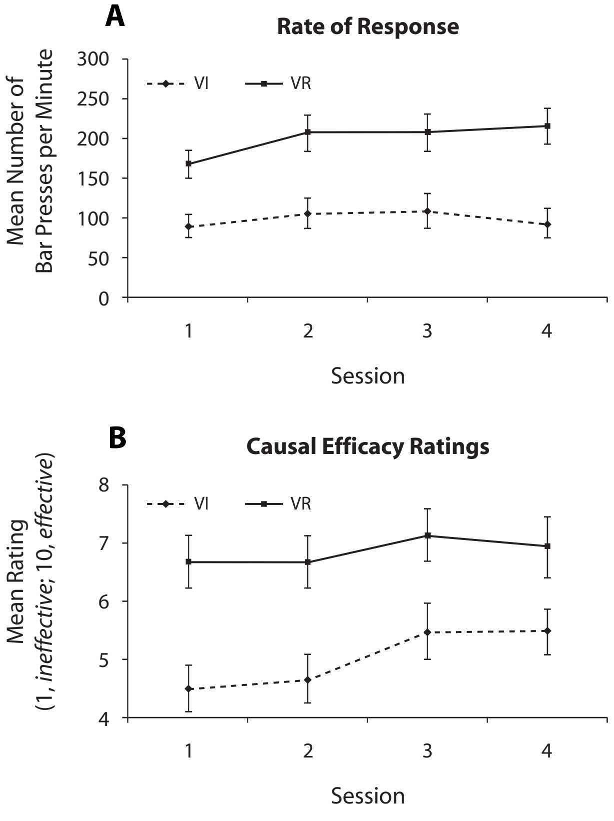

Figure 8. Results from Experiment 3. (A) Mean responses per minute across each session for each schedule. (B) Mean ratings of causal efficacy over the four sessions to each schedule. Error bars represent the mean standard error. VI, variable interval; VR, variable ratio.

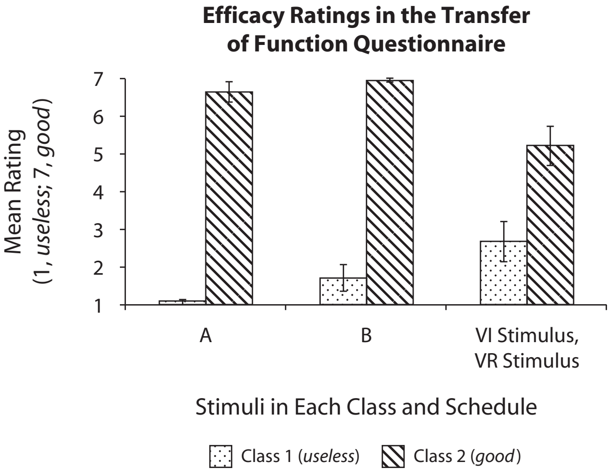

Figure 9. Results from Experiment 3. Mean rating for the stimuli in each class (Class 1, *useless*; Class 2, *good*) and the discriminative stimuli that signaled each schedule. Error bars represent the mean standard error. VI, variable internal; VR, variable ratio.

outcome was necessarily the same in the VI (.08) and VR (.08) schedules.

Derived Stimulus Relations Testing

All participants passed the emergent relations test with an overall mean of 98.61% (SD = 3.40) class-consistent responding. This number is significantly more correct responses than would be expected by chance [$t(21) = 67.02$]. Three participants needed a second training phase before reaching the test criterion (i.e., 83% correct).

Transfer of Function Questionnaire

The mean ratings (*useless–neutral–good*) for each nonsense stimuli in each class and the two discriminative stimuli that signaled the different schedules can be seen in Figure 9. These data show that lower ratings (more *useless*) were made for Class 1 than for Class 2 stimuli. These differences were found to be significant for each cross-class pair [A1–A2, $t(21) = 20.02$; A1–B2, $t(21) = 58.82$; B1–A2, $t(21) = 8.45$; B1–B2, $t(21) = 14.52$]. The discriminative stimulus that signaled the VI schedule (low response rates and low ratings of causal efficacy) was given lower ratings than the discriminative stimulus that signaled the VR schedule (high response rates and high ratings of causal efficacy). This difference was also significant [$t(21) = 2.45$].

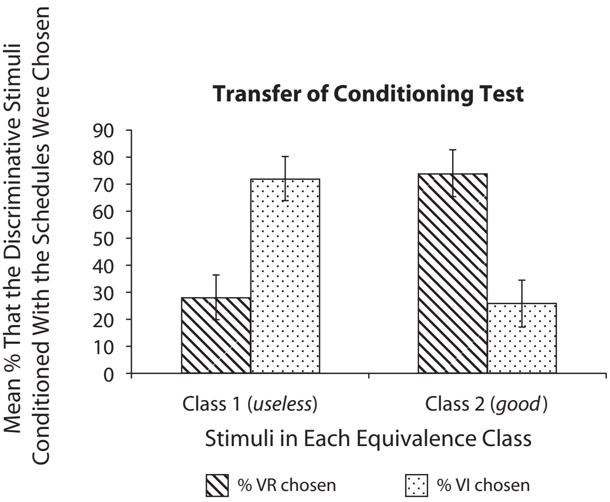

Figure 10. Results from Experiment 3. Mean percentages that the stimuli (colored circles) conditioned with each schedule were chosen with stimuli from each equivalence class (Class 1, *useless*; Class 2, *good*). Error bars represent the mean standard error. VI, variable interval; VR, variable ratio.

Transfer of Conditioning Test

Figure 10 reveals that the participants chose the discriminative stimulus associated with the high causal efficacy schedule (VR) in the presence of the stimuli in the *good* equivalence class significantly more than chance and significantly more than they chose the discriminative stimulus associated with the VI schedule with the members of *good* equivalence class [$t(21) = 2.79$].

In summary, response rates were higher on the VR than on the VI schedule, and the ratings of causal efficacy given to responses emitted on the VR schedule were higher than those given to responses emitted on the VI schedules. This result was obtained despite the equated probabilities of an outcome following a response on the two schedules, suggesting that this is not as important a factor as reinforcement or response rate in determining causal efficacy ratings. The findings also replicated those of Experiments 1 and 2: Evaluative learning took place; the discriminative stimuli present during each of the schedules of reinforcement were later rated according to the causal efficacy ratings given to the associated schedules. Again, it is still unclear exactly which factors control the rating of the discriminative stimuli. It could be differences in reinforcement rate, response rate, or causal efficacy ratings between the two schedules of reinforcement. Further research is needed to determine the processes involved. The participants also showed that they would categorize these discriminative stimuli with existing equivalence classes with the same function (i.e., *good* or *useless*), despite the fact they had never previously been directly related.

GENERAL DISCUSSION

In the present series of experiments, we investigated the manner in which various schedules of reinforcement produce ratings of causal efficacy and whether these ratings would transfer to a stimulus presented with each schedule in an evaluative learning paradigm. In addition, we aimed to demonstrate whether these stimuli would join established equivalence classes, through derived stimulus networks, with the same evaluative function, even though these stimuli had not been directly associated with either of the schedules.

The results of all three experiments showed a general trend of evaluative learning in the schedules task. The discriminative stimuli paired with each schedule were shown to acquire the valence of the participants' causal efficacy ratings. The stimuli were then categorized, through derived relational networks, with either the *good* equivalence class or the *useless* equivalence class in accordance with their evaluative learning. The participation of new stimuli in derived stimulus networks could be a possible mechanism behind the acquisition of evaluative judgments without direct experience of the event.

This series of experiments replicated previous research (Reed, 1994, 1999, 2001a, 2001b, 2003) in which causal efficacy ratings mirrored the pattern of rates of response, even when the rates of reinforcement (Experiment 2) and the probability of an outcome given a response (Experiment 3) were equated. That is, when response rates were high, ratings of causal efficacy were also high, and when response rates were low, ratings of causal efficacy were low. It may be that when the schedule is based on a strong response-based rule, participants are more sensitive to the relationship between the rate of response and the rate of reinforcement (they follow the molar view; Baum, 1973). On a DRH or VR schedule, this relationship is linear: The participants have to respond fast for reinforcement to be delivered. The strong relationship between responding and reinforcement (the outcome) could also lead to strong ratings of causal efficacy being made on this schedule. In contrast, once responding has reached a specified level on the DRL (in Experiment 1) and VI (in Experiment 2 and 3) schedules, increases in the rate of responses are not connected to increases in the rate of reinforcement. This may lead to low rates of responding and to lower causal efficacy ratings.

An alternative view is the molecular theory, which emphasizes the effect of reinforcement on the immediately preceding pattern of responding (see Peele, Casey, & Silberberg, 1984). On a DRL or VI schedule, as the pause from the last response becomes longer, the more likely it is that an outcome will be delivered for the next response. Therefore, long interresponse times are reinforced. However, on a DRH or VR schedule, there is no such favoring of long interresponse times, because reinforcement is just as likely to follow a burst of responding as to follow a long pause of responding. From this, it may be the local context of responding at the time of reinforcement that is the determinant of response rates, causal judgments, and the valence of the associated discriminative stimulus. On the DRH (Experiment 1 and 2) and VR (Experiment 3) schedules on which high rates of response were emitted just before reinforcement, high rates of response, high ratings of causal efficacy, and positive ratings of the associated stimulus (color) were produced. In contrast, on the DRL (Experiment 1) and VI (Experiment 2 and 3) schedules on which low rates of response were required prior to reinforcement, low rates of response, low ratings of causal efficacy, and negative ratings of the associated stimulus were found. According to this argument, to produce high ratings of causal efficacy, the participants have to respond at a high rate just before reinforcement (DRH, VR). One possibile reason that this molecular aspect impacts ratings of causal effectiveness is that the high salience of the response (consisting, in the case of VR and DRH schedules, of a group of responses emitted in close temporal proximity) causes the participants to rate this aspect of the environment more strongly than when the response is less salient (one response emitted in temporal isolation).

In the present article, the data suggest that it is the molar relationship between responding and reinforcement that had a more significant role in determining the causal efficacy ratings. This is shown in Experiment 2, in which the differences in causal efficacy ratings were much smaller between each schedule of reinforcement, and the difference between the ratings of the discriminative stimuli associated with each schedule was found to approach significance only after we equated the rates of reinforcement. This result suggests that the molar reinforcement rate does play a role in determining causal efficacy ratings and discriminative stimuli associated with each schedule. However, one potential reason for the molar control over causal efficacy ratings may have been that the experimental instructions encouraged the participants to vary their responding. Reed (2007) showed that, in rats, more variance between the animals' rates of responding tended to generate sensitivity to the molar characteristics of the schedule that they were exposed to. However, the inclusion of verbal instructions may mean that the observed differences in response rates were rule governed rather than contingency shaped. One way to test this might be to allow the schedules themselves to generate different rates of responding by repeating the procedure without any instructions regarding response rate. This was not done in this study, because a lot of researchers have found that humans can be insensitive to contingencies of reinforcement in terms of responding in a way that is characteristic of nonhumans. Baron, Kaufman, and Stauber (1969) found that participants who were given instructions concerning contingencies were more likely to respond in a schedule-typical manner than were uninstructed participants. Raia, Shillingford, Miller, and Baier (2000) also found that participants who were given accurate instructions about schedule properties were more likely to produce response patterns typical of nonhumans than were participants that were given inaccurate instructions. They also found that when instructions were lean (relative to rich or no instructions at all), this led to behavior that was most responsive to the reinforcement contingencies of the schedule. Because we were concerned with the transfer and generalization of schedule properties in the present experiments, instructions were included to ensure that the participants would respond in a typical pattern.

In future research, it may be interesting to further investigate the differences between evaluative conditioning in humans and nonhumans. The evaluative conditioning task used in the present article has some similarities to the procedures that have been employed to assess conditioned reinforcement value in nonhumans. In these studies, neutral stimuli may become secondary reinforcers if they are presented along with a primary reinforcer, such as food. The variables that determine the value of these secondary reinforcers have been widely investigated, with rate of reinforcement playing an important role (Herrnstein, 1964). However, Nevin, Grace, Holland, and McLean (2001) showed that, in pigeons, the preference for the secondary reinforcer depended, in part, on the response rate, as well as on the reinforcer rate. In their subjects, they also showed a preference for a VI schedule over a VR schedule, even when the reinforcer rates were equal between the two. This is the opposite of the findings in the present Experiment 2, in which the participants showed a preference for the DRH schedule over the VI schedule, despite equal rates of reinforcement. Such issues require further investigation.

All three of the present experiments replicated previous evaluative conditioning studies of humans (Baeyens et al., 1990; Zellner et al., 1983), showing that the valence of the discriminative stimuli became altered by the valence of the schedules of reinforcement. Stimuli paired with schedules that produced high ratings of causal efficacy (i.e., DRH in Experiments 1 and 2 and VR in Experiment 3) were rated as *good*, whereas stimuli paired with schedules that produced lower ratings of causal efficacy (DRL in Experiment 1 and VI in Experiments 2 and 3) were rated as *useless*. This finding shows that evaluations of efficacy in a task can transfer to stimuli present during that task. This may model how certain stimuli in the environment come to elicit feelings of hopelessness and helplessness in patients suffering from depression. In the transfer of conditioning tests, it was then shown that the discriminative stimuli associated with the task and the resulting ratings of causal efficacy would join established stimulus equivalence classes, through derived stimulus networks, with the same evaluative function. This is one possible learning-based approach to explaining the development of generalized evaluations (such as effectiveness, helplessness, etc.) without direct experience of each stimulus.

In addition to promoting basic understanding of evaluative learning in a new paradigm and of the factors in the environment that effect causal efficacy, we also explored the effects in the present study in a way that is potentially important to understanding applied contexts, such as psychotherapy. These results inform us about problems that can occur without direct experience (e.g., a fear of flying without ever being on a plane). As is shown in this study, if evaluations of causal efficacy can transfer to objects or people present during an event and these objects or people are closely related to the self, they could cause a negative cascade that would aggravate disorders such as depression or anxiety (Walther et al., 2005). In the present set of experiments, the stimuli associated with either good or bad causal evaluations were shown to be categorized with related stimuli through derived stimulus networks, which could model the processes in disorders, such as depression, in which negative evaluations spread to many aspects (objects, people, events) of a person's life without direct experience.

Future studies could incorporate stricter tests of equivalence and transfer of function. First, participants could be exposed to the test for emergent relations at the end of the experimental session, as opposed to directly after training, as in previous research (Staunton & Barnes-Holmes, 2004). This would ensure that the participants had no exposure to any of the untrained relations before the transfer of function questionnaire or transfer of conditioning test. Second, it may be interesting to examine relations between the discriminative stimuli and stimuli only in equivalence (as opposed to reflexivity or symmetry) relations with the evaluative descriptions. Finally, the transfer of function questionnaire could be strengthened by the participants' rating all of the stimuli before the derived relations training and evaluative conditioning task in order to show that all of the stimuli were neutral and similar until they were put into derived relations with the evaluative descriptions and the causal efficacy ratings.

Future researchers could also examine the effects of varying the instructions given to the participants when they rated their causal efficacy. Instead of rating how effective they felt at getting points, they could be asked how much control they felt that they had in producing points. If similar results were found using words such as *effective* and *control*, we could confidently relate efficacy judgments to perception of control judgments. It may also be useful to introduce a rating system that does not involve numbers per se, because crude numerics may encourage the participants to rank the discriminative stimuli as better or worse. Instead, a visual analogue scale could be used.

Further research is needed replicating these findings using a clinical population. Alloy and Abramson (1979) found differences in judgments of contingencies between responses and outcomes between depressed and nondepressed students. The depressed participants were found to judge the contingencies accurately, whereas the nondepressed participants overestimated the degree of contingency between their responses and outcomes. These results may be replicated in depressed and nondepressed participants' ratings of causal efficacy, where higher ratings may be found in a nondepressed population than in a depressed population. Further investigation of the role that contingency awareness may play in producing the differences seen between participants' ratings and categorization in the present study is needed. Individual differences have been found in how participants weight information in causal learning and decision making tasks (Osman & Shanks, 2005). These differences may be similar to those in the present study that affected which aspect of the schedule (causal efficacy, response rate, probability of an outcome given a response) transferred to the stimuli presented during the task and later generalized through derived stimulus networks, but further research is needed to test this prediction.

In conclusion, this was the first series of experiments in which the properties of schedules of reinforcement (causal efficacy ratings) in an evaluative learning paradigm were used and that showed that such properties could transfer to novel stimuli present during a schedules task. The stimuli signaling each schedule were then shown to join equivalence classes with the same evaluative function through derived relational networks, despite the fact that these classes were never directly related to the schedules task. This finding is important for the understanding of clinical disorders and of how problems can occur without direct experience, such as those observed in anxiety disorders, phobias, and depression.

AUTHOR NOTE

The research described here was supported by a research grant from the United Kingdom Economic and Social Research Council. Correspondence concerning this article should be addressed to C. Dack, Department of Psychology, Swansea University, Singleton Park, Swansea, SA2 8PP Wales (e-mail: charlottedack@hotmail.com).

REFERENCES

ALLAN, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, **15**, 147-149.

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

BAEYENS, F., EELEN, P., VAN DEN BERGH, O., & CROMBEZ, G. (1990). Flavor–flavor and color–flavor conditioning in humans. *Learning & Motivation*, **21**, 434-455. doi:10.1016/0023-9690(90)90025-J

BARON, A., KAUFMAN, A., & STAUBER, K. A. (1969). Effects of instructions and reinforcement feedback on human operant behavior maintained by fixed-interval reinforcement. *Journal of the Experimental Analysis of Behavior*, **12**, 701-712. doi:10.1901/jeab.1969.12-701

BAUM, W. M. (1973). The correlational-based law of effect. *Journal of the Experimental Analysis of Behavior*, **20**, 137-153. doi:10.1901/jeab.1973.20-137

BECK, A. T. (1978). *The depression inventory*. Philadelphia: Center for Cognitive Therapy.

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.

BUSH, K. M., SIDMAN, M., & DE ROSE, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior*, **51**, 29-45. doi:10.1901/jeab.1989.51-29

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.

DYMOND, S., & REHFELDT, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *Behavior Analyst*, **23**, 239-254.

FULCHER, E. P., MATHEWS, A., MACKINTOSH, B., & LAW, S. (2001). Evaluative learning and the allocation of attention to emotional stimuli. *Cognitive Therapy & Research*, **25**, 261-280. doi:10.1023/A:1010732328104

HERMANS, D. (1998). Evaluative conditioning. *Gedragstherapie*, **31**, 3-6.

HERRNSTEIN, R. J. (1964). Secondary reinforcement and rate of primary reinforcement. *Journal of the Experimental Analysis of Behavior*, **7**, 27-36. doi:10.1901/jeab.1964.7-27

LAZARUS, R. S. (1991). *Emotion and adaption*. New York: Oxford University Press.

LEVEY, A. B., & MARTIN, I. (1975). Classical conditioning of human "evaluative" responses. *Behaviour Research & Therapy*, **13**, 221-226. doi:10.1016/0005-7967(75)90026-1

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 & Adolescent Psychology*, **35**, 292-301.

NEVIN, J. A., GRACE, R. C., HOLLAND, S., & MCLEAN, A. P. (2001). Variable-ratio versus variable-interval schedules: Response rate, resistance to change, and preference. *Journal of the Experimental Analysis of Behavior*, **76**, 43-74. doi:10.1901/jeab.2001.76-43

OSMAN, M., & SHANKS, D. R. (2005). Individual differences in causal learning and decision making. *Acta Psychologica*, **120**, 93-112. doi:10.1016/j.actpsy.2005.04.003

PEELE, D. B., CASEY, J., & SILBERBERG, A. (1984). Primacy of interresponse-time reinforcement in accounting for rates under variable-ratio and variable-interval schedules. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**, 149-167. doi:10.1037/0097-7403.10.2.149

RAIA, C. P., SHILLINGFORD, S. W., MILLER, H. L., & BAIER, P. S. (2000). Interaction of procedural factors in human performance on yoked schedules. *Journal of the Experimental Analysis of Behavior*, **74**, 265-281. doi:10.1901/jeab.2000.74-265

REED, P. (1994). Influence of the cost of responding on human causality judgments. *Memory & Cognition*, **22**, 243-248.

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 & 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. (2003). Human causality judgments and response rates on DRL and DRH schedules of reinforcement. *Learning & Behavior*, **31**, 205-211.

REED, P. (2007). Response rate and sensitivity to the molar feedback function relating response and reinforcement rate on VI+ schedules of reinforcement. *Journal of Experimental Psychology: Animal Behavior Processes*, **33**, 428-439. doi:10.1037/0097-7403.33.4.428

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

ROZIN, P., MARKWITH, M., & MCCAULEY, C. (1994). Sensitivity to indirect contacts with other persons: 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

SCHERER, K. R. (1993). Neuroscience projections to current debates in emotion psychology. *Cognition & Emotion*, **7**, 1-41.

SIDMAN, M. (1971). Reading and auditory–visual equivalences. *Journal of Speech & 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, NJ: 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

STAUNTON, C., & BARNES-HOLMES, D. (2004). Relational frame theory, semantic priming and event related potentials: Building bridges between cognitive neuroscience and behavioural psychology. *Irish Psychologist*, **30**, 96-102.

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., NAGENGAST, B., & TRASSELLI, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition & Emotion*, **19**, 175-196. doi:10.1080/02699930441000274

ZAJONC, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, **35**, 151-175. doi:10.1037/0003-066X.35.2.151

ZELLNER, D. A., ROZIN, P., ARON, M., & KULISH, C. (1983). Conditioned enhancement of human's liking for flavor by pairing with sweetness. *Learning & Motivation*, **14**, 338-350. doi:10.1016/0023-9690(83)90021-8

NOTE

1. In the instructions that each participant saw, Color 1 and Color 2 represent whichever color combination that the participant was assigned in order to signal each schedule of reinforcement.

(Manuscript received October 17, 2008;
revision accepted for publication May 31, 2009.)