Published in final edited form as:

J Exp Psychol Anim Behav Process. 2010 April ; 36(2): 165–171. doi:10.1037/a0016484.

Extracting Functional Equivalence from Reversing Contingencies

Mimi Liljeholm and Bernard W. Balleine

Department of Psychology University of California, Los Angeles

In two experiments assessing acquired equivalence, human participants were initially presented with four cues, two of which were paired with one outcome and two of which were paired with a second outcome. These contingencies were then reversed across several training blocks such that, while each cue was paired equally often with each of the two outcomes *across* blocks, cues A and B always signaled the same outcome *within* blocks (as did cues C and D). In both experiments, performance on a subsequent transfer discrimination was enhanced when participants were required to generalize between stimuli that had been paired with the same outcome within each block of training. Additional tests did not yield evidence of a bias towards a specific set of cue-outcome contingencies in either experiment. Moreover, interviews conducted at the end of Experiment 2 revealed that performance on the transfer discrimination was enhanced only in participants that discovered the equivalence relationships during initial training. The results challenge simple associative, as well as attentional, accounts of acquired equivalence and favor the view that this effect is mediated by comparisons of the similarity of adjacent cue-outcome structures.

How do humans and other animals represent the similarity of stimuli that share functional, rather than physical, properties? It is well known that pairing distinct cues with a common outcome enhances subsequent generalization between them; an effect referred to as *acquired equivalence*. Thus, if two cues (A and B) are paired with the same outcome and a third cue (C) is paired with a different outcome (i.e., $A \rightarrow O1$, $B \rightarrow O1$, $C \rightarrow O2$), any novel treatment subsequently applied to A (e.g. $A \rightarrow O3$) tends to generalize more to cue B than to cue C. This effect has been demonstrated in both human (e.g., Hall, Mitchell, Graham & Lavis, 2003) and non-primate animals (e.g., Honey & Hall, 1989) and has inspired a number of associative theories of the underlying representations (e.g., Hall, 1991, 1996).

For example, Honey & Hall (1989; Hall et al., 2003) have argued that initial training may result in the associative retrieval of O1 by both cues A and B; on $A \rightarrow O3$ trials, the associatively retrieved representation of O1 may then become associated with O3 (i.e., mediated conditioning), providing a simple means for cue B, but not cue C, to activate a representation of O3 on test. Alternatively, Hall (1991) discussed the possibility that initial training results in a dynamic configuration of the features of both cues and outcomes (i.e., cues A, B and C come to respectively activate representations AO1, BO1 and CO2), suggesting that subsequent generalization might reflect Pearce's (1987) rule, which specifies that the similarity between configural cues is determined by the product of the ratio of common elements to total elements for each cue.

Correspondence: Mimi Liljeholm Department of Psychology, UCLA Box 951563 Los Angeles, CA 90095-1563 Tel: 310 825-8712 Fax: 310 206-5895 mlil@ucla.edu .

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/xan.

Note that, although these associative theories differ in several respects, they share the basic premise that acquired equivalence depends on the extent to which cues activate overlapping representations of the sensory, or motivational, features of their common outcome. From this perspective, if all cues were paired equally often with each outcome, and so became equally associated or configured with each outcome representation, the acquired equivalence effect should be abolished. We assessed this prediction in two experiments using human participants.

Experiment 1

The design of Experiment 1 was based on a study by Hall et al. (2003; Exp 1), in which human participants received initial training with four visual cues; two of which (A & B) were followed by a red rectangle (O1) and two (C & D) by a green rectangle (O2). Hall et al. (2003) found that a subsequent transfer discrimination, in which the previously trained cues signaled which of two motor responses (R1 vs. R2) should be performed, was more readily acquired when cues that had been paired with the same outcome in the first phase required the same motor response (i.e., acquired equivalence). Moreover, when an outcome (e.g., O1) had been shared, during initial training, by cues that signaled the same motor response (e.g., R1) in the subsequent transfer discrimination, O1 was capable of eliciting R1 in a final test phase.

Hall et al. (2003) interpreted the results from their final test as evidence for mediated conditioning and attributed performance on the transfer discrimination to the development of an excitatory association between the, associatively retrieved, representation of O1 and R1. Note, however, that although consistent with a mediated conditioning account, the ability of O1 to elicit R1 does not provide conclusive evidence of a direct association between these two events. Alternatively, in the final test, O1 may have retrieved cues A and B and these retrieved cue representations may then have elicited R1 (note that this account eliminates the need to assume that associatively retrieved events are learned about as if though they were present).

Whether this particular scenario is the case or not, the possibility remains that performance on the transfer discrimination was independent of the processes underlying response selections in the final test. To assess this possibility, in the current study, we attempted to eliminate retrieval based on specific cue-outcome contingencies, while preserving the equivalence relationships between cues.

As in Hall et al.'s experiment, the initial training phase involved contingency learning with four visual cues (see Figure 1) and two outcomes (a red and a green colored square). However, in the current study, this first phase was divided into 40 training blocks with each of the cues presented only once in each block; in half of the blocks (Type 1), cues A and B were followed by O1 and cues C and D were followed by O2, whereas in the remaining blocks (Type 2) these contingencies were reversed such that cues A and B signaled O2 and cues C and D signaled O1. All block boundaries were explicitly delineated by a screen instructing participants to "Press the space-bar to continue to the next block", and the order of the two types of blocks was randomized. Although participants were not required to make any overt responses during this phase, the task, as described to them, was to "figure out which (outcome) is signaled by which cue in each block".

Given the many rapid reversals occurring across blocks throughout this first phase, it is unlikely that participants selectively formed associations, or configurations, based on the contingencies presented in one but not the other type of block; in other words, each cue should have been equally associated, or configured, with each outcome representation by the end of training. Thus, if acquired equivalence depends on the extent to which cues activate overlapping representations of outcomes with which they were previously paired, it should not emerge using this procedure. However, *within* each training block, cues A and B always signaled the *same* outcome (whereas the other outcome was common to cues C and D). Indeed, except at

the very point of a reversal, this equivalence held true throughout training for any comparison of the similarity of a given cue-outcome structure (e.g., $A \rightarrow O1$) with the cue-outcome structures occurring on adjacent trials. Thus, on the basis of such local comparisons, one might predict that the representation of functional equivalence should remain relatively intact across reversals.

Note that, while the equivalence sometimes holds true even for comparisons of cue-outcome structures that straddle block boundaries (since the reversals occur randomly across blocks), it *always* holds true for within-block comparisons. The explicit delineation of blocks and the instructions to predict outcomes based on cues “in each block” were included to encourage participants to focus on such comparisons. Of course, this verbal emphasis on events occurring within blocks reveals nothing about the specific contingencies or about cue equivalence. It would be a mistake, therefore, to attribute any obtained equivalence effect to verbal instructions. This distinction is particularly important in light of recent efforts to demonstrate the propositional nature of acquired equivalence by generating the effect using verbal instructions alone (i.e., Smyth, Barnes-Holmes, & Barnes-Holmes, 2008).

Equivalence was assessed using a transfer discrimination test, again based on Hall et al.’s (2003) study; specifically, for half of the participants (Group Consistent), cues that had been paired with the same outcome within each block of phase 1 were trained with the same motor response, whereas for the other half (Group Inconsistent), cues that shared an outcome in each block of phase 1 were trained with different motor responses. If a representation of equivalence emerged as a result of phase 1 training, then it should facilitate acquisition of the transfer discrimination in Group Consistent relative to Group Inconsistent (for the sake of brevity, we will refer to a superior performance by Group Consistent as facilitation, but we note that such results may also reflect interference in Group Inconsistent). Alternatively, if equivalence was abolished by the reversal of contingencies, then there is no basis for predicting a difference between groups on their acquisition of the transfer discrimination.

To ensure that the rapid reversals had indeed eliminated associative, or configural, retrieval specific to the contingencies presented in one but not the other type of block, and to detect the presence of any recency effects (i.e., selective recall of the most recently experienced contingencies), the final test of Hall et al.’s study, in which participants were presented with each of the two outcomes and were required to choose between the two responses established during the transfer discrimination, was also included in the current experiment. In terms of a mediated conditioning account, if, for example, the contingencies from the last block of initial training dominated, such that each outcome representation was selectively activated by a subset of cues (e.g., $O1$ was activated only by cues A and B) then, in Group Consistent, $O1$ should come to elicit the response required to A and B during the transfer discrimination. In contrast, in Group Inconsistent, cues A and B require different responses, and so $O1$ should be as likely to elicit one response as the other. Consequently, the two groups should differ on their response selection in the final test.

Method

Participants—Thirty-two undergraduates at the University of California, Los Angeles (UCLA), participated to obtain credit in an Introductory Psychology course

Procedure—The four visual cues are illustrated in Figure 1. The assignment of visual stimuli to the respective cues (i.e., A , B , C and D) was semi-counterbalanced as listed in Figure 1. Each visual cue was presented once in each of the 40 blocks making up the first phase, for a total of 160 trials. Cues remained on the screen for 1000 ms and were immediately followed by the colored squares. Each colored square remained on the screen for 1000 ms and was then

followed by a 1500 ms duration screen, instructing participants to “Wait for the next cue”. At the beginning of the experiment, participants were presented with the following instructions:

“In the first phase of this experiment we want you to learn to predict which of two different colored squares is signaled by which of four distinct cues. In each block of trials, each of the four cues will be presented and followed by either a red or a green colored square. After viewing four consecutive trials, one with each cue, you will be asked to press any key to move on to the next block of trials. Your task is to figure out which color, red or green, is signaled by which cue in each block. In the beginning you will have to guess, but eventually you will learn to correctly predict which color square will appear based on the given cue. You will not be required to indicate your predictions about what color square will appear given a particular cue in this phase, however it is extremely important that you still make these predictions, and pay close attention, on each trial. You will need to use this knowledge later on in the experiment”

Note that, because the reversals occurred randomly across the blocks, the explicitly delineated block boundaries could not serve as a contextual cue for the reversals. After completing phase 1 training, participants were presented with the following instructions:

“In the next phase, we want you to categorize the four different cues into one of two categories. The two categories correspond to the key labeled ‘2’ and the key labeled ‘8’ respectively. On each trial, one of the four cues will appear on the screen. Depending on which cue it is, you should press either the key labeled ‘8’ or the key labeled ‘2’. After pressing a particular key, you will receive feedback about whether you were right or wrong.”

Each of the four visual cues was presented 12 times during the categorization phase, for a total of 48 trials. At the end of the second phase, participants were presented with the following instructions:

“In the next phase we want you to categorize each of the two colored squares into one of the two categories. On each trial, a colored square will appear on the screen. Depending on which color it is, red or green, you should press either the key labeled ‘8’ or the key labeled ‘2’. You will NOT be given any feedback in this phase.”

Each of the two colored squares were presented 4 times in the third phase, in random order, and were separated by a screen instructing participants to press any key to continue to the next trial.

Results and Discussion

No data were collected during the first phase. The mean percent correct scores for both groups from the second phase are shown in Table 1. It is apparent that discrimination performance was facilitated in Group Consistent compared to Group Inconsistent, indicative of acquired equivalence. Statistical analysis confirmed this description. A Group (Consistent vs. Inconsistent) $\times$ Stimulus (counterbalancing of visual cues) analysis of variance performed on the percent correct scores revealed a main effect of Group, $F(1,28)=7.01$, $p<0.05$, $\eta^2=.20$, but neither an effect of the counterbalancing of stimuli, nor an interaction, F 's < 1.0 .

The results from the third phase were analyzed in terms of whether or not response selections were ‘appropriate’ given the cue-outcome pairings presented in the last block of phase 1. Specifically, if a participant in Group Consistent had experienced, say, $A \rightarrow O1$ and $B \rightarrow O1$ in the last block, then $R1$ (which was the response required to both cues A and B during the transfer discrimination in this group) would be the appropriate response to $O1$ in phase 3. For the analogous participant in Group Inconsistent (i.e., one that had experienced $A \rightarrow O1$ and $B \rightarrow O1$ trials in the last block) $R1$ was again coded as the appropriate response to $O1$: recall

that, since participants in this group had in fact been required to perform R2 in response to cue B, a mediated conditioning account predicts that the number of appropriate responses should be greater in the Consistent than the Inconsistent group.

The group means for appropriate responses made during the third phase are presented in Table 1. There was no consistent pattern in terms of the specific outcome-response decision; the mean “appropriate response” score was approximately 0.5 and did not differ between groups, $F < 1.0$. This suggests that participants’ selection of the first response to an outcome was random and was not biased by the contingencies presented in the last block of phase 1. More generally, the results indicate that acquired equivalence is independent of the extent to which cues associatively, or configurally, activate overlapping representations of common outcomes, and suggest that the effect might instead be mediated by comparisons of the similarity of adjacent cue-outcome structures.

It is possible however, that participants in Group Consistent initially recalled only the contingencies from the last block of phase 1 (resulting in a facilitation of the transfer discrimination due to mediated conditioning) but that, as phase 2 progressed, a tendency to recall the reversed set of contingencies induced *proactive interference* (e.g., Treichler, 2005), explaining why the group mean response selections from phase 3 were at chance level. To assess this possibility we divided the second phase into 4 blocks of 12 trials and performed a Group (Consistent vs. Inconsistent) $\times$ Block (1-4) analysis of variance on the percent correct scores. The analysis revealed that, although there were significantly fewer errors in the Consistent than the Inconsistent Group, and although errors decreased significantly across blocks ($p < 0.05$), there was no group by block interaction, $F < 1.0$.

Presumably, any account that attributes facilitated performance of the phase 2 discrimination by Group Consistent to selective recall of contingencies from the last block of phase 1 while, at the same time, attributing the chance performance by the same group during phase 3 to a failure to recall those same contingencies at some later stage in phase 2 would be somewhat challenged by the persistence of the group difference throughout the transfer discrimination. Nonetheless, to more thoroughly explore the merits of this alternative account, Experiment 2 provides a direct measure of the degree to which participants entered the transfer discrimination with a bias towards the most recently experienced contingencies and, further, assesses the relevance of this potential bias to any obtained equivalence effect.

Experiment 2

In order to assess whether recall of the cue-outcome contingencies during the early stage of the transfer discrimination was dominated by events presented in the last block of phase 1, half of the participants in Experiment 2 (Group Probe) received a probe test before the start of the transfer discrimination, in which they were asked to predict the outcome that would follow each cue by pressing one of two keys. After the probe test these participants continued on to phases 2 and 3 which were exactly as in Experiment 1 (half of the participants in Group Probe received the consistent discrimination whereas the other half received the inconsistent one). The remaining participants (Group No Probe) proceeded with phases 2 and 3 without receiving the probe test (again, these participants were randomly assigned to consistent and inconsistent conditions).

In addition to the probe test, in an effort to eliminate any potential recency bias and thus enable an assessment of acquired equivalence in the *absence* of selectively activated outcome representations, a brief, (3 min.) distracter task was presented to all groups at the end of phase 1. Of course, to the extent that the distracter task attenuated recency effects, it may also have enhanced proactive interference. More importantly, the presence of either or both memory biases does not rule out the possibility that participants’ discovery of the equivalence

relationships in phase 1 was the source of the difference between the consistent and inconsistent groups on the transfer discrimination; such biases would simply prevent us from ruling out alternative processes, such as mediated conditioning. Indeed, perhaps the most straightforward way to establish whether any obtained equivalence effect is due to discovery of the equivalence relationships during initial training is to simply ask participants if they discovered the relationships, and assign them to groups (i.e., "Discovery" and "No Discovery") according to their answers.

If superior performance by Group Consistent on the transfer discrimination is due to discovery of the stimulus relations, then the Consistent-No Discovery and Inconsistent-No Discovery groups should not differ on this measure, nor would one expect the Inconsistent-Discovery group to differ from either of the two No Discovery groups. One would, however, predict superior performance by the Consistent-Discovery group relative to each of the other three groups (note that these predictions presume that the proportion of participants who discover the relationships is roughly the same for the consistent and inconsistent groups). In contrast, if discrimination performance depends on mediation by specific outcome representations – a process that, as discussed, may be both facilitated and masked by the dynamics of memory – then the consistent and inconsistent groups should differ to the same extent in the Discovery and No Discovery conditions. Given these diverging predictions, at the end of Experiment 2, participants were asked if, during the first phase, they had discovered some rule that had allowed them to make accurate predictions about which color would follow which cue. To avoid any bias or confusion introduced by the probe test, only participants in Group No Probe were presented with this question.

Method

Participants—Ninety-two undergraduates at UCLA participated to obtain credit in an Introductory Psychology course.

Procedure—The instructions for the first phase were the same as in Experiment 1. At the end of the first phase, participants were provided with a single sheet of 20 simple algebra problems, the distracter task, and were told to complete as many of the problems as they could. After 3 minutes, the experimenter removed the sheet and instructed participants to press any key on the computer keyboard to return to the previous task. At this point, participants in Group Probe were presented with the following instructions:

"Now that you have completed the math problems we will return to the cues and colored squares from the first phase of the experiment. As before, your task is to predict which colored square is signaled by which cue. However, this time we want you to press a key to let us know what your prediction is. If you think that the cue presented on the screen should be followed by a red square, press 'R'; if you think the cue should be followed by a green square, press 'G'. You will NOT receive any feedback in this phase."

Each of the four cues was presented once, in random order. The cues remained on the screen until a response was given (i.e., the participant pressed either the "R" or "G" key) and each cue was followed by a 1000 ms duration screen instructing participants to "Wait for the next trial".

After this probe test, participants in the Probe group received the same instructions as those that preceded the phase 2 discrimination in Experiment 1. Participants in the No Probe group received these same instructions immediately after completing the distracter task. Phases 2 and 3 then proceeded as in Experiment 1, for both groups. At the end of the experiment, participants in the No Probe group were asked if, in the first phase, they had discovered some rule that had

allowed them to accurately predict which color would follow which cue. If a participants' answer did not reflect the cue equivalence, the experimenter described it (i.e., stated that there had been two pairs of cues and that cues within a pair signaled the same color in each block of phase 1, even though the specific color signaled by each cue reversed across blocks). The participant was then asked again if they had noticed the, now described, relationships during phase 1 training.

Results and Discussion

As in Experiment 1, no data were collected during the first phase. The responses from the probe test were coded in terms of whether (1) or not (0) they reflected recall of the contingencies presented in the last block of phase 1. The mean recall score was 0.57 and did not differ significantly from chance (i.e., 0.5), $t(45)=1.16$, $p=0.30$, nor did it differ significantly between the consistent and inconsistent groups, ($t < 1.0$), suggesting that participants did not enter the second phase discrimination with a bias towards the contingencies presented in the last block of phase 1.

The mean percent correct scores from the second phase and the mean 'appropriate' response selections from the third phase are shown for all groups in Table 2a. It is apparent that, as in Experiment 1, discrimination performance in Groups Consistent was superior to that in Groups Inconsistent, indicative of acquired equivalence. However, contrary to the findings from Experiment 1, the mean 'appropriate' responses appear to be lower in the consistent than in the inconsistent group. This is consistent with the possibility of mediated conditioning based on proactive interference (i.e., on recall specific to the set of contingencies *not* presented in the last block of phase 1).

A Probe $\times$ Consistency $\times$ Counterbalancing analysis of variance performed on the percent correct scores in phase 2 revealed a main effect of Consistency, $F(1,84)=6.42$, $p<0.05$, $\eta^2 = .07$, but no other significant main effects or interactions, F 's < 1.0 . In contrast, and as in Experiment 1, the same analysis performed on the phase 3 response selections did not yield a significant difference between the consistent and inconsistent groups, $F(1,84)=1.1$, $p=0.30$, nor were there any other significant differences, all F 's < 1.0 . Thus, the pattern of interest demonstrated in Experiment 1 (i.e., the emergence of an acquired equivalence effect in the absence of any reliable mediated conditioning effects) was replicated in Experiment 2.

Recall that, at the end of Experiment 2, participants in the No Probe group ($n=46$) were asked if, in the first phase, they had discovered some rule that had allowed them to accurately predict which color would follow which cue; 17 of 46 participants were able to generate a rule consistent with the cue equivalence from phase 1 when presented with this question. An additional 4 participants did not spontaneously generate any rule, but once prompted by the experimenters' description of the equivalence (see method section), they claimed to have had knowledge of it during phase 1 training (in subsequent analyses, these 4 participants were grouped together with those that spontaneously generated the equivalence answer). The remaining 25 participants did not spontaneously generate the equivalence rule, nor did they recognize it from the experimenters' description.

Importantly, the proportion of participants who discovered the equivalence was the same (46 %) in groups consistent and inconsistent, indicating that self-reports about the phase 1 relationships were unaffected by subsequent training. The 46 participants were divided into four groups (Consistent-Discovery; $n=11$, Inconsistent-Discovery; $n=10$, Consistent-No Discovery; $n=13$, and Inconsistent-No Discovery; $n=12$). The mean percent correct scores and mean 'appropriate' response selections for these four groups are presented in Table 2b. Interestingly, it appears that the equivalence effect (i.e., superior performance on the transfer discrimination by Group Consistent relative to Group Inconsistent) only emerged in the group

of participants who reported the equivalence, or claimed knowledge of it once described. Moreover, unlike performance on the transfer discrimination, the difference between the consistent and inconsistent groups on the phase 3 response selections did not appear to depend on discovery. Planned comparisons performed on the percent correct scores from the transfer discrimination revealed no significant difference between the Consistent-No Discovery and Inconsistent-No Discovery groups, nor did either of these two groups differ significantly from the Inconsistent-Discovery group (both, $t's < 1$). In contrast, the mean percent correct score in Group Consistent-Discovery differed significantly from that of the other three groups, $t(42) = 3.78$, $p < 0.001$. Again, the same analysis applied to the response selections from phase 3 revealed no significant differences, $t's < 1.0$. This pattern of results indicates that, at least with the procedures employed in the current series, the acquired equivalence effect is independent of processes involving mediation by specific outcome representations.

General Discussion

In two experiments assessing acquired equivalence, human participants were initially presented with four cues, two of which were paired with one outcome and two of which were paired with a second outcome (i.e., $A \rightarrow O1$, $B \rightarrow O1$, $C \rightarrow O2$, $D \rightarrow O2$). These contingencies were then reversed (i.e., $A \rightarrow O2$, $B \rightarrow O2$, $C \rightarrow O1$, $D \rightarrow O1$), and continued to reverse across a large number of randomly ordered, and explicitly delineated, training blocks, such that, whereas cues A and B always signaled the *same* outcome *within* each block (as did cues C and D), each cue was paired equally often with each of the two outcomes across blocks. In both experiments, performance on a subsequent transfer discrimination was enhanced when it required participants to perform the same motor response to stimuli that had been paired with the same outcome *within* each block of training (Group Consistent), relative to when those stimuli required different responses (Group Inconsistent), indicative of acquired equivalence.

Given the many rapid reversals occurring throughout the initial training phase, it is unlikely that participants selectively formed associations, or configurations, based on a specific set of contingencies and, consequently, theories that attribute acquired equivalence to mediation by shared outcome representations, whether associatively or configurally retrieved, cannot account for these results. Of course, this conclusion hinges on the assumption that the repeated reversals did indeed eliminate selective retrieval of common outcome representations. To verify this conjecture, both experiments included a final test phase in which participants were presented with the outcomes from the first phase and were required to choose between the two responses established during the transfer discrimination. Recall that the mediated conditioning account (e.g., Hall et al, 2003) attributes enhanced performance on the transfer discrimination to the retrieval of a common outcome by cues that require the same motor response, and the formation of an association between the retrieved outcome representation and the performed response. Consequently, to the extent that the Consistent and Inconsistent Groups differ on the transfer discrimination, they should also differ with respect to their response selections in the final test. Contrary to this prediction, and to the findings by Hall et al. (2003), neither of the two experiments in the current series yielded any such evidence for associative mediation.

Nonetheless, in order to assess more directly whether recall during the transfer discrimination was dominated by a specific set of contingencies, Experiment 2 included a probe test, administered before the start of the transfer discrimination, in which participants were asked to indicate their predictions of which outcome followed which cue. Performance on this probe test did not deviate significantly from chance, ruling out a systematic retrieval, or recall, of common outcomes at the outset of the transfer discrimination. Finally, interviews conducted at the end of Experiment 2 revealed that performance on the transfer discrimination was enhanced only amongst participants who discovered the equivalence relationships during initial training. Taken together these results provide converging evidence for the hypothesis

that, at least with the procedures employed in the current series, the acquired equivalence effect is independent of processes involving mediation by specific outcome representations. In the following paragraphs, we will consider the application of some alternative accounts to the current results.

One advantage of the configural view over that of associative mediation is that the former can be applied to more complex demonstrations of acquired equivalence, involving bi-conditional discriminations. Specifically, if, in the current experiments, additional stimulus features had reversed with the reversal of contingencies (i.e., $AX \rightarrow O1$, $BX \rightarrow O1$, $AY \rightarrow O2$, $BY \rightarrow O2$; $CX \rightarrow O2$, $DX \rightarrow O2$, $CY \rightarrow O1$, $DY \rightarrow O1$), then our results could be explained by the formation of configural representations that encode all of the events present on a given trial: For example, the presentation of A, X and O1 would result in the configural representation AXO1. Generalization would be greater between A and B than between A and D because the configural representations activated by A and B (e.g., AXO1 and BXO1) share more elements than do the configural representations activated by A and D (e.g., AXO1 and DY O1). Honey and Ward-Robinson (2002) demonstrated acquired equivalence in rats using this type of bi-conditional discrimination, and developed a connectionist implementation of Hall's (1991) configural account to explain their results. However, unless it is assumed that relations between stimuli can form part of stimulus representations, given the absence of any explicit conditional stimulus features (i.e., X and Y) in the current study, this connectionist analysis is as challenged by the current results as the associative theories previously described.

An important alternative approach to acquired equivalence is provided by attentional theories of animal discrimination learning (e.g., Lawrence, 1949; Mackintosh, 1975). The application of an attentional analysis to acquired equivalence starts with the assumption that some features are shared only by subsets of cues (e.g., that cues A, B, C and D are, respectively, composed of features aX, bX, cY, and dY). Bonardi, Graham, Hall and Mitchell (2007) presented human participants with two highly similar snowflake shapes and two red rectangles of slightly different luminance. In Group S, the two shapes (S1 and S2) were paired with one outcome (O1) whereas the two red colors (C1 and C2) were paired with a different outcome (O2). In Group D, S1 and C1 were paired with O1 and S2 and C2 were paired with O2. In a subsequent discrimination task in which participants were required to make one response (R1) to S1 and C2 and a different response (R2) to S2 and C1, the performance of Group D was superior to that of Group S. Bonardi et al. noted that, since S1 and C2 had been paired with different outcomes (as had S2 and C1) in *both* groups, simple associative mediation could not account for this difference. They argued instead that, in Group S, initial training allowed features common *only* to the two shapes, as well as features common *only* to the two colors, to command special attention, since those features were the best predictors of the outcome on each trial, and that this impaired performance on the subsequent discrimination. However, the stimuli used in the current experiments were counterbalanced, and constructed specifically to ensure that their features were either common to all cues or unique to each. In other words, the features shared by cues A and B were also shared by cues C and D. It is difficult, therefore, to see how our results could be explained by differential attention to shared stimulus features.

Alternatively, it is possible that the representation of cue equivalence is acquired across instances of similarity mappings between cue-outcome pairs occurring on adjacent trials (or within explicitly delineated blocks). We assume that these mappings are structured (Halford, 1993), entailing representations not just of objects (e.g., A and O1), but also of relations (e.g., *outcome-of*) and bindings of objects to relations (e.g., *outcome-of*(O1, A), with the positions of O1 and A specifying the direction of the relationship. From this perspective, if, say, the trials $A \rightarrow O1$ and $B \rightarrow O1$ are encountered within a block, the similarity of the structures *outcome-of*(O1, A) and *outcome-of*(O1, B) will be mapped, and the additional relation *same*(*outcome-of*-A, *outcome-of*-B) will be abstracted (see Kroger, Holyoak & Hummel, 2004). Then, if the

relation $A \rightarrow O2$ is encountered in a subsequent block, retrieval of the relational schema *same* (outcome-of-A, outcome-of-B) will create the expectancy that $B \rightarrow O2$ should also be observed. Finally, this relational structure will provide a basis for the difference between groups on the transfer discrimination; that is, *same*(outcome-of-A, outcome-of-B) will facilitate learning A-R1, B-R1 but interfere with learning A-R1, B-R2.

Although the abstraction of functional equivalence based on structured similarity mapping falls outside the scope of traditional associative and connectionist architectures, symbolic connectionist systems (e.g., Hummel & Holyoak, 1997, 2003) suggest how such a process could be implemented in a neural network. Recall that, in the statement *outcome-of*(O1, A), the positions of O1 and A specify the direction of the relationship; specifically that O1 is the outcome of A and that A is the predictor, or producer, of O1. According to the *LISA* model (Learning and Inference with Schemas and Analogies; Hummel & Holyoak, 1997; 2003), the binding of relational roles (e.g., *outcome*) to their objects (e.g., O1) is explicitly represented by synchronous neuronal firing, allowing for the mapping of relational structures in working memory, which in turn supports the formation and storage of conjunctive (cf. associative) codes in long-term memory. The dynamic binding, and resulting independence, of roles to their objects permits relational invariance and enables, for example, generalization of the relation *same*(outcome-of-A, outcome-of-B) across contingency reversals. Thus, while a complete account of relational representation goes beyond the scope of the current paper, we note that this approach may help advance our understanding of acquired equivalence, particularly with respect to the apparent outcome invariance demonstrated in the current series.

Acknowledgments

This research was supported by a grant from the NIMH # 56446 to BWB. We thank Keith Holyoak for his valuable comments.

References

- Bonardi C, Graham S, Hall G, Mitchell C. Acquired distinctiveness and equivalence in human discrimination learning: Evidence for an attentional process. *Psychonomic Bulletin & Review* 2005;12:88–92. [PubMed: 15945202]
- Doumas LAA, Hummel JE, Sandhofer CM. A theory of the discovery and predication of relational concepts. *Psychological Review* 2008;115:1–43. [PubMed: 18211183]
- Hall, G. Perceptual and associative learning. Clarendon Press; Oxford, England: 1991.
- Hall G. Learning about associatively activated stimulus representations : Implications for acquired equivalence and perceptual learning. *Animal Learning & Behavior* 1996;24:233–255.
- Hall G, Mitchell C, Graham S, Lavis Y. Acquired equivalence and distinctiveness in human discrimination learning: Evidence for associative mediation. *Journal of Experimental Psychology: General* 2003;132:266–276. [PubMed: 12825640]
- Halford, GS. Children's understanding: The development of mental models. Erlbaum; New York: 1993.
- Honey RC, Hall G. The acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology: Animal Behavior Processes* 1989;15:338–346. [PubMed: 2794870]
- Hummel JE, Holyoak KJ. A symbolic -connectionist theory of relational inference and generalization. *Psychological Review* 2003;110:220–264. [PubMed: 12747523]
- Hummel JE, Holyoak KJ. Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review* 1997;104:427–466.
- Kroger JK, Holyoak KJ, Hummel JE. Varieties of sameness: The impact of relational complexity on perceptual comparisons. *Cognitive Science* 2004;28:335–358.
- Lawrence DH. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology* 1949;39:770–784. [PubMed: 15398590]

- Mackintosh NJ. A theory of attention: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review* 1975;82:276–298.
- Pearce JM. A model for stimulus generalization in Pavlovian conditioning. *Psychological Review* 1987;94:61–73. [PubMed: 3823305]
- Smyth S, Barnes-Holmes D, Barnes-Holmes Y. Acquired equivalence in human discrimination learning: The role of propositional knowledge. *Journal of Experimental Psychology: Animal Behavior Processes* 2008;34:167–177. [PubMed: 18248123]
- Treichler FR. Successive reversal of concurrent discriminations by macaques (*Macaca mulatta*): proactive interference effects. *Animal Cognition* 2005;12:92–96.

Figure 1.

The visual stimuli used as cues in Experiments 1 and 2, with the counterbalancing of the assignment of visual stimuli to cues A, B, C and D respectively listed below.

Table 1

Mean percent correct scores on the transfer discrimination, and mean 'appropriate' response selections from phase 3 of Experiment 1, with standard errors shown in parenthesis.

Groups	Phase 2	Phase 3
Consistent	82.81 (3.6)	50.78 (12.0)
Inconsistent	66.56 (4.8)	51.56 (11.7)

Table 2

Results from Experiment 2: Mean percent correct scores on the transfer discrimination and mean 'appropriate' response selections from phase 3 shown a) for all groups and b) for participants in Group No Probe, grouped according to the discovery and consistency conditions. Standard errors are shown in parenthesis.

a)

Groups		Phase 2	Phase 3
No Probe	Consistent	83.68 (2.8)	36.46 (7.7)
	Inconsistent	77.65 (3.0)	44.32 (8.3)
Probe	Consistent	85.13 (2.6)	39.77 (9.8)
	Inconsistent	77.17 (2.6)	50.52 (9.3)

b)

Groups		Phase 2	Phase 3
Discovery	Consistent	93.37 (2.1)	40.91 (13.1)
	Inconsistent	80.00 (3.0)	50.00 (11.7)
No Discovery	Consistent	75.48 (3.6)	32.69 (9.2)
	Inconsistent	75.69 (4.2)	39.58 (12.8)