Psychonomic Bulletin & Review
2005, 12 (1), 88-92

Acquired distinctiveness and equivalence in human discrimination learning: Evidence for an attentional process

CHARLOTTE BONARDI, STEVEN GRAHAM, and GEOFFREY HALL
University of York, York, England

and

CHRIS MITCHELL
University of New South Wales, Sydney, New South Wales, Australia

In a first stage of training, participants learned to associate four visual cues (two different colors and two different shapes) with verbal labels. For Group S, one label was applied to both colors and another to both shapes; for Group D, one label was applied to one color and one shape, and the other label to the other cues. When subsequently required to learn a task in which a given motor response was required to one of the colors and one of the shapes, and a different response to the other color and the other shape, Group D learned more readily than Group S. The task was designed so that the associations formed during the first stage of training could not generate differential transfer to the second stage. The results are consistent, however, with the proposal that training in which similar cues are followed by different outcomes will engage a learning process that boosts the attention paid to features that distinguish these cues.

In a recent report, Hall, Mitchell, Graham, and Lavis (2003) made use of experimental designs and theoretical concepts derived from studies of associative learning in animals to further the analysis of acquired equivalence and distinctiveness effects in human discrimination learning. In their basic experimental design (based on one used with pigeons by Bonardi, Rey, Richmond, & Hall, 1993; see also Kaiser, Sherburne, Steirn, & Zentall, 1997), human participants received initial training in which four different geometrical figures (A, B, C, and D) were used to signal two different outcomes. Presentations of A and B were both followed by, for example, the presentation of the nonsense syllable *wug*; presentations of C and D were both followed by the nonsense syllable *zif*. In the next stage, the participants were required to learn a discrimination. In the *consistent* condition, they had to make one motor response (e.g., to press a key on the left of a keyboard) to presentations of A and of B, and a different motor response (to press a key on the right) to C and to D; in the *inconsistent* condition, one response was required to A and C and the other response to B and D.

The discrimination was acquired more readily in the *consistent* than in the *inconsistent* condition. That is, performance was superior when the task required the participants to make the same response to cues that had shared a common outcome in the first stage of training, and different responses to cues that had been trained initially with different outcomes. The observation that training in which two cues are associated with a common event can enhance generalization between them has been called the *acquired equivalence effect*; the observation that discrimination between two similar cues will be facilitated by prior training in which each has been associated with a different outcome has been called the *acquired distinctiveness effect* (see Hall, 1991, for a review).

Hall et al. (2003) considered two possible explanations for their results, one based on an extension of standard associative learning principles and the other on learned changes in attention. The associative account assumes that in the first stage of training each of the four cues will become associated with its outcome, so that on subsequent presentations A and B will both tend to activate the representation of *wug*, and C and D the representation of *zif*. These outcome representations will thus be activated during the discrimination stage of the procedure. When the subject learns to make a particular response to a given cue (e.g., to respond "left" to A), the associate of A will be activated and will also become a cue for performing that response. Since B also activates this associate, a tendency to make the same response will be elicited immediately when B is presented, which will facilitate performance for participants in the *consistent* condition. For participants in the *inconsistent* condition (who are required to respond "right" to B), this tendency will need

This work was supported by Unilever Research. S. G. is now at the National University of Singapore and Cognitive Neuroscience Laboratory, Singapore General Hospital, Singapore. Correspondence concerning this article should be addressed to G. Hall, Department of Psychology, University of York, York YO10 5DD, England (e-mail: g.hall@psych.york.ac.uk).

to be overcome and will detract from efficient discrimination performance. Although the terminology is different, this interpretation is in principle the same as that offered many years ago by Hull (1939) in his analysis of the "problem of stimulus equivalence" (see also Miller & Dollard, 1941).

The attentional analysis starts from the assumption that each of the critical cues (A, B, C, and D) will share features with each of the others; those shared only by A and B we shall call *p*, and those shared only by C and D we shall call *q*. In the first stage of training, feature *p* will uniquely signal the outcome that follows both A and B, and similarly, *q* will uniquely signal the outcome that follows C and D. Subjects keen to anticipate the outcome of any Stage 1 trial may thus learn to focus attention on these predictive features. To do so would facilitate learning of the Stage 2 task by subjects in the *consistent* condition, since this task requires them to make one response to both of the cues that contain *p* and a different response to both of the cues that contain *q*. The notion that acquired distinctiveness effects might depend on attentional processes also has a long history, having been proposed for the case of animal discrimination learning by Lawrence (1949; see also Mackintosh, 1975; Sutherland & Mackintosh, 1971) and, in a rather different form, for that of human discrimination learning by Gibson (e.g., 1969).

Although the basic effect demonstrated by Hall et al. (2003) can be explained in both associative and attentional terms, these authors favored the former interpretation on the basis of the results generated by a further phase of testing in their experiment. In this phase, the subjects were presented again with the stimuli that had been used as outcomes in the first stage of training (i.e., *wug* and *zif*, for the example described earlier) and were asked to make a motor response (the left or right keypress). Those given the *consistent* condition in Stage 2 reliably made the left response to *wug* and the right response to *zif*. This is just the pattern of behavior that would be expected if, as is postulated by the associative theory, the associatively activated representations of the nonsense syllables had become associated with the responses required (to Cues A and B, and to Cues C and D, respectively) during discrimination training. But, although this observation is consistent with the associative account, it does not necessarily disprove the attentional alternative: There is no reason that an attentional learning process should not be operating alongside the associative mechanism. The experiment to be described in this article was designed to provide unambiguous evidence of the operation of an attentional process by modifying the design of the experiment reported by Hall et al. in a way that precluded transfer on the basis of the associative mechanism.

The design of the experiment (outlined in Table 1) was based on one conducted by Delamater (1998), with rats as the subjects. The initial phase of training involved four stimuli, corresponding to A, B, C, and D of the Hall et al. (2003) study. The critical feature of the present experiment, however, was that these stimuli fell into two pairs differing along different stimulus dimensions:

Table 1
Experimental Design

Stage 1	Stage 2	Stimulus–Associate–Response Combinations
	Group S	
Sn1 → *x*	Sn1 → R1	Sn1–*x*–R1
Sn2 → *x*	Sn2 → R2	Sn2–*x*–R2
Co1 → *y*	Co1 → R2	Co1–*y*–R2
Co2 → *y*	Co2 → R1	Co2–*y*–R1
	Group D	
Sn1 → *x*	Sn1 → R1	Sn1–*x*–R1
Sn2 → *y*	Sn2 → R2	Sn2–*y*–R2
Co1 → *x*	Co1 → R2	Co1–*x*–R2
Co2 → *y*	Co2 → R1	Co2–*y*–R1

Note—Sn1 and Sn2 represent the "snowflake" stimuli of Figure 1; Colors Co1 and Co2 represent two shades of red (pale and dark, respectively); *x* and *y* represent the syllables *wug* and *zif* (counterbalanced). For half the participants, Response 1 (R1) was pressing a key on the left of a keyboard and Response 2 (R2) was pressing a key on the right; for the remaining participants, the key assignments were reversed. The rightmost column shows the combinations of stimulus (Sn1, Sn2, Co1, or Co2), associate (*x* or *y*), and response (R1 or R2) present in each condition. Note that in neither of the groups is *x* or *y* uniquely associated with a particular response.

shape (two "snowflake" patterns; see Figure 1) and color (two different shades of red).The four stimuli are presented in Table 1 as Sn1 and Sn2 (the snowflakes), and Co1 and Co2 (the colors). It was assumed that discriminating between these two dimensions would be trivially easy for our participants, whereas discriminating between members of a pair would be much more difficult. This amounts to assuming that Sn1 and Sn2 have salient common features, as do Co1 and Co2, but that the features that are shared by a color and a shape are very low in salience (or even nonexistent). Thus, the main determinant of performance on the categorization task that constituted the test phase of the design (see Table 1) would be the ability to discriminate Sn1 from Sn2 and Co1 from Co2. As in the experiment by Hall et al., the first stage of training consisted of trials in which the critical cues (in this case Sn1, Sn2, Co1, and Co2) were followed by the nonsense syllables. For Group S (for "same") Sn1 and Sn2 were both associated with one syllable (*x* in the table) and Co1 and Co2 were both associated with the other (*y*). For Group D (for "different"), one of the shapes (Sn1) and one of the colors (Co1) was associated with *x*; the other shape (Sn2) and the other color (Co2) were followed by *y*. The question of interest was how these different forms of prior training would influence the participants' ability to perform the categorization task.

The associative account has no grounds for predicting any difference between the two groups in these circumstances. For both groups, it may be assumed that each shape and each color will become associated with the nonsense syllable that follows it in Stage 1 and that the representations of these syllables will be activated when the shapes and colors are presented in Stage 2. However, the tendency to form an association between the evoked representation and the motor response required in

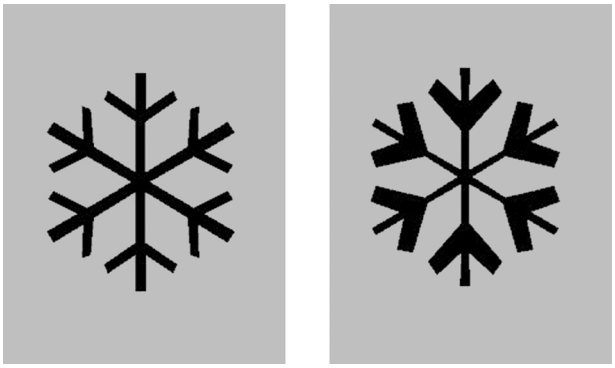

Figure 1. The "snowflake" patterns used as Sn1 (left) and Sn2 (right).

Stage 2 will not be able to facilitate performance of the Stage 2 task for either group, since in neither group is a given syllable uniquely associated with a particular response. As Table 1 shows, both groups must make a given response (R1) to Sn1 and Co2, and in both groups these stimuli had been associated with different nonsense syllables in Stage 1. Similarly, both groups must make another response (R2) to Sn2 and Co1, cues that had again been previously associated with different syllables in Stage 1. Such associations as might be acquired by the representations of the syllables would be likely to detract from accurate Stage 2 performance, but they would do so equally in both groups.

According to the attentional account, the training given to Group S should allow the features common to the two shapes and the features common to the two colors to command special attention, since these are the best predictors of the outcome of each trial. This would hinder performance on the Stage 2 categorization task, since these common features must be ignored if the subjects are to learn to make different responses to each of the two shapes and to each of the two colors. For Group D, on the other hand, no such interference should occur. For this group, it is the unique features of Sn1 and Sn2, and of Co1 and Co2, that are the best predictors of *x* and *y*, respectively; accordingly, it is these unique features that should come to command attention as a result of Stage 1 training. (We have assumed that there is no salient feature shared by Sn1 and Co1, so the fact that these stimuli share a common outcome in Stage 1 training should not impair their ability to become associated with different responses in Stage 2.) Furthermore, since attention to these unique features is necessary for accurate performance on the categorization task, the attentional account predicts that Group D should outperform Group S.

METHOD

Participants

Twenty-four participants, the majority of whom were undergraduates at the University of York, took part in the experiment. They were randomly allocated to two equally sized groups (S and D).

Apparatus

The experiment was conducted in an experimental carrel using a personal computer operating Windows 95 and equipped with a mouse, a keyboard, and a Viglen monitor (model 950T). The monitor screen was positioned at eye level, about 0.5 m from the participant. The programs that controlled stimulus presentation were written using Microsoft Developer Studio in Visual C+ + 4.0. All the stimuli were presented in the center of the screen on a gray background. The snowflake stimuli (Sn1 and Sn2) were black and approximately 7× 7 cm; Sn1 had thin and Sn2 had thick arms (see Figure 1). The color stimuli (Co1 and Co2) were pale or dark red rectangles (R/G/B values of 198/0/0 and 255/0/0, respectively). The nonsense syllables *wug* and *zif* were presented in black 72-point Comic Sans MS font.

Procedure

Before Stage 1 of the experiment, the participants were presented with a set of written instructions informing them that they would receive a series of stimuli (shapes or colors) which would be followed by a syllable (*wug* or *zif*). They were told to try to remember which stimuli and which syllables went together, since they would be asked about this at the end of the experiment. They were asked to concentrate on each stimulus as it appeared and to press the left mouse button whenever they saw the syllable *wug* or *zif*. Thirty-two trials followed, which comprised eight 0.5-sec presentations of each of the color and shape stimuli, immediately followed by one of the nonsense syllables, which remained on the screen until the participant pressed the left mouse button. This initiated the 2-sec intertrial interval, during which the words "Get ready!" appeared in the top left corner of the screen. The sequence of trials was random, apart from the constraint that each cue was presented eight times. For half of Group S, both the colored patches were followed by *wug* and both the snowflakes by *zif*, and for the remaining participants this arrangement was reversed. For half of Group D, one of the snowflakes (Sn1) and one of the colored patches (Co1) was followed by *wug*; the other stimulus of each pair was followed by *zif*. For the remaining participants in Group D, this arrangement was reversed.

At the start of Stage 2, the participants received a second set of written instructions informing them that they now had to decide to which of two categories each of the shape and color stimuli belonged. The categories were defined as *left* (indicated by pressing the "\" key, situated on the far left of the keyboard) or *right* (indicated by pressing the "/" key, situated on the far right of the keyboard). The participants were told that initially they would have to guess but that feedback would be given after each trial, so that by trial and error they should be able to learn to categorize the stimuli accurately. They were told to proceed quickly but to try to avoid making mistakes. Thirty-two trials followed, which comprised eight presentations of each of the colors and each of the snowflake stimuli, presented in random order. The intertrial interval was again 2 sec, and the stimuli remained on the screen until the participants had made a categorization response. When a response was made, the word "Correct!" or "Wrong" appeared in the top left corner of the screen, throughout the duration of the intertrial interval. For half of the participants in each group, the *right* response was required to Sn1 and Co2 and the *left* response to Sn2 and Co1; for the remainder, these assignments were reversed (see Table 1). Reaction times (with 1-msec resolution) were also recorded during this phase.

RESULTS AND DISCUSSION

No data were collected during the first stage of the experiment. The critical results come from the classification data of Stage 2, in which the participants had to classify the various stimuli into the categories *left* and *right*.

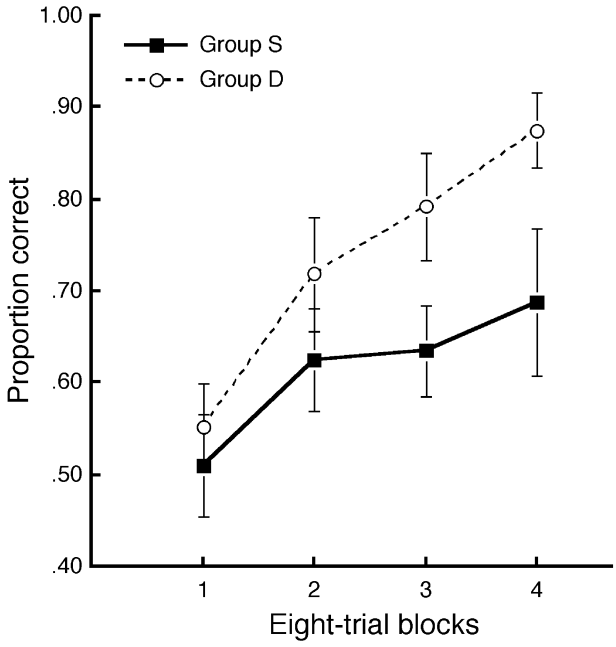

Figure 2. The mean proportion of correct responses made by participants in Group S and Group D in the four eight-trial blocks of the test phase. Vertical bars represent standard errors of the means.

The proportion of correct responses for each group on successive eight-trial blocks of this stage is shown in Figure 2. The figure shows that initial performance was at or near chance in both groups, but that as acquisition occurred with continued training, Group D began to outperform Group S. This description of the results was supported by an analysis of variance with group and block as variables, which revealed a significant main effect of both group [$F(1,22) = 5.06, p = .035$] and blocks [$F(3,66) = 8.92\ p = .000$]. The interaction between the variables was not reliable ($F < 1$). The average reaction time for a correct response was 922 msec for Group S and 972 msec for Group D; these scores did not differ significantly ($F < 1$), demonstrating that the more accurate performance in Group D was not the result of a speed/accuracy trade-off.

The results of this experiment show that acquired equivalence/distinctive effects can be obtained with an experimental design in which the associative mechanism described in the introduction cannot generate a difference between the groups. They thus accord with the outcome of the study by Delamater (1998), in which rats were the experimental subjects. To demonstrate acquired equivalence/distinctiveness in these circumstances is not to show that associative processes do not normally play a role—indeed, there is evidence both for rats (e.g., Honey & Hall, 1989) and for people (e.g., Hall et al., 2003) suggesting that they might. But, with one exception (see below), in previous investigations of these effects using human participants, designs have been used that have permitted explanation in both associative and attentional terms.

The exception is a study recently reported by Le Pelley and McLaren (2003). They made use of a causal learning task in which the cues were the names of foodstuffs, consumption of which might lead to the development of a particular type of allergic reaction (the outcome) in an imaginary patient. Although the details were somewhat more complex, the essence of their design may be summarized as follows: In the first stage of training, Cues A and D were consistently followed by Outcome 1, and Cues B and C by Outcome 2. Each of four other cues, X, Y, V, and W, was followed equally often by Outcome 1 and by Outcome 2. In Stage 2, compound cues were trained as signals for new outcomes: AX and CV were both followed by Outcome 3; BY and DW were both followed by Outcome 4. This arrangement ensured that, during Stage 2, the associatively activated representations of the Stage 1 outcomes occurred equally often in the presence of Outcome 3 and Outcome 4. They cannot, therefore, have any effect on the formation of the Stage 2 discrimination. Nonetheless, a final test showed that the cues that had been consistently followed by a given outcome in Stage 1 (A, B, C, and D) were given higher ratings as being the likely cause of their Stage 2 outcomes than were the cues that had had inconsistent consequences in Stage 1 (X, Y, V, and W).

Le Pelley and McLaren (2003) explained their results in terms of the theory of attention (or stimulus associability) proposed by Mackintosh (1975) in the context of work on conditioning in animals. This theory holds that some aspects of the attention paid to a given cue can be modified by experience, and, in particular, by experience of the cue's reliability as a predictor of other events. A cue that is a good predictor of an outcome will undergo an increase in associability (i.e., it will be better able subsequently to enter into new associations); a cue that is a poor predictor will suffer a loss of associability. It follows that in the study by Le Pelley and McLaren, when a compound such as AX is followed by Outcome 3 in Stage 2, the association between A and the outcome should be formed more readily than that between X and the outcome.

Mackintosh's (1975) theory can be applied readily to the present results. For Group S, it is the common features of the two shape stimuli and of the two color stimuli that reliably predict the outcome in Stage 1, and these features should gain associability. For Group D, on the other hand, the outcome of a Stage 1 trial is predicted by the features that are unique to the two shapes and by those that are unique to the two colors, and it is these features that will gain associability. Group D will therefore be at an advantage when it comes to the Stage 2 task, since accurate performance on that task requires the participants to learn about the features that distinguish Sn1 from Sn2 and Co1 from Co2.

In some respects, the results of the present study provide more persuasive evidence in favor of this attentional

analysis than do those of Le Pelley and McLaren (2003). In their study, the properties of the various cues were assessed in a final test that was presumed to give information about the strength of associations formed over the course of the previous stage of training. An advantage of our procedure was that we were able to monitor directly the performance shown over the course of training and to demonstrate (see Figure 2) that acquisition proceeded more readily for the group for which the critical cues are postulated to have a high level of associability. Relatedly, the use of a separate final test stage raises the possibility that the results obtained might be a consequence of some inferential process that operates at the time of test (as opposed to an associability mechanism that operates to determine acquisition over the previous stage of training). That is, the subjects in Le Pelley and McLaren's study might have been able to look back on information acquired about the cues in both of the previous stages of training and make a response on this basis (judging, for instance, that a cue that was unreliable in Stage 1 should not be trusted as a potential cause of the Stage 2 outcome.) It is difficult to see how a process of this sort could be responsible for the results reported here.

The account of attentional factors proposed by Mackintosh (1975) is just one of several possibilities, but it gains an advantage over alternatives in the way in which it describes changes in associability as dependent on the extent to which cues (or aspects of cues) are good predictors of their outcomes. The interpretation offered by Gibson (1969), for example, also assumes that subjects will come to increase the attention they pay to features that distinguish stimuli and learn to ignore those that do not. But the process responsible for this is assumed to be quite independent of associative learning—mere exposure to the stimuli should be enough to achieve it. Without further elaboration, therefore, there is no reason for this theory to predict any difference between our S and D groups, given that both received the same exposure to the cues in Stage 1 of training. What remains a problem for Mackintosh's theory is that direct tests, using animal conditioning procedures, of its central proposition (that a predictive cue undergoes an increase in associability) have failed to provide support for it (see, e.g., Hall & Pearce, 1979; Pearce & Hall, 1980). This matter remains to be resolved.

REFERENCES

Bonardi, C., Rey, V., Richmond, M., & Hall, G. (1993). Acquired equivalence of cues in pigeon autoshaping: Effects of training with common consequences and with common antecedents. *Animal Learning & Behavior*, **21**, 369-376.

Delamater, A. R. (1998). Associative mediational processes in the acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology: Animal Behavior Processes*, **24**, 467-482.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Hall, G. (1991). *Perceptual and associative learning*. Oxford: Oxford University Press, Clarendon Press.

Hall, G., Mitchell, C., Graham. S., & Lavis, Y. (2003). Acquired equivalence and distinctiveness in human discrimination learning: Evidence for associative mediation. *Journal of Experimental Psychology: General*, **132**, 266-276.

Hall, G., & Pearce, J. M. (1979). Latent inhibition of a CS during CS–US pairings. *Journal of Experimental Psychology: Animal Behavior Processes*, **5**, 31-42.

Honey, R. C., & Hall, G. (1989). Acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology: Animal Behavior Processes*, **15**, 338-346.

Hull, C. L. (1939). The problem of stimulus equivalence in behavior theory. *Psychological Review*, **46**, 9-30.

Kaiser, D. H., Sherburne, L. M., Steirn, J. N., & Zentall, T. R. (1997). Perceptual learning in pigeons: Decreased ability to discriminate samples mapped onto the same comparison in many-to-one matching. *Psychonomic Bulletin & Review*, **4**, 378-381.

Lawrence, D. H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology*, **39**, 770-784.

Le Pelley, M. E., & McLaren, I. P. L. (2003). Learned associability and associative change in human causal learning. *Quarterly Journal of Experimental Psychology*, **56B**, 68-79.

Mackintosh, N. J. (1975). A theory of attention: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, **82**, 276-298.

Miller, N. E., & Dollard, J. (1941). *Social learning and imitation*. New Haven, CT: Yale University Press.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, **87**, 532-552.

Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press.

(Manuscript received November 14, 2003;
revision accepted for publication February 12, 2004.)