Psychon Bull Rev (2013) 20:385–390
DOI 10.3758/s13423-012-0346-3

BRIEF REPORT

Guilt by association and honor by association: The role of acquired equivalence

Mikaël Molet · Jessica P. Stagner · Holly C. Miller · Thierry Kosinski · Thomas R. Zentall

Published online: 4 December 2012

Abstract Guilt by association and honor by association are two types of judgments that suggest that a negative or positive quality of a person or object can transfer to another person or object, merely by co-occurrence. Most examples have been demonstrated under conditions of direct associations. Here, we provide experimental evidence of guilt by association and honor by association via indirect associations. We show that participants may treat two individuals alike if they have been separately paired with a common event using an acquired-equivalence paradigm. Our findings suggest that association fallacies can be examined using a paradigm originally developed for research with nonhuman animals and based on a representation mediation account.

Keywords Guilt by association · Honor by association · Acquired equivalence

Guilt by association and honor by association occur when two people or objects are judged to be similar due to their co-occurrence. If the association is negative, it is called guilt by association, and if it is positive, it is called honor by association.

M. Molet (✉) · T. Kosinski
Département de Psychologie, Université Lille–Nord de France,
Domaine universitaire du "Pont de Bois", rue du Barreau,
BP, 60149, 59653 Villeneuve d'Ascq Cedex, France
e-mail: mikael.molet@univ-lille3.fr

J. P. Stagner · T. R. Zentall
University of Kentucky, Lexington, KY, USA

H. C. Miller
KU Leuven, Leuven, Belgium

An emerging view appears to be that guilt by association and honor by association may depend on conditioning processes (Walther & Langer, 2010). For example, people often form negative attitudes toward anyone who conveys unpleasant information (e.g., Manis et al. 1974): The unpleasant information coincides with the messenger; hence, the messenger is perceived adversely (a kill-the-messenger effect). Similarly, people who describe others negatively or positively are themselves often perceived unfavorably or favorably, respectively (e.g., Skowronski et al. 1998; see also Gawronski & Walther, 2008). That is, these individuals become associated with negative or positive attributes. Skowronski et al. 1998. refer to this effect as spontaneous trait transference.

The mere-ownership or endowment effect is also consistent with these examples. For example, Thaler (1980) found that randomly assigned owners of an object appear to value the object more than do randomly assigned nonowners of the object. In a related vein, Bliss-Moreau et al. (2010) had participants listen to different speakers emit positive, negative, or neutral words. Next, the participants had to decide whether another series of primed neutral words, such as "seat," were positive or negative. The prime was spoken by one of the previous speakers. The results showed that primes that were spoken by a person who had previously uttered positive words were perceived positively. Thus, voices associated with positive words tended to imbue other words with a positive valence.

These phenomena can be described in conditioning terms as a change in the value of a conditioned stimulus (CS) caused by its co-occurrence with an unconditioned stimulus (US) that has affective value (Levey & Martin, 1975). For example, in the kill-the-messenger effect, the messenger (CS) experiences a revaluation by being associated with unpleasant news (US). In spontaneous trait transference, the communicator (CS) is associated with the (un)pleasant description made about a third party (US).

In research perhaps even more relevant to the present study, Walther (2002) demonstrated that the (dis)liking of a person could be developed following the person's pairing not with affective information (e.g., the kill-the-messenger effect), but with a second person who had been paired with affective information. More specifically, two neutral faces ($F_1 - F_2$) were paired in Phase 1, and then F_2 was paired with an affective stimulus (either positive or negative) in Phase 2. During testing of F_1, an evaluative response appropriate to the affective stimulus was observed, even though F_1 itself had never been directly paired with the affective stimulus. It was hypothesized that the pairing of F_1 with F_2, and the subsequent pairing of F_2 with the affective stimulus, resulted in the formation of an excitatory association between each element by the following associative chain: F_1–F_2–affective stimulus. Specifically, F_2 became excitatory in Phase 2 of training, and then, since F_1 had been associated with F_2 in Phase 1, this produced an emotional conditioned evaluation to F_1 at test. This finding can be described as a case of higher-order conditioning (see Rizley & Rescorla, 1972), which extended the demonstration of guilt by association and honor by association to more than the co-occurrence of an attitudinal person/object with an evaluated experience.

To date, studies of conditioning related to guilt by association and honor by association have focused, for the most part, on first-order conditioning (e.g., Manis et al., 1974), in which an individual has been directly paired with positive or negative information, and on higher-order conditioning, in which two individuals have been paired together, and subsequently one has been associated with negative or positive information that has then transferred to the other stimulus (e.g., Walther, 2002). In the present research, we asked whether guilt by association and honor by association can occur under conditions in which two individuals have not been directly presented together. Currently, it is not known whether people who share a common, presumably irrelevant, association can come to share a positive or negative characteristic that has been attributed to only one of them. We propose that the acquired-equivalence paradigm is a good candidate for investigating this question. Acquired equivalence refers to the phenomenon that when two stimuli are associated with a common event, these stimuli can come to be regarded as equivalent. Thus, subsequent learning about one stimulus can be transferred to the other stimulus in the absence of direct training.

Research with animals has suggested that an equivalence relation will often develop in this way. This equivalence relation can be demonstrated following training by associating one of the two stimuli with a new, unrelated outcome and finding that the second stimulus is also associated with the new outcome (e.g., Urcuioli et al. 1989; Zentall et al., 1992; Zentall et al. 1995). In Phase 1, Urcuioli et al. reinforced pigeons' choice of Comparison W following presentation of Samples A and B, and reinforced choice of the alternative Comparison X following presentation of two other samples, C and D (formally, A→W, B→W and C→X, D→X). The pigeons were then trained to associate one pair of those samples, A and C, with new comparison stimuli Y and Z (A→Y and C→Z). Evidence for emergent stimulus relations was found when the remaining samples from the original training (i.e., B and D) were then presented with the new comparisons (Y and Z). In the absence of any explicit training, most of the pigeons chose Y when the sample was B and chose Z when the sample was D. Such evidence of transfer is the hallmark of an acquired equivalence between cues trained with common comparison stimuli.

The explanation of acquired equivalence has typically been based on an associative chain mediated by the representation of the common event, the so-called representation mediation hypothesis (e.g., Gluck & Myers, 2001). The idea is that during initial training, the representations of stimuli paired with the same outcome become modified to enhance generalization between them. This approach assumes that associative links or category assignments are stored in memory as representations of the stimuli. In this way, at test, the presentation of one stimulus evokes retrieval of stimulus representations from memory, and a response is given on the basis of the theoretical link or category assignment stored in the retrieved representation. In Urcuioli et al. (1989), Stimuli A and B were first paired with a common event W (A→W, B→W), presumably causing event W to be linked to or incorporated within memory representations of A and B (the same rationale is true for Samples C and D and Comparison X). Then, when A was paired with a different event Y, Y was linked to or incorporated with the representation of A. Thus, when B was presented, it was presumed to retrieve event W, but event W should lead to retrieval of the representation of A, which in turn should lead to retrieval of Y. In this way, it is possible for an emergent relation to develop between B and Y, as observed by Urcuioli et al. (the same rationale can be used for the emergent relation between Sample D and Comparison Z).

This kind of acquired equivalence has also been demonstrated in humans (e.g., Hall et al. 2003; Smyth et al. 2008). However, to the best of our knowledge, none of the researchers who have studied such transfer in the equivalence literature have spoken (in)directly to the point of guilt and honor by association. Instead, the previous investigators have used neutral events as the comparison choices. In one study, although aversive stimuli (e.g., unpleasant pictures of mutilated faces) were used as the comparison stimuli, they were directly trained, so the test did not involve the transfer of training of positive and negative attributes (Tyndall et al. 2009).

In the present research, we asked whether guilt and honor by association would occur by using a simple

functional-equivalence design, analogous to designs used in animal studies. Participants were first trained to associate sample images of two faces, F_1 and F_2, with choice of a blue water bottle (comparison stimulus), and two other faces, F_3 and F_4, with choice of a green water bottle. The question was, if they were to learn to associate F_1 with negative behavior and F_3 with positive behavior, would they then attribute the respective behaviors to F_2 and F_4? If, in the absence of direct experience, F_2 were associated with negative behavior and F_4 with positive behavior, this would suggest that the F_1 and F_2 faces and the F_3 and F_4 faces had acquired functional equivalence. Whereas previous research had shown the influence of guilt and honor by association when two individuals were directly associated, the goal of the present research was to determine whether information about one person would transfer to another when the people were not directly associated, but merely linked through their common association with a presumably irrelevant stimulus.

Method

Participants

The participants were 24 undergraduate students (ages 18–23 years; eight of each sex) from the University of Kentucky.

Apparatus

The participants performed the experiment using a Dell Latitude E540 computer. We used a many-to-one matching-to-sample task that was programmed using Visual Basic (see Fig. 1). Four women's faces were used as samples and two images of water bottles as comparisons in Phases 1, 2, and 3 (Training 1); subsequently, one sentence describing positive behavior ("Reads a book out loud to residents of a nursing home") and one describing negative behavior ("Slammed the door in the face of a girl scout selling cookies") were used as comparisons in Phases 4 (Training 2) and 5 (transfer test). All of the stimuli were counterbalanced across participants.

Procedure

A mouse click on the Play button caused presentation of the sample stimulus. After an interval of 500 ms, the comparison stimuli were presented simultaneously below the samples. When the participant clicked on the correct choice button (located below the comparison stimuli), the word "correct" appeared between the comparison stimuli; when the incorrect choice button was clicked, the word "incorrect" was displayed. The feedback messaged remained visible for 1.0 s.

The experiment was conducted in five phases. In Phase 1 (see "Training 1" in Fig. 1), the participants' choice of one comparison (water bottle) was correct following presentation of one sample (F_1), and choice of the other comparison was correct following presentation of the other sample (F_3). A minimum of 12 trials were presented, six with each sample. These training trials were presented in random order, with no more than two of the same samples being presented consecutively. The left or right position of the comparisons was also randomized. The participants remained in Phase 1 until they reached a criterion of 5/6 correct trials for each of the two samples. Participants who did not reach criterion in 12 trials were given an additional 12 trials of training. The procedure used in Phase 2 was the same as that in Phase 1, but with new sample faces (F_2 and F_4; again see "Training 1" in Fig. 1) and the same comparison stimuli. Choosing one comparison in the presence of one sample (F_2) was reinforced, and choosing the other comparison in the presence of the other sample (F_4) was reinforced. The participants remained in Phase 2 until they had reached a criterion of 5/6 correct trials for each of the two samples. Phase 3 involved a mixture of Phase 1 and Phase 2 trials—at least 24 trials, six with each of the four samples. The participants remained in Phase 3 until they had reached a criterion of 5/6 correct trials for each of the four samples.

The procedure used in Phase 4 was similar to that of Phase 1 (or Phase 2, counterbalanced among participants), but new comparisons were presented (i.e., the positive and negative behaviors). For trials with one of the sample faces (e.g., F_1), choice of the positive behavior was correct, and for trials with the other sample face (e.g., F_3), choice of the negative behavior was correct (see "Training 2" in Fig. 1).

Phase 5 consisted of one test trial with each of the two samples from Phase 3 that had not been presented in Phase 4 (e.g., F_2 and F_4), but with the same comparisons from Phase 4 (positive and negative behaviors). The task of the participants was to indicate who performed a positive behavior and who performed a negative behavior. No feedback was given on the test trials (see "Transfer Test" in Fig. 1).

Results and discussion

The results of the experiment are summarized in Fig. 2. An equivalence response occurred whenever participants responded to the untrained sample face (e.g., F_2→negative behavior) in the same way that they had been trained to respond to the sample face associated with the same water bottle (e.g., F_1→negative behavior). Such an equivalence response is represented by a "1," and "1, 1" indicates that an equivalence response was made on both the first and the second test trial. A "0" represents the absence of an

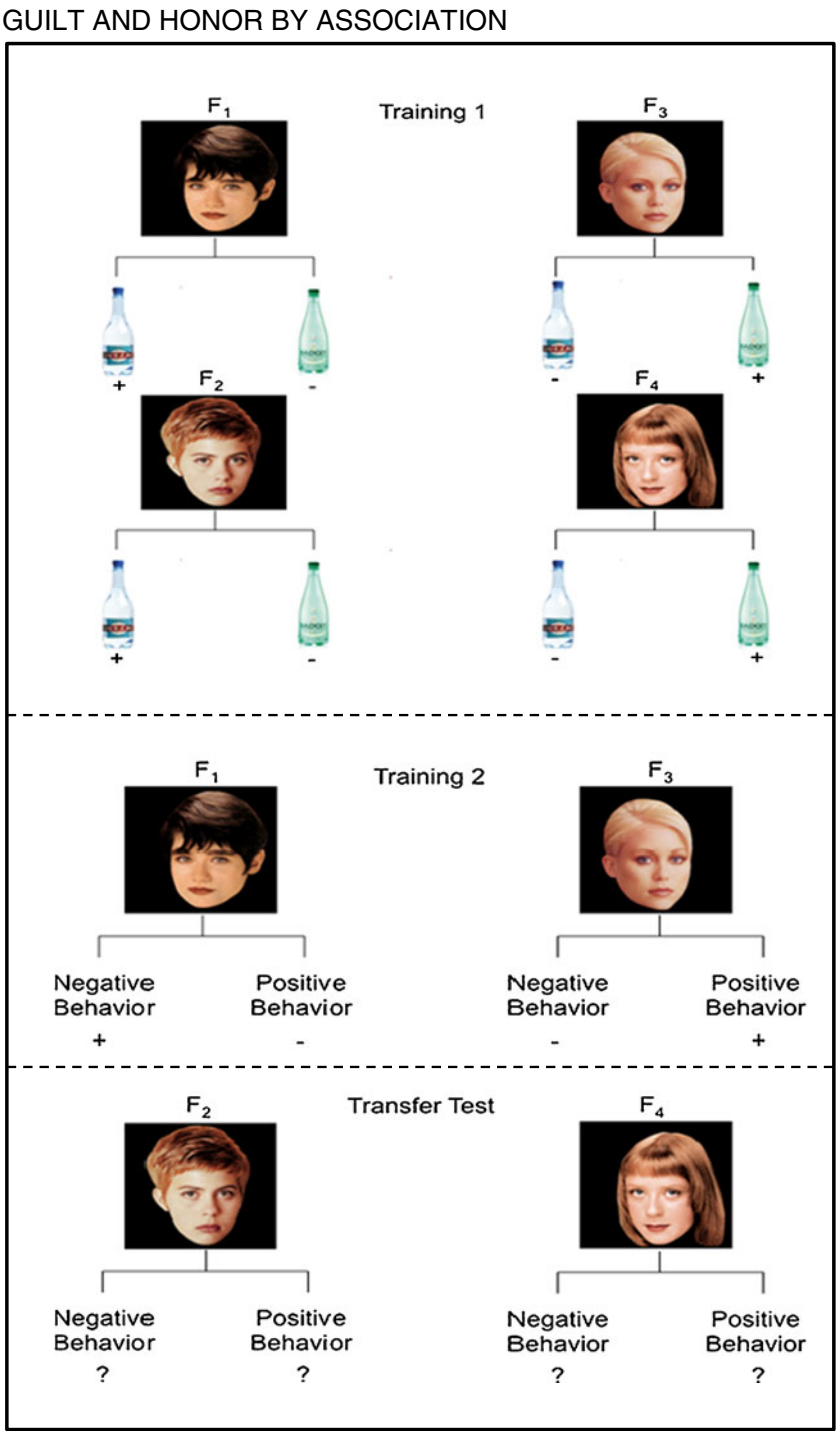

Fig. 1 Experimental design. Participants were trained on many-to-one matching to sample: In Training 1, the faces refer to the samples, and the water bottles refer to the comparisons. First, two samples were trained (Phase 1); then two more samples were trained (Phase 2); and finally, all four samples were combined (Phase 3). In Training 2 (Phase 4), positive and negative behaviors served as the comparisons with one pair of samples, and in transfer testing (Phase 5), the remaining samples from Training 1 were presented with the comparisons from Phase 4 (see details in the text). The "+" and "–" signs refer, respectively, to correct and incorrect choice responses

equivalence response. As can be seen in the figure, the majority of the participants made an equivalence response on both test trials (17 out of 24). Only one participant made an equivalence response on the first test trial but not on the second test trial, and another participant showed the reverse pattern. The five remaining participants did not make an equivalence response on either test trial. A binomial test indicated that a significant number of participants made an equivalence response on both the first test trial (p = .008) and the second test trial (p = .008). Furthermore, with regard to equivalence, if one considers only two response patterns (the 17 participants with a consistent 1, 1 equivalence pattern vs. the seven participants with one of the three other patterns—1, 0; 0, 1; and 0, 0), a binomial test indicated that the effect was still significant (p = .021). Thus, the majority of participants made a consistent equivalence response.

The present findings are consistent with previous conditioning research. For example, Walther (2009) observed that

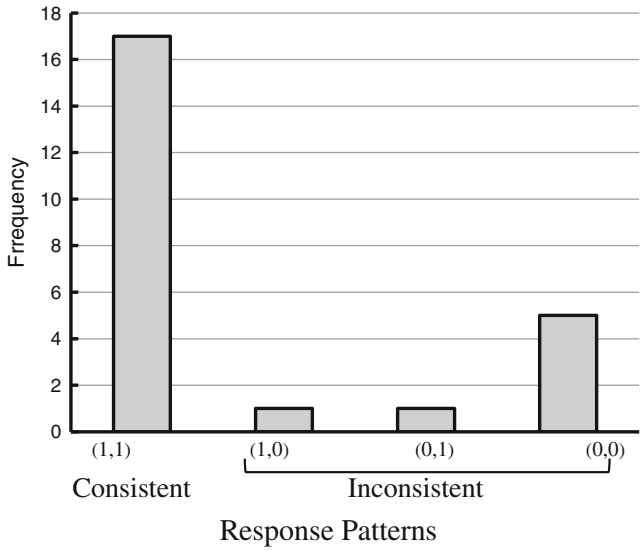

Fig. 2 The numbers of participants who were consistent with emergent equivalence (a score of 1, 1—equivalence on both the first and second trials) and inconsistent with emergent equivalence (a score of 1, 0 or 0, 1—equivalence on one trial but not the other—or of 0, 0—equivalence on neither trial)

participants may treat two individuals alike if they have been paired together. However, the present findings go beyond previous demonstrations in showing that participants may treat two individuals alike, even if the individuals have been separately paired with a common event. Thus, subsequent learning about one person can be transferred to the other person in the absence of a direct association.

What is the process underlying the guilt- and honor-by-association effects in the present study? According to the representation mediation hypothesis, Faces 1 and 2 were first paired with the blue water bottle ($F_1 \rightarrow B$, $F_2 \rightarrow B$), presumably causing event B to be linked to or incorporated within memory representations of F_1 and F_2 (similarly, Faces 3 and 4 were presumed to be linked with the green water bottle). Then, when F_1 was paired with a positive behavior (P), P was linked to or incorporated within the representation of F_1. Thus, when F_2 was presented, it was presumed to retrieve event B, but event B was assumed to retrieve a representation of F_1, which in turn should lead to retrieval of P. In this way, it was possible for an emergent relation to develop between F_2 and P, as we observed in the present experiment (the same rationale could be used for the emergent relation between F_3 and negative behavior). Balance theory (Heider, 1944) suggests that people are motivated to see their world in balanced states, and so will be motivated to judge/perceive stimuli in ways in which such balance will be achieved. One of those ways is to see similar people as liking similar things, as having similar traits, or as doing similar things. The social psychology literature provides support for many of the predictions made by balance theory, and the results of the present research are also consistent with this theory.

What do these results suggest about the affect, cognition, and/or motivation that is responsible for this equivalence effect? That is, why are the stimuli treated as functionally equivalent? Three mechanisms provide possible explanations: (1) It could be that the effect is driven by affect. If two faces are seen as being similar (Training Phase 1) and one comes to be liked/disliked (Training Phase 2), then the other will also become liked or disliked (transfer phase). (2) Alternatively, it could be that the effect is driven by trait judgments. If the faces are seen as being similar (Training Phase 1) and one comes to be seen as having a given trait (Training Phase 2), then the other will also be seen as having the given trait, which will affect judgments of trait-related behavior (transfer phase). (3) Finally, it could be that the effect is driven by assumed behavior similarity. If the faces are seen as being similar (Training Phase 1) and one engages in a given behavior (Training Phase 2), then it will be assumed that the other will engage in the same behavior or in a behavior with similar implications (transfer phase). Thus, the present research does not indicate what mechanism underlies the functional equivalence found in the present research. However, it does indicate that people who share an independent association can come to share an unrelated positive or negative characteristic that is later attributed to only one of them, on the basis of a representation mediation mechanism.

In fact, research with groups has suggested that the kind of transfer found in the present study can occur with regard to the traits of group members (Crawford et al. 2002). In Crawford et al.'s study, although insufficient evidence was provided about the traits of the group members, participants attributed the traits of some group members to the other members of the group because of their common group membership.

Finally, one could argue that the training task also produces more than an association between two individuals. For example, an association exists between $F_1 + F_2$ and $F_3 + F_4$ (which is intended), but also differentiation between ($F_1 + F_2$) and ($F_3 + F_4$). In essence, we have created two-person groups that have intramember similarity and intergroup differentiation. This is no longer a simple person–person link. Given this possibility, do the present results show a case of failed group-level source monitoring leading to misidentification of behaviors? Perhaps the differentiation between the groups is more important than the similarities within the groups, so that greater attention is paid to what separates the groups.

No data concerning the participants' conscious motivations while making their choices were collected in the present experiment. Thus, consciously controlled reliance on the common element (i.e., the association between faces and a

common water bottle) cannot be ruled out. However, even if the participants consciously based their judgments on the common element, their decision could be considered representative of their real-life behavior, and this association could seem reasonable to them. According to Hamilton and Sherman (1996), character judgment induces online information processing in which early information serves as a basis for the individual impression, and subsequently learned information is assimilated to fit this impression in a memory-based fashion. Nevertheless, we acknowledge that the functional-equivalence effects reported here were obtained using a fully controllable measure. Thus, it would be worth investigating whether the present effects could be replicated using a more indirect behavioral measure such as the implicit relational assessment procedure (IRAP; Barnes-Holmes et al., 2006) to minimize demand compliance, because the IRAP requires participants to respond quickly and accurately in ways that are either consistent or inconsistent with their putative attitudes.

To the best of our knowledge, the present research represents the first attempt to link the principle of acquired equivalence with the association fallacies research literature (i.e., guilt by association and honor by association). We have provided experimental evidence of guilt by association and honor by association, in which a positive or negative judgment of a person is formed because a positive or negative quality of one person is transferred to another by means of an independent and presumably irrelevant association. Our findings suggest that association fallacies can be the product of acquired equivalence and can be examined using a paradigm originally developed for research with nonhuman animals. We hope that the present research stimulates further theoretical speculations and empirical investigations.

References

Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? Developing the implicit relational assessment procedure (IRAP) as a direct measure of implicit beliefs. *Irish Psychologist, 32,* 169–177.

Bliss-Moreau, E., Owren, M. J., & Barrett, L. F. (2010). I like the sounds of your voice: Affective learning about vocal signals. *Journal of Experimental Social Psychology, 46,* 557–563.

Crawford, M. T., Sherman, S. J., & Hamilton, D. L. (2002). Perceived entitativity, stereotype formation, and the interchangeability of group members. *Journal of Personality and Social Psychology, 83,* 1076–1094.

Gawronski, B., & Walther, E. (2008). The TAR effect: When the ones who dislike become the ones who are disliked. *Personality and Social Psychology Bulletin, 34,* 1276–1289.

Gluck, M. A., & Myers, C. E. (2001). *Gateway to memory: An introduction to neural network modeling of the hippocampus in learning and memory*. Cambridge, MA: MIT 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.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review, 103,* 336–355.

Heider, F. (1944). Social perception and phenomenal causality. *Psychological Review, 51,* 358–374.

Levey, A. B., & Martin, I. (1975). Classical conditioning of human "evaluation" responses. *Behavior Research and Therapy, 13,* 221–226.

Manis, M., Cornell, S. D., Moore, J. C., & Jeffrey, C. (1974). Transmission of attitude relevant information through a communication chain. *Journal of Personality and Social Psychology, 30,* 81–94.

Rizley, R. C., & Rescorla, R. A. (1972). Associations in second-order conditioning and sensory preconditioning. *Journal of Comparative and Physiological Psychology, 81,* 1–11.

Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology, 74,* 837–848.

Smyth, S., Barnes-Holmes, D., & Barnes-Holmes, Y. (2008). Acquired equivalence in human discrimination learning: The role of propositional knowledge. *Journal of Experimental Psychology. Animal Behavior Processes, 34,* 167–177.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization, 1,* 39–60.

Tyndall, I. T., Roche, B., & James, J. E. (2009). The interfering effect of emotional stimulus functions on stimulus equivalence class formation: Implications for the understanding and treatment of anxiety. *European Journal of Behaviour Analysis, 10,* 121–140.

Urcuioli, P. J., Zentall, T. R., Jackson-Smith, P., & Steirn, J. N. (1989). Evidence for common coding in many-to-one matching: Retention, intertribal interference, and transfer. *Journal of Experimental Psychology. Animal Behavior Processes, 15,* 264–273.

Walther, E. (2002). Guilt by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology, 82,* 919–934.

Walther, E., & Langer, T. (2010). For whom Pavlov's bell tolls: Processes underlying evaluative conditioning? In J. P. Forgas, J. Cooper, & W. D. Crano (Eds.), *The psychology of attitudes and attitude change: An introductory overview* (pp. 59–75). New York, NY: Psychology Press.

Zentall, T. R., Sherburne, L. M., Steirn, J. N., Randall, C. K., Roper, K. L., & Urcuioli, P. J. (1992). Common coding in pigeons: Partial versus total reversals of one-to-many conditional discriminations. *Animal Learning & Behavior, 20,* 373–381.

Zentall, T. R., Sherburne, L. M., & Urcuioli, P. J. (1995). Coding of hedonic and nonhedonic samples by pigeons in many-to-one delayed matching. *Animal Learning & Behavior, 23,* 189–196.