Understanding Guilt-by-Association: A Review of the Psychological Literature on Attitude Transfer and Generalization

Kate A. Ratliff
University of Florida


In his 1573 book, The Garden of Pleasure, the British writer James Sandford wrote: “He that goeth to bedde wyth Dogges, aryseth with fleas.” This is the first-known English-language record of a saying my grandmother repeated often: “If you lie down with dogs, you’ll get up with fleas.” This refers, of course, to the idea that we should be careful of the company we keep, either because we may be led astray by our disreputable associates, or, more relevant to this chapter, because other people might think we have been led astray, and our reputation will suffer as a result. Some of us might decide our love for dogs is worth the risk of fleas, but my grandmother’s advice is not entirely incorrect—stimuli often do take on meaning in the absence of direct experience through their relationships with other stimuli (Shanks, 1995). A growing body of research on impression formation speaks to attitude transfer, an effect by which evaluations of one individual transfer to another individual who is associated\(^1\) in some way (Ratliff & Nosek, 2008).

Attitude transfer is a specific instantiation of generalization, a principle in classical learning theory by which a response conditioned to one stimulus generalizes to other stimuli that are similar (see Till & Priluck, 2000, for a review). In early studies of classical conditioning, Pavlov (1927) found that dogs would salivate (a conditioned response) at the sound of a bell (a

\(^1\) Here I use the term associated in the general, lay sense to mean connected, linked, or related in some way. The definition contains no assumption about the presence or absence of deliberate, propositional reasoning about the relationship between the stimuli, a point I will return to in greater detail later.
conditioned stimulus) after repeated pairings of the bell with food (an unconditioned stimulus). Subsequent studies showed that a bell with a different tone than the original would also produce the salivation response through generalization. According to these theories, the strength of generalization is directly proportional to how similar a new stimulus is to the original conditioned stimulus (i.e., the generalization gradient); a new bell is more likely to elicit salivation to the extent that it is similar in tone to the original bell (Klein, 2019).

In studies of learning, the experimental situation is tightly controlled and the observed behavior is relatively unambiguous (e.g., milliliters of saliva produced, number of times a lever is pushed). But understanding the role of generalization in people’s evaluations of other people is not so straightforward. First, understanding evaluations of people is complicated by the fact that target’s behavior and traits are inherently ambiguous and context-dependent (Uleman, 2005). Further, whereas a sound can be made more or less similar to another sound by manipulating frequency in hertz (i.e., pitch), judgements of the similarity between two people can be influenced by the perceiver’s motivations, expectations, prejudices, personality, and situational constraints. And, of course, those of us studying person perceptions are continually attuned to the discrepancies between what people think, say, and do. Thus, documenting attitude generalizations in person perception has unique challenges in comparison to our behaviorist colleagues’ observations of whether pigeons learn to peck disks of differing color wavelengths (Blough, 1967) or whether the fear that a rodent learns in a box with a square roof transfers to a box with a triangular roof (Huckleberry, Ferguson & Drew, 2016).

The goal of this chapter is to provide an overview of attitude generalization and transfer in person perception to better understand when, why, and how people use information about one person or group to judge related others, and how these processes can form and maintain group-
based prejudices and stereotypes by “spreading” evaluative information—particularly negative information—across group members. The chapter will focus on what we have learned about attitude generalization and transfer, primarily in person perception, under the following general themes: lay beliefs about attitude generalization and transfer; evidence for attitude transfer on direct and indirect measures; associative and propositional explanations for attitude transfer effects; and similarity, categorization, and valence effects in attitude transfer. The chapter concludes with a discussion of related phenomena that may have implications for understanding attitude transfer and generalization, such as transference (Andersen, Glassman, and Chen, 1995), cognitive balance (Heider, 1958), spontaneous trait transfer (Skowronski, Carlston, Mae, & Crawford, 1998), and stereotyping (Hamilton & Trolier, 1986).

Lay Beliefs about the Acceptability of Attitude Transfer

Over the years, my lab has explored people’s lay beliefs about the acceptability of attitude transfer among people with various kinds of relationships. In one study, we collected demographic information from U.S. American participants along with several individual difference measures. Participants also responded to the stem: “To what extent is it acceptable to use information about one person to form an impression of [X]”, where X was 24 possible relationships, individually presented, including those of a coincidental nature (e.g., person standing at the same bus stop), those that include a choice (“their best friend”) and those that differ in family closeness (e.g., “sibling” or “cousin”). Table 1 shows the average response, on a 7-point scale ranging from 1 = Very Unacceptable to 5 = Very Acceptable.

---

2 The Attitudes and Social Cognition Lab at the University of Florida has accumulated a considerable amount of data on attitude transfer that never made it into manuscripts; we are using this chapter as an opportunity to present some of those findings that we think might be interesting or useful to others. We set up a project page on the Open Science Framework to share our unpublished study materials and data—https://osf.io/xas3w/. You can find more details about the studies there.
Table 1. Average perceived acceptability of attitude transfer by relationship type. Participants (N = 1,848) responded to the stem “To what extent is it acceptable to use information about one person to form an impression of [X]” on a 7-point scale ranging from 1 = Very Unacceptable to 5 = Very Acceptable

<table>
<thead>
<tr>
<th>Relationship Type</th>
<th>Mean Transfer Acceptability (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Someone standing at the same bus stop</td>
<td>1.72 (0.86)</td>
</tr>
<tr>
<td>Someone of the same race</td>
<td>1.72 (0.89)</td>
</tr>
<tr>
<td>Someone who is the same gender</td>
<td>1.82 (0.91)</td>
</tr>
<tr>
<td>Someone who lives in the same large city</td>
<td>1.85 (0.90)</td>
</tr>
<tr>
<td>Someone shopping at the same grocery store</td>
<td>1.88 (0.90)</td>
</tr>
<tr>
<td>Someone in the same large lecture class</td>
<td>1.88 (0.88)</td>
</tr>
<tr>
<td>Their co-worker</td>
<td>1.91 (0.88)</td>
</tr>
<tr>
<td>Their employee</td>
<td>2.01 (0.90)</td>
</tr>
<tr>
<td>Their cousin</td>
<td>2.02 (0.90)</td>
</tr>
<tr>
<td>Someone in the same small lecture class</td>
<td>2.05 (0.94)</td>
</tr>
<tr>
<td>Someone who lives in the same neighborhood</td>
<td>2.07 (0.95)</td>
</tr>
<tr>
<td>Their employer</td>
<td>2.08 (0.96)</td>
</tr>
<tr>
<td>Someone who lives in the same small town</td>
<td>2.10 (0.96)</td>
</tr>
<tr>
<td>Someone shopping at the same clothing store</td>
<td>2.11 (0.97)</td>
</tr>
<tr>
<td>Someone on the same sports team</td>
<td>2.22 (1.00)</td>
</tr>
<tr>
<td>Someone who likes the same music</td>
<td>2.31 (1.06)</td>
</tr>
<tr>
<td>Someone who attends the same church</td>
<td>2.44 (1.05)</td>
</tr>
<tr>
<td>Their sibling</td>
<td>2.44 (1.06)</td>
</tr>
<tr>
<td>Their child</td>
<td>2.54 (1.12)</td>
</tr>
<tr>
<td>Someone who belongs to the same political party</td>
<td>2.60 (1.08)</td>
</tr>
<tr>
<td>Someone who belongs to the same sorority</td>
<td>2.63 (1.11)</td>
</tr>
<tr>
<td>Their parent</td>
<td>2.65 (1.12)</td>
</tr>
<tr>
<td>Their romantic partner</td>
<td>2.78 (1.14)</td>
</tr>
<tr>
<td>Their best friend</td>
<td>2.80 (1.15)</td>
</tr>
</tbody>
</table>

There are a few observations about these lay beliefs that I would like to highlight, with a clear caveat that these are largely post-hoc explanations. Overall, people see it as unacceptable to use information about one person to inform their opinion of another; there is no relationship type for which the average response reaches the neutral scale mid-point. This finding is consistent with the idea that most people find it unreasonable and immoral to use information about one person to judge another (Banaji & Bhaskar, 2000), and with the observation from my lab that people report their judgements about the acceptability of attitude transfer being more influenced
by logic ($M = 3.40$) and fairness ($M = 3.39$) than by social norms or conventions ($M = 2.58$), 
$d_{\text{diff}} > 0.20$ (scale = 1 to 5).

In addition, people who agree that attitude transfer is acceptable for one relationship are likely to agree for other relationships. Treating the ratings as a scale, the reliability is $\alpha = .96$. The intraclass correlation (ICC) between scale items is $.49 (p < .0001)$. That said, there is variability in the extent to which ratings correlate with one another; the range of correlations between any two of the 24 items ranges from $r = .25$ (best friends and people at the same bus stop) at the lowest to $r = .68$ (parent and child; $p < .0001$) at the highest.

**Individual Differences in Lay Beliefs about the Acceptability of Attitude Transfer**

In the study for which data are presented in Table 1, neither overall acceptability ratings nor discrimination between the relationships with the highest (best friends) and lowest (people standing at the same bus stop) ratings are related to participant age, gender, education, racial/ethnic group, political orientation or religious identity.

Although demographics do not seem to moderate belief about the acceptability of attitude transfer, ideological and motivational orientations do. For example, there is a small, positive correlation ($r$ ranging from .18 to .24) between political conservatism and the extent to which attitude transfer is perceived to be acceptable and the extent to which information about one group member is used to inform evaluations of another. Other factors that are associated with beliefs about the acceptability of attitude transfer in our unpublished data include Personal Attitude Stability (the extent to which people believe their own attitudes are stable; Xu et al., 2020), Personal Need for Structure (the desire to structure the world in simple, more manageable form; Thompson, Naccarato, Moskowitz, & Parker, 2001), Right-Wing Authoritarianism (a set of attitudes including dogmatism, a preference for conformity, willingness to coercively enforce
behavioral standards, punitiveness toward enemies, and strong concern with hierarchy; Altemeyer, 1988; Costello et al., 2021), and essentialist beliefs (the belief that group membership is immutable, described in more detail below; Haslam, Bastian, Bain & Kashima, 2006). We have focused largely on U.S. Americans in our work, though it is likely that beliefs about the acceptability of attitude transfer—like other naïve theories of impression formation (Shimizu, Lee, & Uleman, 2017) differs across cultures and subcultures.

*Entitativity and Lay Beliefs about the Acceptability of Attitude Transfer*

The pattern of acceptability ratings presented in Table 1 are consistent with theories about group *entitativity*—the extent to which a collection of individuals is perceived as being a coherent, unified entity (i.e., “groupiness”; Campbell, 1958). Judgements of entitativity reflect that some assemblages of people, like those standing at the same bus stop, are less likely to be seen as a single, meaningful unit than others, such as people playing together on the same sports team (Hamilton, 2022, this volume; Lickel et al., 2000). While judgements of group entitativity do have some overlap with judgments of similarity between group members (Dasgupta, Banaji, & Abelson, 1999), they are distinct (Crump et al., 2010). I will return to this point later in the chapter. There are two dominant approaches to understanding entitativity: The “essence-based” and “agency-based” approaches (Brewer, Hong, & Li, 2004; see Agadullina & Lovakov, 2018, for a meta-analysis and review).

*Essence-Based Entitativity.* According to the essence-based approach to entitativity (Crump et al., 2010), groups are high in entitativity to the extent that they are perceived as homogenous, particularly in physical or mental/trait characteristics. Haslam, Rothschild, and Ernst’s (2002) conception of essentialism is an example of the essence-based approach, focusing on the underlying, inherent nature of social categories. Consistent with this conceptualization,
participants generally see it as more acceptable to use information about one person to judge their family members than non-family, and the distance among family members matters; for example, transfer between siblings is seen as more acceptable than transfer between cousins (Cohen’s $d = 0.43$). This is consistent with the finding that mock jurors see criminal defendants as being more likely to be guilty to the extent that they are seen as similar—without providing a definition of similarity—to family members who have been convicted of a crime (Rerick, Livingston, & Miller, 2021).

In line with the essence-based approach to entitativity, we also measured endorsement of psychological essentialism—the belief that differences between (racial) groups are immutable and naturally occurring (Haslam et al., 2006). Participants completed an 8-item racial essentialism scale (Pauker, unpublished), which included items such as “knowing what race someone is tells you a lot about their abilities and traits” and “Race is determined by biological factors such as genes and hormones.” There was a significant, positive correlation between essentialist beliefs and ratings of the acceptability of using information about one person to form an impression of another person of the same race ($r = .28$, $p < .0001$). A comparable correlation was observed between essentialism and the average of all acceptability ratings ($r = .25$, $p < .0001$), suggesting that essentialist views of groups in general—and racial groups in particular—are related to judgements about the acceptability of attitude transfer.

**Agency-Based Entitativity.** The agency-based approach to entitativity considers the group's heterogeneity, motivations, intentions, and level of interaction among group members, with the latter factor argued to be the most important (Brewer et al., 2004). Group size factors into these judgments. For example, transfer acceptability ratings are higher when considering people living in a small town compared to a big city ($d = 0.27$) and students in a small lecture
class compared to a large one \((d = 0.19)\). Narrower categories also seem to promote transfer; for example, people believe it is more acceptable to judge people interchangeably who are shopping in the same clothing store compared to the same grocery store \((d = 0.25)\).

*Entitativity and Attitude Transfer Effects*

Crawford, Sherman, and Hamilton (2002) more directly demonstrated that the transfer of traits from one individual group member to another is dependent on perceived group entitativity. They presented participants with valenced information about one group member and then gave people the opportunity to evaluate a second group member. The group was manipulated between-subjects to be high in entitativity (i.e., made up of similar people with shared background, attitudes, and personalities) or low in entitativity (diverse people with different backgrounds, attitudes, and personalities). Participants treated the group members as interchangeable only when the group was perceived as highly entitative. Presumably, when a social group was thought to be large and diverse, participants recognized that one individual is not representative of all group members and resisted the transference of traits—and presumably the valenced evaluation implied by those traits—from one person to another person in the same group.

It is clear from this exploration of lay beliefs and acceptability judgments that there is variation in perceptions of how acceptable it is to use information about one person as the basis of evaluating another who is related in some way. However, a deliberate judgement that a particular association between two people is not a sufficient basis for judgements ignores that those two people *are* in fact associated, whether by group membership, identity, family relationship, shared interests, or proximity. Thus, even when attitude transfer is deemed unacceptable and deliberately resisted, we might expect generalization to occur anyway through
processes that are spontaneous (i.e., without instruction or intention to make them; Uleman, 1987) or difficult to control (see Moors & De Houwer, 2007, for an overview of the features of automaticity).

**Evidence for Attitude Transfer on Direct and Indirect Measures of Evaluation**

Participants in my earliest studies of attitude transfer (Ratliff & Nosek, 2008) were first exposed to an attitude formation paradigm in which they read behaviors performed by Reemolap, a member of the group *Laapians*, and Vabbenif, a member of the group *Niffians*; there is considerable evidence that people spontaneously form impressions of people based on their behavior (i.e., spontaneous trait inference; Uleman, Newman, & Moskowitz, 1996). One of these original group members performed predominantly positive behaviors and the other predominantly negative behaviors, manipulated between-subjects. Participants were then given minimal information about two new individuals, Bosaalap and Ibbonif, belonging to the same groups, for example:

*Ibbonif is a sculptor and very much enjoys gardening, hiking, and playing card games. Ibbonif is kind and thoughtful, but tends to be slightly greedy at times.*

*Bosaalap is a painter and very much enjoys cooking, hiking, and listening to music. Bosaalap is warm and considerate, but tends to be slightly dishonest at times.*

Pretesting showed these descriptions to be evaluated as similarly valenced; the descriptions were also randomized across participants, ensuring that any differences in attitudes toward the new people could be a function only of attitudes formed toward the original people from the induction phase.

To evoke low entitativity, we described the groups as being large, diverse, and made up of many kinds of people who do many kinds of activities; it was plainly stated that Bossalaap and Reemolap, and Ibbonif and Vabbenif, had never met one another. At the same time, the
group members in this study did share some physical features (see image below). Participants were then tasked with evaluating either Reemolap and Vabbenif (the original people) or Bosaalap and Ibbonif (the new people). Self-reported evaluations of the new people were not influenced by the behavioral descriptions of the original people; however, evaluations of the new people measured with the Implicit Association Test (IAT; Greenwald & Lai, 2020; Greenwald, McGhee, & Schwartz, 1998) were equal in strength and direction to evaluations of the original people, a phenomenon that, for now, we will refer to as implicit attitude transfer. This effect was replicated as part of the Reproducibility Project: Psychology (Open Science Collaboration, 2015). Together, we interpreted these findings as suggesting that, even in the absence of self-reported attitude transfer, we might expect generalization on indirect measures of attitudes that capture evaluations that are (relatively) more spontaneously generated or difficult to control.

**An Associative Learning Explanation for Attitude Transfer and Generalization Effects**

There were two assumptions that guided our original interpretation of these findings. The first assumption is that attitudes may be acquired through simple associative learning mechanisms (Cacioppo, Marshall-Goodell, Tassinary, & Petty, 1992; Eagly & Chaiken, 1993; Olson & Fazio, 2002; Walther, Nagengast, & Traselli, 2005; Uleman, Saribay, & Gonzalez, 2008). Associationism provides an explanation for the fact that much of our knowledge is more complex than a simple summary of direct experience. For example, you might speculate about someone’s personality—even if you have not met them—based on what you know about someone else who belongs to the same social group.

Associative learning is a descriptive term referring to the type of learning that occurs anytime a relationship is detected between multiple concepts or events in an organism’s environment (Shanks, 1995). The importance of associationism and associative learning in social
psychology is evident in the very definition of an attitude as an “association between a concept and an evaluation – positive or negative, favorable or unfavorable, or desirable or undesirable” (Fazio, 1986, p. 214). This assumption led to the hypothesis that, even if someone consciously rejects the idea of evaluating the new colleague based on evaluations of someone else from the same group, the association between them exists, and may lead to generalization anyway. This hypothesis is consistent with the Associative-Propositional Evaluation (APE) Model (Gawronski, 2022, this volume; Gawronski & Bodenhausen, 2008, 2011) whereby an associative learning mechanism leads to the automatic formation of mental associations between stimuli that are linked in some way. The APE model also posits a propositional mechanism whereby deliberate reasoning about the truth value of the linked stimuli is possible.

The second assumption guiding our original interpretation of the Ratliff and Nosek (2008) attitude transfer findings is related to the first—that propositional knowledge is better assessed through direct measures and associations through indirect measures. Direct measures of attitudes involve asking participants to self-report their evaluations. Indirect measures, on the other hand, either do not alert participants to what is being measured, or reduce participants’ deliberative control over their responses, even if they are aware of what is being measured (De Houwer, 2006).

Ratliff and Nosek’s (2008) findings—and subsequent findings replicating the attitude transfer effect (Chen & Ratliff, 2015; Hawkins & Ratliff, 2015; Ratliff & Nosek, 2011)—are consistent with a dual-process argument as follows: An association to the original and

---

3 This is where the term association gets confusing. The original and new group members are inherently associated because we have said they are members of the same group (i.e., they are associates). But here I refer to mental associations, links between group members, represented in the mind (De Houwer et al., 2013), whether or not an individual intends or wants for them to form.
new group members forms automatically based on their relationship (e.g., shared group membership, physical resemblance, temporal or special proximity of learning about them). This association occurs through generalization and/or second-order conditioning: generalization is where a response conditioned to one stimulus generalizes to other stimuli that are similar (see Till & Priluck, 2000), and second-order conditioning, sometimes called a spreading attitude effect, where an association between Stimulus A (Reemolap) and Stimulus B (negative behavior) and an association between Stimulus A (Reemolap) and Stimulus C (Bosaalap) leads to an association between Stimulus B (negative behavior) and Stimulus C (Bosaalap), despite their never being presented together (Walther, 2002). In either case, the new group members automatically take on the evaluation of the original group members. Thus, one assumption about associations is the formation of attitudes toward novel group members via their relationship with known group members—if a known group member (Reemolap) is viewed negatively, that negativity will transfer to a novel group member (Bosaalap).

We then assume that people can (and usually will; see Table 1) apply a rule that it is unfair or illogical to evaluate one member of a large, diverse group based on another person’s actions, and so self-reported attitudes reflect differentiation between the new and old group members; that is, we deliberately reject the negative evaluation of group member B. A second assumption about associations now comes into play—that lingering negativity toward group member B can be assessed with the IAT, because that measure is particularly suited to assess

---

Recent arguments point to the importance of specifying what constitutes automaticity in a given context (Van Dessel, 2020). Candidate characteristics of automaticity include responses that occur in ways that are uncontrolled, unaware, efficient, or fast (Moors & De Houwer, 2006); in this case, intentionality is the most relevant factor. Intentionality refers to the extent to which performance on a measure can be controlled when motivated to do so. In early attitude transfer studies we assumed a failure of intentionality due to the dissociation between what people self-reported (i.e. no attitude transfer) and what their performance on the IAT (i.e. attitude transfer); in later studies we directly manipulated intentionality (e.g., Hawkins & Ratliff, 2015), a point I return to momentarily.
association between concepts (e.g., Reemolap, Vabbenif) and evaluative attributes (e.g., Good, Bad).

To recap the argument, self-reported evaluations of the new people were not influenced by the behavioral descriptions of the original people; however, evaluations of the new people measured with the IAT were equal in strength and direction to evaluations of the original people. We made two assumptions about associations to explain these findings—that the positivity or negativity of original group members transferred to new group members via their association, and that the association between the new group member and positivity or negativity could be assessed by the IAT. The original data and subsequent studies are consistent with these assumptions, as described below.

Ratliff and Nosek (2008) found that, after a delay of several days, self-reported evaluations showed evidence of generalization too, suggesting that as resources to prevent generalization (e.g., clear knowledge of who did what) decline, simple associations may play a bigger role in evaluations of new group members. Further support for this interpretation came from subsequent studies showing that there was a stronger correlation between the IAT and self-reported evaluations of the new people at Time 2 than existed immediately after learning about the group members (Ratliff & Storbeck, unpublished data). These findings are consistent with findings that cognitive resources are important for social inference; for example, Kubota et al. (2014) showed that people were more likely to make dispositional judgments about people’s behavior when experiencing physiological stress that disrupted executive functioning.

Consistent with other demonstrations of difficulty preventing formation of conditioned attitudes on indirect measures (Gawronski, Balas, & Creighton, 2013) manipulations to induce intentional control over attitude transfer on the IAT failed to do so. In five studies, Hawkins and
Ratliff (2015) demonstrated that IAT scores assessing attitudes toward new consumer products (studies 1 and 2; see also Ratliff, Swinkels, Klerx, & Nosek, 2012) or new group members (studies 3-5) showed evidence of attitude transfer even when participants were explicitly instructed to be fair and to avoid generalization. An accountability manipulation telling participants that they would be required to explain their decision at the end of the study (e.g., Thompson, Roman, Moskowitz, Chaiken, & Bargh, 1994) did not prevent implicit attitude transfer, nor did priming egalitarian goals using an objectivity writing task (Moskowitz & Li, 2011), regardless of whether the objectivity manipulation was presented before the attitude induction, after the attitude induction but before introduction to new group members, or after the attitude induction and introduction to new group members but before completing dependent measures. Two additional studies manipulated the relationship between the original and new people to disrupt implicit attitude transfer; however, there were still transfer effects on the IAT when the original and new people were described as enemies (study 7) or as strangers at a bus stop (study 8).

Internal meta-analysis confirmed that, in all eight studies, implicit and explicit evaluations formed toward novel consumer products and social group members. Although these attitudes influenced subsequent evaluations of new consumer products, the transfer of evaluations to new people was largely avoided in those studies with social groups. Implicit evaluations, on the other hand, transferred readily between consumer products and social group members, even in those conditions where deliberate control processes should have been engaged.

---

5 In these studies, we inferred from the pattern of self-reported evaluations that people intended to avoid attitude transfer; however, it is possible that these online studies with volunteer participants simply did not engage the kind of strong commitment to egalitarian goals that would be necessary to control more spontaneous attitude transfer.
Challenges to an Associative Learning Explanation for Attitude Transfer and Generalization Effects

In recent years, both assumptions used to explain basic attitude transfer effects on direct and indirect measures—that attitude generalization can be explained through “simple” associative learning mechanisms and that the IAT measures associations between evaluations and attitude objects—have been convincingly challenged (Bading, Stahl, & Rothermund, 2020; Corneille, Mierop, Stahl, & Hüttter, 2019; De Houwer, 2014; De Houwer 2019; Högden & Unkelbach, 2020; Van Dessel et al., 2020). Challenges to the assumption that attitude formation and generalization can be explained primarily through associative learning stem largely from studies of evaluative conditioning, which can be functionally defined as a change in liking of a neutral stimulus (conditioned stimulus; CS) because of its pairing with a valenced stimulus (unconditioned stimulus; UCS; Gast, Gawronski, & De Houwer, 2012).

In the attitude transfer paradigm described previously (Chen & Ratliff, 2015; Hawkins & Ratliff, 2015; Ratliff & Nosek, 2008; 2011), EC-like effects are observed in evaluations of the original people—the initial targets of the attitude formation paradigm used to induce attitudes toward the original group members before the introduction of new people from the same group. Both implicit and explicit attitudes are consistent with the valence of the information presented in the induction; that is, attitudes toward the group described as performing predominantly

---

Thus, the conclusion from these studies is better summarized as “attitude transfer is difficult to control” than as “attitude transfer cannot be controlled”.

6 In a typical evaluative conditioning paradigm, neutral stimuli (UCS) are paired with valenced stimuli (CS) though typically in an incidental way (e.g., they appear on the screen at the same time). In our attitude induction paradigm (adapted from Gregg, Seibt, & Banaji, 2006), participants are given descriptive behavioral information about the group members that directly implies trait characteristics (e.g., Reemolap helped an elderly person cross the street). So the group member is indeed “paired” with the behavioral information, though in a way that is not common for an EC study.
positive behaviors are more positive than those toward the group described as performing predominantly negative behaviors; these effects are generally very large ($d > 1.0$).

It is generally assumed that conditioning effects are due to the formation of associations between the CS and UCS (De Houwer, 2018). The associative learning hypothesis has strong face validity (Corneille & Stahl, 2019) and empirical support. For example, Hu, Gawronski, & Balas (2017) simultaneously presented participants with pharmaceutical products and negative health conditions, varying whether the product was described as causing or preventing the health condition. Although self-reported evaluations of the pharmaceutical product reflected the relation between the condition, responses on an indirect measure reflected the co-occurrence of the product with the health condition regardless of the relationship (see Moran & Bar-Anan, 2013, for similar findings). More recently, and consistent with the finding from Hawkins and Ratliff (2015), Gawronski and Brannon (2021) found that instructions to counteract stimulus co-occurrence effects were ineffective in preventing evaluative conditioning. Together, these results support the idea that “mere” associations between stimuli could lead to attitude transfer between stimuli that are related in some way.

On the other hand, behavioral and physiological evidence suggests that evaluative conditioning is mediated by propositional representations that specify knowledge about how stimuli are related to one another (relational knowledge; De Houwer, 2018, 2019; Mitchell, De Houwer, & Lovibond, 2009)—for example, $A$ predicts $B$, $A$ causes $B$, $A$ goes with $B$, $A$ prevents $B$, $A$ is like $B$, $A$ is opposite $B$, etc. Whether (and how) the valence of one stimulus influences evaluations of another stimulus may depend on this propositional knowledge about how the stimuli are related to one another.
For example, Zanon, De Houwer, Gast, & Smith (2014; Experiment 1) presented participants with one set of non-words paired with positive stimuli and another set paired with negative stimuli, and then told them that the meaning of the non-words was the opposite that of the stimuli with which it was paired. For associative learning, this relational information should not matter – an IAT assessing evaluations of the novel word should reflect the valence of the paired word; however, participants showed an IAT effect demonstrating more positivity toward the non-words paired with negative over those paired with positive, thereby reflecting the instructed pairing (i.e., a proposition) rather than the observed pairing (i.e., an association). Along these lines, Kurdi, Krosch, and Ferguson (2020) found that implicit evaluations of novel moral agents reflected the valence of the outcome their actions produced but also inferences about the actor’s mental state; for example, IAT scores reflected more negativity toward someone who intentionally caused another person to fall off of a bridge compared to someone who caused the same outcome without meaning to do so. Finally, mere instructions about a learning procedure leads to IAT scores that are at least as large as experiencing the actual learning procedure (e.g., Kurdi & Banaji, 2017; Smith, Calanchini, Hughes, Van Dessel, & De Houwer, 2019; Van Dessel, De Houwer, & Smith, 2018). Taken together, these and other studies provide strong evidence that implicit measures do not only reflect simple associations (e.g., object + good), but can incorporate additional logical information.

A Propositional Account of Attitude Transfer Effects

It does not require extensive mental gymnastics to reinterpret attitude transfer effects in accordance with prominent propositional models like the Integrated Propositional Model (IPM; De Houwer, 2014, 2018). According to the IMP, and as described previously, a proposition is a mental representation of information about how stimuli are related (e.g., A predicts B, A causes
B, A goes with B, A prevents B, A is like B). Propositions have inherent truth value (i.e., the potential to be true or untrue) which can be accepted or rejected in any given situation through inferential reasoning. For example, propositions about the relationship between group members in the Ratliff and Nosek (2008) attitude transfer paradigm might be Bosaalap goes with Reemolap, or Bosaalap is like Reemolap. That proposition would be accepted or rejected based on what one believes to be true about the acceptability of using information about Reemolap to evaluate Bosaalap. For example, the proposition that Reemolap goes with or is like Bosaalap may be accepted if Reemolap and Bosaalap belong to a group that is more entitative compared to a less entitative group (Hawkins & Ratliff, 2015), an outgroup to the perceiver compared to an ingroup (Chen & Ratliff, 2012), or a group that is joined voluntarily (e.g., choosing to join the military) compared to one that is not (e.g., drafted into the military; Vitiello & Ratliff, unpublished data). This inferential reasoning about the propositions is then reflected in self-reported evaluations of the new person where it is either accepted (i.e., if Reemolap is like Bosaalap) or rejected (i.e., Reemolap is not like Bosaalap).

So far, there is no divergence in what the IPM and dual-process or associative models, including the APE model (Gawronski & Bodenhau sen, 2006, 20011), would predict. But on its surface the IPM would predict sensitivity to deliberate reasoning about group members to be reflected on both self-report and the IAT. So why does the IAT reflect guilt-by-association even when inferential reasoning about propositions would say otherwise? De Houwer (2014) describes several ways in which the IMP can account for effects that might appear associative, for example, discrepancies between behavior (in this case, IAT performance) and people’s self-reported propositions (in this case, consciously rejecting the idea that information about one group member should be used to evaluate another group member).
The IPM assumes that propositions are activated automatically from memory, which allows propositional models to mimic the behavior of an associative-based network. Thus, it is possible to obtain partial automatic retrieval of the proposition; that is, retrieval of propositions form memory may be incomplete. In the case of attitude transfer, someone may have formed and memorized the proposition that Reemolap is not like Bosaalap, which is what they self-report, but the “not” is dropped during automatic retrieval and what is retrieved from memory during performance of the IAT is instead that Reemolap is like Bosaalap. De Houwer (2014) also proposes the possibility that automatically retrieving an older or rejected proposition from memory can lead to effects that run counter to currently held propositions. For example, the general proposition that things that look similar behave similarly, or that things in proximity to one another are similar may be rejected in favor of a more egalitarian proposition, such as things that look similar sometimes behave differently. So, although Reemolap is like Bosaalap is an older, rejected proposition, it may be retrieved from memory during IAT performance nonetheless. It could also be that the things that look similar behave similarly proposition is more well-rehearsed and, validated more frequently, or is a more general rule than the less practiced, exception-to-the-rule proposition that things that look similar sometimes behave differently; these may be other ways in which one proposition may be feeblter—and therefore less likely to be automatically activated—than another.

The IMP has been criticized for conflating the concepts of consciousness, intentionality, and effort within the umbrella of propositional reasoning (Uleman, 2009) and for being too flexible to be falsifiable (Kurdi & Dunham, 2020), and others have argued that the IMP can better explain conditioning though it does offer an alternative explanation for attitude transfer effects that generates novel hypotheses about when we would expect to see attitude transfer
effects. Importantly, however, although this model presents a challenge to our original assumptions about the associative nature of attitude transfer effects, and the ability of the IAT to assess newly formed associations between evaluations and group members, it does not undermine the existence of attitude transfer as an effect that can help us to understand the formation and maintenance of group-based attitudes and stereotypes.

**Similarity, Categorization, and Valence Effects in Attitude Transfer**

*Similarity between Stimuli Impacts Generalization.* The role of similarity in generalization is a topic big enough to carry its own chapter, so here I will focus only on select findings that are particularly relevant for impression formation. Fazio, Eiser, and Shook (2004) created BeanFest, a computer game where the goal is to accumulate points by making correct decisions about which beans to approach and which to avoid. Approaching a positive bean increases points and approaching a negative bean decreases points. During the game, the beans are viewed on a 10 X 10 matrix where the x-dimension is the shape of the bean (ranging from perfectly circular to oblong) and the y-dimension is the number of speckles the bean has (ranging from one to ten). Thirty-six beans, in six regions of the board, were presented to the participants; each selected bean produced a positive or negative outcome after it was presented. Generalization was measured by having participants evaluate 64 novel beans that varied in similarity to the known beans (via Euclidean distance from the novel bean to the nearest bean in the matrix).

The results clearly demonstrated a generalization effect: The more that novel beans visually resembled known beans, the more they were evaluated similarly to the known beans (Fazio et al., 2004). This finding is consistent with early work on generalization gradients in
learning, and also with the *shared features principle*, which refers to the idea that when stimuli share one feature, people often assume they share others as well (Hughes, De Houwer, Mattaveli, & Hussey, 2020). By this logic, if two individuals share some features (e.g., physical resemblance, group membership), people will assume they share others. In this case, if behavioral valence is the only other information one has about one group member, it makes sense that it would be the feature shared between individuals. Attention to shared features is also important for generalization; generalization from Stimulus A to Stimulus B was stronger when the two stimuli were similar on dimension to which participants were instructed to pay attention compared to when the two stimuli were similar on a less salient dimension (Spruyt, Klauer, Gast, De Schryver, & De Houwer, 2013). Alves et al. (2020) also found differentiation effects, where generalization occurs more strongly between stimuli that are distinctly related to one another. Based on this, we might expect stronger attitude transfer among group members who are members of only one group compared to those with multiple group memberships.

There are also open questions about the role of actual versus perceived similarity in attitude transfer/generalization. Verosky and Todorov (2010) demonstrated that attitude generalization increased as a function of objective facial similarity (i.e., novel faces morphed with known faces at 20% or 35%; see also Kraus & Chen, 2010, and Günaydin et al., 2012). This linear effect is consistent with classic learning theory’s emphasis on generalization gradients (Klein, 2019). On the other hand, Gawronski and Quinn (2013) showed that—on self-report and an evaluative priming task (Fazio, Jackson, Dunton, & Williams, 1995)—the valence of a known individual equally generalized to a novel individual whose face was morphed at 50% or 100% resemblance to the original face. There is also evidence that, for White perceivers, IAT performance indicates that evaluations of a novel group member reflect greater attitude transfer
from an original group member when the group members are Black relative to White (Ratliff & Nosek, 2011). This may be due to an outgroup homogeneity effect where the Black group members seem to visually resemble one another though they actually do not. It is also possible that, consistent with entitativity effects described previously, groups with members that are perceived as being more visually similar to one another are also seen as more entitative, coherent, and unified (i.e., “group-y”), or as sharing an essence.

There are still other examples where no similarity at all seems to be required for attitude transfer to occur. For example, Hebl and Mannix (2003) found that a male individual seated next to an overweight woman was denigrated in a hiring context substantially more than if he was seated next to an average-sized woman. This effect held even when it was made clear that there was no relationship between the applicant and the overweight woman, suggesting that, at least in some cases, proximity is all that is required to observe attitude transfer effects (see also Hawkins & Ratliff, 2015; Pryor, Reeder, & Monroe, 2012).

One possibility is that generalization and higher-order conditioning are separable processes by which attitudes transfer from one group member to another. A generalization explanation for attitude transfer effects would predict that more objective resemblance between new and old group members should lead to stronger transfer effects. In higher-order conditioning a stimulus only has to be associated in some way with another stimulus to take on its valence, for example, by sharing category membership, whether or not it is similar to that original (Walther, 2002). An interesting avenue for future exploration would be to compare the extent of attitude transfer based on different types of group member similarity (e.g., visual resemblance vs. shared beliefs vs. biological relationships). Such tests would be interesting for better understanding
transfer effects but could also shed light on the role of associations versus propositions in attitude transfer.

**Stimulus Valence.** Negative information is generally more influential than positive information in evaluation (Cacioppo, Gardner, & Berntson, 1997). For example, little unfavorable information is needed to confirm a negative stereotype about a group, but quite a bit of favorable information is needed to form a positive stereotype or to disconfirm a negative stereotype (Rothbart & Park, 1986). Negativity also seems to be more “contagious” than positivity (Rozin & Royzman, 2001; Boydstun, Ledgerwood, & Sparks, 2019). For example, most people refuse to wear a sweater believe to be worn by Adolf Hitler, even if Mother Theresa also wore it. Early evidence that negative evaluations also generalize more readily than positive evaluations came from follow-up studies using Fazio and colleagues’ BeanFest paradigm. Overall, generalization is moderated by information extremity and valence (Shook, Fazio, & Eiser, 2007). Similarity to beans that produced a more extreme point gain or loss mattered more than similarity to those with a more tempered outcome, but this was less true for negative beans; that is, it takes less similarity to a negative bean to be judged negatively than it takes similarity to a positive bean to produce an equally strong positive evaluative outcome. Ratliff and Nosek (2011, Study 1) observed similar effects using the novel groups attitude transfer paradigm (Ratliff & Nosek, 2008). Compared to positive information, one group member’s negative behavior had a stronger influence on evaluations of a novel group member; this effect was observed on both self-report measures (a small effect) and the IAT (large effect size).

Although there is a general tendency toward negativity bias in attitude generalization, there is also variability (Fazio, Pietri, Rocklage, & Shook, 2015). Negative attitudes generalize more strongly for some people, positive for some people, and others show no asymmetry in
valence weighting. Pietri et al. (2013) demonstrated that this weighting bias in generalization is related to behavioral manifestations of rejection sensitivity (level of concern about and perceived likelihood of interpersonal rejection; Downey and Feldman, 1996), threat assessment (judgments of the likelihood that an ambiguous situation will become negative or threatening; Riskind et al., 2000), and risk tolerance (a preference for high-risk/high-reward over low-risk/low-reward options; Wallach, Kogan, & Bem, 1962). Correlations between these measures and valence weighing are not high (ranging from $r = .22$ to $r = .38$), but it is noteworthy that attitude generalization in a novel computer game would relate to performance-based measures at all.

**Stereotyping and Other Related Phenomenon**

A variety of other phenomenon may be related to attitude transfer but are just outside the scope of this chapter. Anderson and colleagues’ (Andersen, Glassman, and Chen, 1995) work builds on the Freudian idea of transference where a new person activates a representation of a significant other based on one or more common features, and then the new person is assumed to share traits with the significant other. For example, participants evaluate a target person who is physically similar to their romantic partner as having the same personality traits as the romantic partner (Glassman & Anderson, 1999). Work on cognitive balance (Heider, 1958) may also be relevant in person perception. For example, the motivation to maintain coherence among attitudes would predict that people will like others who are liked by those they feel positively about or who are disliked by those they feel negatively about, and that people will dislike those who are disliked by people they feel positively about or who are liked by those they feel negatively about; Gawronski, Walther, and Blank (2005) found just this. Another related phenomenon is spontaneous trait transfer (STT; Skowronski, Carlston, Mae, & Crawford, 1998). In STT, which is frequently described as an associative process (Uleman et al., 2008), the
valence of the information that Person A uses to describe Person B transfers to Person A, and thus Person C would evaluate Persons A and B similarly.

And of course, attitude transfer is highly related to stereotyping. Many explanations of group-based stereotyping explain stereotyping as a hierarchical generalization from beliefs about the group to beliefs about the individuals within that group. For example, Secord (1959) defines stereotyping as “a categorical response, i.e. membership is sufficient to evoke the judgment that the stimulus person possesses all the attributes belonging to that category” (p. 309). Others define stereotypes as characteristics that are associated with either a group (Katz & Braly, 1933) or as characteristics that are associated with an individual due to her or his group membership (Hamilton & Trolier, 1986). Stereotyping, like other forms of generalization, can be either accurate or inaccurate. Part of the purpose of a superordinate group is to described features that are shared among members of that group (Medin, Goldstone, & Genter, 1993; Tversky, 1977); however, in the case of stereotyping, those features may be distorted or erroneously applied to group members who do not share them.

The studies I have described here focus primarily on a person-to-person attitude transfer; however, we might also consider the possibility of person-to-group transfer and how such transfer processes might contribute to the formation and maintenance of stereotypes. Hamilton et al. (2015) demonstrated that perceivers draw spontaneous trait inferences about groups just as they do about individuals. Further, evaluations of a single group member influence evaluations of the group itself. Henderson-King and Nisbett (1996) demonstrated that participants in a study with an unkind Black confederate were subsequently less likely to sit near a different Black confederate. Fazio and Olson (2006) used an evaluative conditioning paradigm in which they presented participants with faces of Black or White individuals paired with positive or negative
pictures; participants who saw Black faces paired with positive words, and White faces paired with negative words, were subsequently more positive toward different Black people; that is, evaluations of the individual exemplars generalized to the category. Similar exemplar-to-category generalization effects have been demonstrated in adults with fictitious aliens and employees at a company (Glaser & Kuchenbrandt, 2017). Stark, Flache, and Veenstra (2013) showed that negative attitudes toward an individual outgroup member contributed to students’ attitudes toward that person’s ethnic group, and Skinner, Olson, and Meltzoff (2020) found that adults’ biases in favor of or against one individual can influence children’s evaluations of the groups to which those adults belong. These kinds of person-to-group generalization processes may also contribute to the success (or failure) of intergroup contact, a strategy for prejudice reduction that involved bringing together people from different groups to increase interaction (Allport, Clark, & Pettigrew, 1954; see Levy-Paluck, Porat, Clark, & Green, 2020 for a review).

Concluding Remarks

In Don Quixote, Miguel de Cervantes wrote: Tell me what company you keep, and I will tell you what you are (translated from Spanish). You probably noticed that I used a lot of related terms throughout the chapter: attitude transfer, attitude generalization, spreading attitude effect, higher-order conditioning, transference, stigma-by-association, etc. Although each of these has slightly different meanings, and is the preferred nomenclature in different research traditions, they all share the same fundamental idea—that evaluations of one individual may transfer to another who is related in some way. Sometimes this is intentional and people believe it is acceptable to judge people based on the actions of another. Many would agree, say, that choosing a known ax murderer as a best friend says something about you as a person. We would also probably agree that having one time stood next to an ax murderer at a public bus stop probably
does not say anything about who you are as a person (other than that you are lucky to have survived the encounter!). But there is plenty of evidence that people’s actual evaluations and behaviors are less concerned with deliberate judgements of fairness and may be influenced by relationships between group members despite intentions to the contrary.

Attitude transfer can have substantive real-world consequences. Imagine a teacher who judges a student’s academic performance based on the grades of their friends, or a jury that judges a criminal based on whether he has family members who have been convicted of a crime (as in Rerick, Livingston, & Miller, 2021). Prior collaborators of scientists who are guilty of scientific fraud—who themselves have no connection at all to misconduct—face a citation penalty of ~9% in the aftermath (Hussing & Pellens, 2018). In very tangible ways, attitude transfer matters. Further, transfer of attitudes from one person to another, or from one person to a whole group, could be a key mechanism in the formation of intergroup attitudes and can help to explain how stereotypes and prejudices are maintained. Further, on the positive side, the observation that attitudes toward groups and group members can be changed through evaluations of single group members could be a promising strategy for prejudice and stereotype-reduction interventions.

Author’s Note:

I would like to thank Colin Tucker Smith and Sarah Olshan for their helpful comments on this chapter.
References


