

Running Head: Implicit Misattribution Model of EC

## **The Implicit Misattribution Model of Evaluative Conditioning**

David S. March & Michael A. Olson

University of Tennessee, USA

Russell H. Fazio

The Ohio State University, USA

*In press. Social Psychological Bulletin. Special issue on “Evaluative Conditioning – Theoretical Accounts”.*

Manuscript draft date: 4 June 2018

Word count (main text): 9187

Address correspondence to:

Michael A. Olson

Department of Psychology

University of Tennessee

olson@utk.edu

865-974-8264

### Abstract

Evaluative conditioning (EC) refers to a change in one's attitude toward an object based on its contiguous pairing with other positive or negative objects. EC can, in principle, occur through multiple mechanisms, some more and some less thoughtful. We argue that one relatively low-thought route through which EC produces evaluative change is implicit misattribution. Our Implicit Misattribution Model (IMM) is premised on research indicating: a) attributional thinking is pervasive and relatively automatic, b) affective experiences are pervasive and relatively automatic, and c) errors in automatic attributional processing can lead to misattribution of affect from one object to another, resulting in the latter object taking on the affect produced by the former. Research employing the "surveillance paradigm" we developed provides support for the model, particularly its key moderating variable, source confusability. We further discuss assumptions of the model in terms of encoding, storage, and retrieval/application of the conditioned attitude, as well as the role of contingency awareness and other central issues in the EC literature.

## The Implicit Misattribution Model of Evaluative Conditioning

### **Opening Observations**

Attitudes are positive and negative evaluations, stored in memory, that provide a summary of what is good and bad about all manner of objects, from people to products to positions on politics. Under specific conditions, they are reliable predictors of behavior. Thus, attitudes are an essential component to our understanding of the mind (Cooper, Blackman, & Keller, 2016). Naturally, a fundamental question surrounding attitudes is their origins.

The field of social psychology, once defined as the study of attitudes (Thomas & Znaniecki, 1918), has for decades pursued this question from a variety of theoretical views with correspondingly diverse methodological approaches (see Cooper et al., 2016; Maio & Haddock, 2014, for reviews). Attitude researchers widely agree that attitudes are determined by multiple mechanisms, from thoughtful, belief-based reasoning to simple, associative linkages (not to mention genetic and evolved mechanisms). The various theories of attitude formation make unique predictions about the conditions under which a given mechanism of attitude formation, as opposed to other mechanisms, might operate. Accordingly, the mechanisms through which attitudes form are inextricably linked to the contexts in which those mechanisms presumably operate.

This is also the case with evaluative conditioning (EC); it is an effect to be explained (De Houwer, 2007). Attitudes can come about through EC via the contiguous pairings of an attitude object (conditioned stimulus; CS) and positive or negative objects in the environment (unconditioned stimulus; US). Despite the common endpoint (attitude formation or change), such

effects can theoretically come about through a variety of mechanisms, and features of the context in which the CS and US are jointly experienced can facilitate the operation of one mechanism over others (see Jones, Olson, & Fazio, 2010). We believe that the latter point is particularly worth highlighting. In this special issue, where researchers articulate various models designed to explain the mechanisms underlying various EC effects, none do so in a contextual vacuum. Most EC researchers employ only one or a few of several established EC paradigms (e.g., the picture-picture paradigm: De Houwer, Baeyens, Vansteenwegen, & Eelen, 2000; the surveillance paradigm: Olson & Fazio, 2001). These paradigms are essentially contexts that can facilitate or inhibit the operation of one mechanism or another. We believe it is important to recognize that like all research investigating the origins of attitudes, the theoretical views that specify mechanisms underlying EC effects are not paradigm neutral. In other words, theories about mechanisms underlying EC must consider the paradigm in which those mechanisms operate. Mechanism and paradigm are inevitably entangled.

We see it as progress that this special issue is premised on the possibility of multiple mechanisms. Instead of much past research that offered simplistically broad statements about the qualities that characterize EC generally (e.g., that it does or does not require contingency awareness, or is or is not prone to extinction; see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010), the current issue allows for nuance that arguably better reflects the likelihood of multiple mechanisms, each with its own operating principles and conditions. That is, EC may or may not require contingency awareness or be prone to extinction, and the unique features of differing EC paradigms may lend themselves to some mechanisms over others.

In this article we will focus on a particular model of EC, our Implicit Misattribution Model (IMM; first discussed in Jones, Fazio, & Olson, 2009). Before we do, we want to be clear

about the implications of the above analysis. Our model does not claim to explain the universe of EC effects. Certainly, other models offer compelling explanations of the process(es) underlying EC under different conditions than those we typically study (e.g., De Houwer, 2009).

Furthermore, as paradigm and mechanism are dependent, it is important to note that the model was developed in the context of the EC paradigm we developed (Olson & Fazio 2001, 2002, 2006). Like any research paradigm, it was not developed arbitrarily; we had certain theoretical views and empirical goals in mind when creating it. In what follows, we address the key assumptions of the IMM, its specific hypotheses, and, critically, the relevance of the research paradigm in which we test them. We then consider the additional questions posed by the editors of the special issue.

### **The Implicit Misattribution Model (IMM)**

The IMM provides an account of how attitudes form as a result of the contiguous presentation of a CS and US. In brief, it assumes that people routinely engage in attributional (i.e., causal) processing of objects in their environment (Kelley & Michela, 1980). That is, they attempt to locate the causes of events and experiences—including the affective experiences that pervade daily life. Such processing often occurs implicitly (i.e., without intention or awareness of the process; see Gilbert, 1989, for a review). Critical to the IMM, attributional processing can go astray; people often fail to accurately identify the causes of affective experiences. When affect generated from some object in the environment (the US) is mistakenly and automatically attributed to some other object (the CS), EC occurs through implicit misattribution. The model argues that implicit misattribution of affect is more likely to the extent that one experiences what we have come to call “source confusability,” that is, ambiguity as to the origins of experienced

affect. We describe each of these components of the model, support for its underlying assumptions, and its implications for EC in more detail below.

It is telling that our model rests on notions of misattribution of affect rather than, say, the statistical or predictive learning notions that one finds in the traditional animal and human learning literature. Perhaps crudely, we embarked on our EC research not with any interest in whether some object might predict some other object or consequence, or form a “stimulus-stimulus” link in memory, but whether we could somehow get the affect from one to “bleed” onto the other. And, coupled with our own intuitions as well as research from the implicit learning literature in cognitive psychology (e.g., Reber, 1989), we suspected that this “bleeding” of affect could occur implicitly. If it did, not only would we be avoiding the “demand awareness” alternative explanation that dominated so much earlier EC research, we would also be identifying a potentially pervasive process: one is constantly being bombarded with both affect-laden and neutral stimuli in nature. If misattribution from the affective sort to the neutral sort occurs implicitly, we could all potentially have a great number of attitudes we neither intended to have nor knew we acquired. In that sense, implicit misattribution of affect might be a ubiquitous source of attitudes.

### **The Pervasiveness of (Mis)Attributions**

Among social psychologists, attribution and its specious cousin, misattribution, are both familiar and well-studied. Indeed, how people come to understand and interpret their environments is a foundational topic of social psychology (Ross & Fletcher, 1985). There is ample evidence, across decades of research, that humans engage in reasoning about the causes of events and experiences (e.g., Weiner, 1985). Most relevant to the present purposes, there is evidence that people engage in such reasoning automatically (Gilbert, 1989). For example, in the

domain of impression formation, people tend to see traits in behaviors, that is, they interpret peoples' behaviors in terms of causally relevant traits without having to walk through, in any deliberative fashion, what a given behavior might imply about the presence of some trait. As Gilbert, Pelham, and Krull (1988) put it, "We see Henry playing poker rather than simply moving his fingers, Herbert cheating rather than simply taking a card from his sleeve, and we are usually unaware of the inferential processes by which such categorizations are achieved" (p. 733). In short, attributional inferences are pervasive, spontaneous, operate highly efficiently, and can occur without awareness.

There is also compelling evidence that when it comes to attributional thinking, we regularly screw it up (Ross, 1977). Perhaps the best-known example is the fundamental attribution error, the failure to account for situational and contextual influences when determining the cause of an individual's behavior (Jones & Harris, 1967; see also Gawronski, 2004). In the context of the IMM, our focus is specifically on the attribution of affect, and there is strong evidence that we sometimes screw this up too. In a classic demonstration, Dutton and Aron (1974) led men crossing a precarious suspension bridge to misattribute their physiological arousal to a contiguous object: a woman positioned directly across it. The same woman aroused less affection among men crossing a more pedestrian bridge. Similarly, a sunny day leads people to report greater life satisfaction than a dreary day (unless their attention is drawn to the source of the contaminant, a point to which we will return; Schwarz and Clore, 1983). There are numerous additional examples of misattribution of affect, both in laboratory and field settings (see Cotton, 1981, for a review).

Research in other domains indicates that the most basic features of objects are open to misattribution, and, speaking to the implicitness of the process, that such errors can occur very

early in processing (Treisman & Gelade, 1980). In visual perception, “feature migration” occurs when a simple visual feature of an object (e.g., its color or shape) is misattributed to a contiguous object. For example, when briefly presented, the color of one letter can be misperceived as the color of a nearby letter (e.g., a green ‘T’ near a blue ‘L’ might be reported as appearing blue; Treisman & Schmidt, 1982). Attitude activation, because it occurs so early (Fazio, 2007), is also likely to precede perceptual binding, suggesting that when object delineation errs, attitudes, like features, may be confused. The attitude may be “assembled” with an object that did not evoke it. To be clear, we do not equate misattribution of affect with feature migration effects. We raise the analogy only to highlight how experiences—whether of basic physical properties or affective ones—are constructive, and that such constructive processes can err.

More relevant to evaluation, perceptual fluency accounts of mere exposure effects indicate that the processing efficiency of a novel visual stimulus increases with multiple exposures, and those exposures need not be ones people can report having seen. That processing fluency, in turn, is misattributed as and subsequently manifested as increased liking for the object (Reber, Schwarz, & Winkielman, 2004; Reber, Winkielman, & Schwarz, 1998; see also Winkielman & Cacioppo, 2001). Even subliminally presented valenced images have been shown to influence how positively individuals evaluate subsequently presented neutral symbols (Murphy & Zajonc, 1993), suggesting that one need not be aware of the source of the affect, let alone engage in some conscious causal reasoning, for misattribution to occur. Indeed, misattribution *must* be implicit if it is to be amiss. Thus, attributional processing is a regular feature of human information processing, it is rife with errors, and those errors can occur without intent or awareness.

### **The Pervasiveness of Affect**

So far, we have made the case for misattribution as a pervasive, low-level, automatic, and sometimes unconscious process that can occur in a variety of perceptual domains (e.g., visual perception, trait inferences, physiological arousal). But is there reason to suspect that implicit misattribution should be particularly likely in the domain of positive and negative affect? We believe the answer is a resounding ‘Yes!’ Why? Mainly, because people are regularly awash in positive and negative affect.

Previously acquired attitudes are pervasive sources of such affect. People have a lot of attitudes because attitudes are so functional for efficient engagement with the environment (Maio & Olson, 1999). As we have conceptualized them, attitudes are summary evaluations of objects that distill evaluative information in memory about objects to simple object-evaluation associations. Attitudes are particularly functional insofar as they provide “ready-aids” to efficient function, steering us toward helpful objects and away from harmful ones. They are “ready” if they are very accessible, that is, are activated automatically—inescapably—upon encountering the object (Fazio, 2007). Decades of work has shown that strong attitudes are indeed capable of automatic activation, that such activation results in the experience of positive and negative affect, and that the activation of such affect is generally unavoidable. The chocolate on the counter activates positive affect, and the garbage on the road activates negative affect, regardless of whether those objects are central to our attention or current goals (though some minimal attention is necessary for processing; Roskos-Ewoldsen & Fazio, 1992).

### **Source Confusability**

So far, we have spoken primarily on the ‘how’ question of EC according to the IMM. Specifically, we have argued that because attributional thinking is pervasive yet error prone, and that positive and negative affect are regularly experienced, people are prone to mistakenly

identify the source of their experienced affect as the CS rather than the US. However, to offer an answer to the ‘when’ question, that is, the conditions under which implicit misattribution occurs, we must introduce another component of the model: source confusability. Source confusability refers to anything, be it the person, the context, or the stimuli one is processing, that increases the difficulty of accurately identifying the source of experienced affect (see Jones et al., 2010). In the IMM, it is the primary moderating variable delineating the likelihood of misattribution.

Operational examples of source confusability include visual and temporal proximity and increased saccadic eye movements between CS and US. We will describe the details of these operationalizations below in a review of the research we have conducted with the surveillance paradigm to test the model.

Considering its importance to the model, it may be helpful to consider what might be considered the opposite of source confusability, something we might call “source clarity.” Take the Schwarz and Clore (1983) study, described earlier, documenting improved life outlook on sunny rather than dreary days. In another condition, these researchers pointed out the weather to respondents prior to posing the life outlook question. In this condition, the effect of weather on life outlook disappeared. Drawing respondents’ attention to the source of a potential judgment contaminant obviated its effect: misattribution interrupted. In a recent example, Ruys, Aarts, Papiés, Oikawa, & Oikawa (2012) ostensibly showed participants subliminal primes (there were no primes) followed by Chinese ideographs, in a variant of the affective misattribution procedure (Payne, Cheng, Govorun, & Stewart, 2005). Participants rated both the pleasantness of the supposed subliminal prime as well as the Chinese ideographs. When participants were told that the source of their affect toward both the subliminal picture and Chinese ideograph could only stem from one source (i.e., the pleasant or unpleasant subliminal prime or their internal affective

state), their ratings of the subliminal prime did not influence their ratings of the target (i.e., no misattribution occurred). But, when participants were told affective cues could reflect either of two sources, their ratings of the subliminal prime affected their ratings of the target ideograph (i.e., misattribution occurred). The salience of a potential judgmental contaminant thus reduces its potential to be confused with another possible source. Similarly, some sources of affect are, by their nature, extremely strong, salient or what we have called “evocative”: a brilliant red Lamborghini or someone crying hysterically, for example. Here, the source of affect is relatively clear compared to subtler, less attention-grabbing sources. And it is probably obvious that affect experienced on Wednesday or in Hoboken is unlikely to be attributed to an event on Saturday or in Knoxville. That is, as temporal or spatial distance between the experienced objects increases, the likelihood of source confusability decreases.

We hope this discussion helps flesh out our conception of source confusability. We would like to add, however, that the list we have provided is not exhaustive, and there are likely to be other yet undiscovered factors that influence source confusability. For now, the critical point is that source confusability is seen as the primary moderator of EC within the mechanism of implicit misattribution. Because the IMM developed through the surveillance paradigm, we describe it next, highlighting aspects we believe contribute to its facilitating misattribution, and summarize some research that supports the IMM.

### **Empirical Approach: The Surveillance Paradigm**

The paradigm derived neither from previous EC research, nor from any particular theory on learning per se. Instead, we confess it derived primarily from own meandering intuitions and discussions with (mostly social psychology) colleagues on how we might create a paradigm to facilitate the “bleeding” of affect from a known object (US) onto a novel one (CS). Only later, in

developing and testing the IMM, did we realize that such a vague metaphor might more precisely be described as misattribution. We also wished to mirror the “real world’s” tendency to regularly bombard the senses with a steady perceptual stream of objects, some of which systematically covary.

We aimed to develop a paradigm with certain features mostly intended to reduce the obviousness of the CS-US pairings. It was important that the pairings be non-obvious for two reasons: (1) we wanted to avoid alternative explanations regarding demand effects, and (2) we assumed that the bleeding of affect would be more likely if people failed to notice the actual source of any experienced affect. We incorporated several features that many other paradigms do not to reduce the obviousness of the pairings, as well as the obviousness of the overall research goal.

First, we created a compelling cover story. We told participants that we were interested in “attention and rapid responding,” and that they would be playing the role of a security guard, watching for suspicious activity. Participants learned that a stream of images, sometimes alone and sometimes in pairs, would appear on the screen, and that their task was to press a key when a pre-specified target item (which was neither CS nor US) appeared. Most images and image pairs were distractors, purportedly meant to make the task more challenging. We believed such a cover story, along with the primary target search task, would deflect attention from the CS-US pairings, yet ensure continued attention to the stimuli.

Secondly, we employed non-repeating USs; the only thing the USs paired with a given CS had in common was their valence.

Third, the USs were not the attention-grabbing, strongly evocative, positive and negative images most paradigms employ. Instead, they were relatively less evocative positive words and images.

We tried to increase the odds of that “bleeding” discussed above in other ways as well. “How best to classically-condition?” we asked our animal learning colleagues. The answer we received was as unanimous as it was counter-intuitive to us: forward pairings (where CS precedes US) are nearly universally superior to backward or simultaneous pairings. We immediately suspected we were chasing a different beast, so we decided on simultaneous CS-US pairings, again, based on intuition. Research we describe later verified our hunch.

A final point regards our choice of CS. Under the well-supported assumption that it is easier to create a new attitude than change a preexisting one, we wanted not only novel objects as CS, we wanted those objects to be potentially meaningful in an evaluative sense—objects toward which one might reasonably experience a positive or negative reaction (i.e., objects that might be a plausible source of experienced affect). Lesser-known Pokémon seemed to fit the bill.

In a typical study, after being told the cover story and introduced to the targets, participants undergo around 400 or so trials (each trial lasting around 1.5 seconds) of the surveillance task. As mentioned, the vast majority of images are fillers, with some targets, and a few blank screens, aimed at creating a sense of a non-rhythmic visual stream. Buried within the stream are the 40 or so CS-US total pairings involving each of 2 CS, one paired with positive USs and the other with negative USs. All participants are thus exposed to the same images; which CS is paired with positive vs. negative USs is counterbalanced between participants.

The paradigm has yielded a number of relevant findings over the years, first and foremost that it can successfully create attitudes (Olson & Fazio, 2001, 2002). Importantly, and unlike

most EC paradigms, most participants are unable to report the CS-US contingencies on both valence and identity memory measures, yet an EC effect is reliably observed among these participants on both direct and indirect measures.

### **Tests of the IMM**

A number of studies designed to test the IMM have been conducted using the surveillance paradigm. Some of these are reported in Jones, Fazio, & Olson (2009), and we will briefly review them here. First, in a correlational study, a remote eye-tracking device provided a means of counting the number of times participants' gaze shifted between CS and US while both were presented on the screen. We reasoned that the likelihood of source confusability would increase to the extent that CS and US were jointly attended to and processed. That is, when affect activated by the US is experienced at the same time one is visually processing the CS, the likelihood of confusing the source of such affect should be higher than if the two were processed at different points in time. Eye-tracking provides good insight into what people are attending to and processing (e.g., Kruschke, Kappenman, & Hetrick, 2005). Thus, we predicted that as the number of eye gaze shifts between CS and US increase, the opportunity for misattribution of affect should also increase, in turn leading to a stronger EC effect. Most participants did, in fact, show regular shifts between CS and US (after all, their task was to survey the presented stimuli to see if they matched the target for which they were to be vigilant), but those who showed a greater number of shifts evidenced stronger EC. A follow-up study attempted to manipulate the eye-gaze shift patterns experimentally. Specifically, in one condition the CS and US rapidly oscillated—they essentially flashed back and forth. Greater conditioning was observed in this condition than in one without the oscillation.

Jones et al. (2009) also report a study involving the manipulation of CS-US spatial proximity by either having CS and US appear very near each other on the screen, or further apart. Spatial processing (i.e., identifying the “where” of “what”) is relatively spontaneous, and is integrated with other information, including affect, early in processing. As two events occur nearer in space, the likelihood that their features will bind increases. For example, recall that feature migration occurs when an aspect of one object (e.g., color) is mistakenly perceived on a nearby object. Evidence shows that such mistakes are more likely to occur as those two objects increase in proximity (Treisman & Schmidt, 1982). Indeed, Jones et al. (2009) found stronger EC effects when the CS and US appeared closer together rather than further apart.

An analogous argument applies to temporal proximity. As we have described, most of our studies involve simultaneous rather than forward or backward CS-US pairings. Forward pairings are most effective at encouraging signal learning, as is the interest of most learning theories of animal behavior. There, CS predicts US, but the two are conceived separately. When presented simultaneously, however, the likelihood that any affect associated with the US is attributed to the CS would arguably increase. Evidence of the superiority of simultaneous presentations comes from a few lines of work. First, Rydell and Jones (2009) manipulated the onset and offset of positive and negative USs when a given CS was paired with both. Although negative USs had more impact than positive USs when both were presented at equal temporal distance, whichever US appeared most closely in time to the CS dominated; CSs more closely paired with positive than negative USs were viewed more positively, and CSs more closely paired with negative than positive USs were viewed more negatively. Secondly, and more directly, Hütter and Sweldens (2013) manipulated whether CS-US pairs were shown simultaneously or not and found that EC in the absence of memory for the contingencies

occurred only during simultaneous presentations. While we suspect that backward conditioning may be more amenable to producing misattribution than forward conditioning, we consider them both inferior to simultaneous presentations. That is, for misattribution to occur, the affect elicited by the US may need to still be sufficiently active during the experience of the CS. Simultaneous presentation obviously lends itself to this condition, as does a backward conditioning paradigm where affect is elicited immediately prior to CS presentation.

Recall that early in this research, we intuited that less “evocative” USs might more effectively lend themselves to misattribution than USs whose valence is extreme, salient, and obvious, and most of the experiments conducted using the surveillance paradigm used USs that were clearly positive or negative, but low in evocativeness. We manipulated US evocativeness in a study reported in Jones et al., and the results were informative. Among participants viewing CS-US pairings of low evocativeness, little contingency memory was reported, and among unaware participants, an EC effect was observed, as we expected. However, among participants viewing pairings involving highly evocative USs, many more participants were able to accurately recall the valence of the US appearing with a given CS. No EC effect was found among unaware participants in this condition, suggesting that no implicit misattribution occurred. Among the contingency aware, EC was observed, but we suspect that it came about through a different, likely propositional process, prompted by participants’ awareness of the CS-US pairings.

Another study reported in Jones et al. provides additional support for the importance of attributional processing in the surveillance paradigm. We made use of the well-documented phenomenon that the more salient (e.g., large, loud, bright, etc.) an object is, the more likely it is to be the recipient of a causal attribution (Pryor & Kriss, 1977). In this study, two versions of our

CS-US stimulus sets were created, one in which the CSs were relatively large and the USs were relatively small, and another in which the CSs were relatively small and the USs were relatively large. Perhaps counter-intuitively, the IMM predicts that the large-CS/small-US images should increase the probability of the CS being seen as the source of the experienced affect, as the CS is more salient in this condition. This is just what we found.

We hope this broad overview of the IMM and the surveillance paradigm provides a general conceptual understanding of them both. Next, we address the remaining specific questions posed by the editors of the special issue.

### **Specific Questions Regarding the IMM**

#### **How does exposure to stimulus co-occurrence change evaluative response?**

As described above, misattribution occurs any time the affect evoked by one object is erroneously attributed to another. Though some have questioned the ubiquity of such occurrences, we suspect they happen quite often and have important consequences (e.g., stigma by association; Pryor, Reeder, & Monroe, 2012). Using an example borrowed from Gast and Rothermund (2011), imagine a new neighbor moves into your apartment building, and prior to gaining personal knowledge about this person, you see them on several occasions speaking with your disliked landlord. You know that it makes sense for a new tenant to speak with their new landlord, yet after witnessing their interactions, you find that you also dislike your new neighbor. This is just one example of the myriad ways that misattribution may occur in “real-world” settings (see also Walther, 2002).

Also, as previously reviewed, selective attention to the CS can influence EC effects. By manipulating eye-gaze shifts and relative salience of either CS or US, one can effectively increase or decrease misattribution of affect. Manipulations of this sort likely affect the content

of the learning process, and not the learning process itself. Therefore, the integration of the CS and the evaluative response elicited by the US are influenced by whether the CS and US are processed in relation to one another. Misattribution occurs through the integration of the CS with the evaluative response elicited by the US. Meaning, simple co-occurrence is not sufficient to produce misattribution for if the capacity to experience affect elicited by the US is in any way lessened, either through task demands or other manipulation, misattribution is unlikely to occur.

### **What is stored in memory?**

Implicit misattribution involves the transfer not of the valence of the US to the CS (i.e., a CS-US or S-S hypothesis), but of the evaluative response (ER) elicited by the US to the CS (i.e., a CS-ER or S-R hypothesis). The IMM holds that such a CS-ER linkage is maintained in memory as part of the summary attitude, which in this case is an association between an object (the CS) and an automatic evaluative response (the ER). The surveillance task we use encourages evaluative responses to the US, and through repeated exposure to CS-US pairings, participants are likely to build an association linking a CS and an ER. The result of such an association may imply that EC resulting from IM is more resistant to US revaluation than other processes involving the encoding of CS-US associations.

Evidence for CS-ER learning comes from research showing that, in the context of some EC paradigms, US revaluation (i.e., when the valence of a US changes after conditioning) does not impact previously acquired EC effects (e.g., Baeyens, Vanhouche, Crombez, & Eelen, 1998). Of course, other studies employing very different learning paradigms have demonstrated US revaluation effects (e.g., Walther, Gawronski, Blank, & Langer, 2009), suggesting an S-S learning process. Our position is that both S-S and S-R learning can create EC effects, depending on the operating process, context, and stimuli.

**Are there important processes that occur over time between initial acquisition and evaluative response?**

The IMM does not make specific predictions regarding any processes that occur between initial acquisition and the ER. There are, however, some interesting points on which we can speculate. We have previously proposed that attitudes resulting from implicit misattribution are relatively affective in nature (Kendrick & Olson, 2012). Attitudes formed by misattribution are therefore more likely to resemble gut intuitions than rational beliefs. This has implications for several memory processes.

Foremost, it implies that consolidation – typically meaning the transfer of some connection from short- to long-term memory within a few hours of sensory input – can take place rather quickly, but that such attitudes may only initially be expressed under certain conditions. As time passes, however, such EC effects, and their associated attitudes, may strengthen. For example, in some of our unpublished work, we have found that EC effects created through misattribution strengthen after a ten-minute delay. Also, a recent meta-analysis (Gawronski, Gast, & DeHouwer, 2015) showed that unreinforced CS presentations diminished EC effects on self-reported evaluations but had no effect on EC effects measured through an evaluative priming measure. Though speculative, this may imply that CS-ER links consolidate rather quickly and persist over time.

**Does EC depend on automatic or deliberate processes during acquisition?**

The misattribution of affect proposed by the IMM will generally occur automatically. That is not to say that deliberate (i.e., propositional) processes cannot co-occur, but that misattribution *must* be implicit as by its very nature, it is amiss, and not privy to the corrective function of propositional reasoning. The model does not suggest that IM occurs inescapably or inevitably,

but the model is clear that IM is incidental, that is, unintentional. Indeed, without source confusion of the origin of experienced affect, IM is unlikely to occur. Recall Ruys et al.'s (2012) findings that only by creating ambiguity in the source of affect would misattribution occur. As discussed above, as source confusability increases, EC effects are more likely to occur. The inverse also applies – as source clarity increases EC effects through IM are less likely to occur. Therefore, the IMM predicts that EC effects resulting from IM depend more on automatic than deliberate processes during acquisition.

We also assume, like attributional processing broadly, that misattribution is efficient, requiring little in the way of cognitive resources. This is not to say, however, that it should be impervious to attentional demands—every cognitive process requires some amount of resources. Some attentional allocation, for example, is necessary for the US to be noticed sufficiently for the activation of its associated affect, and some attention must be allocated to the CS for misattribution of affect to occur. It should also be noted that attention and awareness are not synonymous (e.g., Cohen, Cavanagh, Chun, & Nakayama, 2012). Attention merely refers to some amount of prioritization of processing in this context. We should also note that misattribution can certainly occur as a function of deliberate, careful thought, but given how little participants notice and recall about the pairings in our typical experiments, we consider thoughtful misattribution unlikely in this context.

In the EC literature, questions of automatic vs. deliberate processing tend to evoke the debate regarding associative versus propositional processing (e.g., Mitchell, De Houwer, & Lovibond, 2009). First, it is important to distinguish between how a stimulus is processed from how it is ultimately represented in memory. Although we claim that implicit misattribution results in an object-evaluation association in memory, we are agnostic as to whether the process

that results in that association is best labeled “associative” or “propositional.” We believe the associative-propositional distinction to have lost explanatory value, with proponents of one process readily able to account for findings claimed to be driven by the other. For example, association-learning has been considered a relatively resource-efficient process, but one that only leads to gradual changes in automatic evaluations (e.g., Rydell & McConnell, 2008), and the non-conscious has been assumed to be unable to process conjunctions and negations (e.g., Dijksterhuis & Nordgren, 2006). Yet, people appear very quickly able to process relations beyond simple associations, even non-consciously (e.g., Sklar et al., 2012). Further, such quick propositional reasoning can impact and even “undo” automatic evaluations (Cone & Ferguson, 2015; see De Houwer, 2014). It may be that automatic vs. controlled processing might be entirely orthogonal to associative vs. propositional processing, but in any case, we see little explanatory progress in claiming the implicit misattribution process to be either associative or propositional in nature.

**Does stimulus co-occurrence have different effects on deliberate versus automatic evaluative response?**

As we discuss in greater detail above, the IMM proposes that misattribution creates gut-level intuitions represented as S-R associations, not deliberative, propositionally-based attitudes. In other words, IM that results from the co-occurrences of CS and US affects automatic more than deliberative evaluative responses. And relatedly, because implicit/indirect measures are typically better suited to assess evaluations over which one lacks control, they are the most straight-forward means of assessing such attitudes.

That is not to say that such attitudes should not influence responses on explicit/direct measures. Indeed, we have shown when they can. Affective associations are often the first thing

activated when perceiving an object toward which one has an attitude. Therefore, increasing trust in initial (gut) reactions should lead to higher correspondence between the implicitly formed attitude and an explicit measure. This is precisely what we have found (Kendrick & Olson, 2012). After experimentally creating attitudes toward novel objects (i.e., Pokémon characters), we manipulated participants into thinking of themselves as either more or less intuitive, or more or less expert, at judging Pokémon. Across both manipulations, we found that individuals told they were intuitive or expert exhibited stronger EC effects compared to those told they were more rational or had little expertise. The “expert” and “intuitive” participants, therefore, relied on their automatic evaluative response when subsequently evaluating the stimuli; alternatively, given reason not to trust their gut, low intuition/expert participants’ evaluations were driven more by deliberate evaluative responses. Since IM affects automatic evaluative responses, low intuition/expert participants likely did not rely on these attitudes, and since no EC effects were found in this group, deliberate evaluative responses remained unaffected by the conditioning procedure.

Thus, the IMM assumes that co-occurrence in the surveillance paradigm is more likely to affect automatic than deliberate evaluative responses. People will, however, allow such automatic evaluative responses to influence their deliberate evaluative responses if they are neither motivated nor have the opportunity to do otherwise. In fact, this is a core hypothesis of the MODE model, and much research supports it (see Fazio & Olson, 2014, for a review).

### **What is the role of awareness in EC?**

It should be clear that awareness generally inhibits misattribution, regardless of type of awareness (e.g., valence, identity, demand) or stage of processing (e.g., encoding, storage, expression). Of course, there are devils in the details we describe below.

First, one must distinguish between awareness, attention, and encoding. For successful EC to occur, the CS and US must be attended to for any affect from the US to be activated and any attribution of affect to be made to the CS. Given the spotty history of subliminal or unconscious EC (see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010), one might speculate that awareness of the pairings is therefore necessary for EC to occur. However, from our perspective, very brief (e.g., 15-20 ms or less) presentations of either CS, US, or both, simply reduce the likelihood of *any* amount of attention to and encoding of the stimuli. Weak or questionable evidence of subliminal EC is not evidence that awareness of the pairings is necessary for EC to occur. Regardless of what mechanism a given model considers, the less time, attention, and encoding respondents can possibly allocate to the CS-US pairings, the less likely it is that *any* EC will occur, regardless of process.

That said, awareness of the contingency between the CS and US is likely to lead to accurate attribution of the source of affect as the US, thus precluding misattribution. This is likely to particularly be the case for valence awareness, that is, knowledge of the systematic pairings of CS and USs of a given valence, regardless of whether one remembers the specific US (see Stahl, Unkelbach, & Corneille, 2009). Indeed, noting the source of one's (positive or negative) affect likely increases attention to that source (as the principles of hedonics would imply), thereby decreasing attention to other objects in the immediate environment. If one has grown aware of some pairing scheme, then the US and its valence become all the more salient. As a result, the source of the experienced affect should be appropriately identified as the US. Conceptually, this is equivalent to our size manipulation of salience (Jones et al., 2009). Thus, contingency awareness should reduce EC effects particularly when the paradigm leading to EC relies on a misattribution mechanism; anything that reduced the "mis" in misattribution should

interfere with the result of such a process. Among the demand aware, that is, those who have identified the goal of the research, misattribution seems particularly unlikely. Such individuals must have not only noticed the systematic pairings, but also derived the conclusion that the pairings were supposed to change their attitudes. They are arguably saying something to themselves like, “You want me to associate that puppy’s positivity with this Pokémon, but I know that whatever warm fuzzies I’m feeling come from the puppy.” Of course, EC may still occur under such conditions from other mechanisms, including both demand effects and propositional reasoning.

We readily admit that in other paradigms relying on other mechanisms, awareness is likely to lead to increased EC effects. Though we assume that IMM-based EC effects are likely to be smaller (at least on explicit measures) than EC effects produced through propositional reasoning, we also suspect that the awareness-effect size relation is particularly present for explicit but not implicit measures of EC. As we have shown (Kendrick & Olson, 2012), relying on one’s “gut” (i.e., by encouraging participants to attend to their affect) produces larger explicit EC. There are obvious reasons why one might not trust their gut on an explicit measure. From the perspective of the IMM, it would seem unlikely that respondents would be aware that their attitudes had formed or changed in any way as a function of the CS-US pairings. Indeed, such awareness would suggest intentionally thinking about them. Evidence for the unintentional expression of implicitly formed attitudes via US comes from evidence from our lab using subliminal priming measures (where the CS was never even consciously perceived, yet still influenced responses to target words; Olson & Fazio, 2002), and processing tree paradigms (e.g., Hütter, Sweldens, Stahl, Unkelbach, & Klauer, 2012). However, that is not to say that respondents are unaware of the attitudes that formed as a function of the pairings. Indeed, across numerous studies, they

have reported them on explicit measures.

### **How do relational qualifiers present at the time of acquisition influence EC?**

The IMM suggests that IM is most likely to occur when stimuli co-occur in space and time. The mere co-occurrence of stimuli implies, at least implicitly, a relationship between the CS and US. The linking that results from co-occurrence is not unavoidable, if certain conditions are present. Indeed, anything that draws attention to the systematic pairings has potential to prompt conscious attributional thinking and stifle misattribution. For example, by telling participants the sole source of their felt affect, Ruys et al. (2012) were able to inhibit IM. But, by letting the source remain ambiguous between two origins, IM occurred. Pointing out the source of affect specifies the appropriate attribution for the source of affect.

We also suspect that the introduction of explicit relational qualifiers in EC paradigms (e.g., Fiedler & Unkelbach, 2011; Förderer & Unkelbach, 2012) draws conscious attention to the CS-US pairs and prompts deliberate processing of the sources of any experienced affect. We expect that such focused attention to the CS-US relations reduces the likelihood of misattribution because it focuses attention to the US as the source of experienced affect and hence increases processing of it (thereby decreasing processing of the CS). Focusing attention on CS-US relations likely increases the likelihood of propositional reasoning. It is not surprising to us that these studies tend to show evidence of propositionally-based EC; propositional reasoning is what that research context compels participants to do and provides them ample opportunity to do it.

However, simply becoming aware of a relationship between CS and US is not necessarily sufficient to interfere with misattribution. IM is not inhibited by knowledge of pairings per se, but of the source of affect elicited by the US (as more accurately measured through valence awareness). It is safe to assume that, in most situations, the two types of knowledge are

correlated. Particularly when using lowly-evocative stimuli (as we generally do), we expect that awareness of the pairings can exist without knowledge of the source of implicitly experienced affect. Relational qualifiers that draw attention to the *source* of affect as being something other than the CS are likely to interfere with IM and diminish EC effects. Ultimately, from the perspective of the IMM, the impact of any relational qualifier depends upon the extent to which it affects source confusability.

**Is EC inevitable when stimuli co-occur? Does the model assume specific factors that would moderate EC?**

Whether EC is likely when stimuli co-occur likely depends on several factors, though the answer to the overall question is a resounding “no!”, simple co-occurrence is not by itself a sufficient predictor of misattribution. EC is certainly not inevitable when stimuli co-occur, but certain contexts and person factors are likely to interact with the effects of co-occurrence. Things like goal relevance (which we discuss in further detail in the next section), focus on valence, attentional resources, and source confusability all likely influence the inevitability of EC effects after stimuli co-occur in the context of the IMM. Attending to the CS-US pairings is obviously a prerequisite for IM to occur, and we have shown that increasing co-attendance to both CS and US increases EC effects (Jones et al., 2009).

Source confusability, which we talked about at length in the introduction, is the primary moderator of EC effects in the context of the IMM. If the source of the affect elicited by the US is clear, IM is unlikely. We have shown that making the CS more perceptually salient increases EC effects as drawing relatively more attention to the CS vs the US likely increases source confusability. Inhibiting resource availability at encoding (e.g., Mierop, Hutter, & Corneille, 2016) has also been shown to interfere with EC effects; any time resources are depleted means

less implicit as well as explicit processing.

Several recent studies suggest that EC effects are sometimes goal dependent and that goals active during conditioning affect the likelihood that misattribution will occur. That is, the stimulus dimension being attended can mediate the EC effect. For example, when valence processing is not central to the conditioning task or is undermined by a secondary task, EC effects are depressed (Field & Moore, 2005; Gast & Rothermund, 2011). When participants are tasked with judging the valence of CS-US pairs, EC effects are found. But, when participants are tasked with judging the pairings along a non-valenced dimension (e.g., size), EC effects diminish. This implies that an evaluative goal (which indeed is likely the default disposition) is a likely requisite for IM. Task demands encouraging processing along a non-valent dimension by directing evaluation away from a valence dimension will likely inhibit IM of valence. That said, there are conditions where misattribution of non-valence-related dimensions occur. For example, Olson, Kendrick, & Fazio (2009) increased the accessibility of non-valence-related dimensions (e.g., size, speed) via subliminal priming, which in turn increased misattributions along those dimensions. Our sense is that the dimension of valence is largely a “default” dimension of misattribution because valence is such a central and accessible dimension of meaning. When other dimensions are made salient, misattribution can occur with those dimensions as well.

### **How does verbal information about the CS-US co-occurrence (instruction) change CS evaluation?**

As stated in the previous section on relational qualifiers, any information that draws attention to the pairings has the possibility of reducing IM. Once the source of affect becomes obvious, misattribution becomes unlikely. The model does not make specific predictions about

the effect of verbal instructions, but it seems apparent that a model that specifies an implicit mechanism would have little to say about contexts that make processing more explicit.

Moreover, the absence of actual experienced affect (apart from a singular instruction) reduces even the possibility that any affect would be attributed one way or the other.

**Does the model predict sensitivity to statistical CS-US contingency or CS-US contiguity?**

Misattribution is a process that occurs as a function of CS-US contiguity. Thus, at least in principle, a single CS-US pairing may be sufficient for misattribution. However, multiple CS-US pairings are likely to increase the probability of misattribution, as well as provide opportunity for rehearsal of any newly-acquired association, which should increase its strength (Fazio, 1995). Contingency learning, on the other hand, requires some modicum of attention, across multiple trials, of the statistical relation between CS and US. Other implicit learning mechanisms, particularly those dealing with language acquisition, are uniquely sensitive to such statistical regularities (e.g., Saffran, Aslin, & Newport, 1996). We speculate that minor variations in contingency (e.g., 75% vs. 95%) would have little effect on implicit misattribution.

**Does the model predict that EC would be sensitive to later presentations of CS or US alone (i.e., show extinction)?**

Attitudes formed implicitly through CS-ER association inherently involve an affective response and therefore may be less sensitive to extinction than attitudes formed via CS-US linking would be. The IMM makes no specific predictions regarding sensitivity to later presentations, but current research implies that attitudes formed via IM may be particularly resilient, and perhaps have a slower decay rate than other types of attitudes. Furthermore, unreinforced CS presentations affect self-reported evaluations more than EC effects measured through priming measures (Gawronski, Gast, & De Houwer, 2015). Priming measures are likely

to capture the gut level attitude we assume IM generates. Therefore, it appears that implicitly formed attitudes, particularly those involving CS-ER associations, extinguish relatively slowly.

Of course, no stored representation is immune to decay, and there is no reason to expect that an attitude derived through implicit misattribution would not be subject to the same processes of decay beyond those described in extinction research. If, for example, the evaluative response evoked by the CS does not prove functional (e.g., the object is rarely encountered, or is encountered in contexts where there is little functional value in approaching or avoiding it), then it may lose its propensity toward activation when the object is encountered. For example, Sanbonmatsu, Posavac, Vanous, Ho, and Fazio (2007) found that when people encountered objects toward which they harbored attitudes capable of automatic activation in contexts where the attitudes did not serve any function, the automatic evaluative response weakened.

**Does the model predict any individual differences at any stage (acquisition, storage, expression) of evaluative learning?**

Though no specific predictions are made formally by the model, certain individual differences may be likely to moderate EC effects resulting from IM at acquisition. First and foremost, these would be factors that impact source confusability. For example, chronic ambivalence (Haddock, Foad, Windsor-Shellard, Dummel, & Adarves-Yorno, 2017) involves the regular experience of both positive and negative affect toward a variety of objects. It may be that chronically ambivalent individuals are more likely to misattribute any experienced affect. Other factors that may impact accurate attributional thinking may also lead to misattribution, such as being cognitively taxed either as a result of being anxious/stressed, or intoxicated. We would hasten to add, however, that being taxed, distracted, or intoxicated may decrease attention to the CS and US, thus possibly reducing the likelihood of both the experience of affect and

misattribution to the CS. Also with regards to acquisition, there is research suggesting that individuals high in need for closure (NFC) are particularly likely to misattribute internal experiences (Ecker & Bar-Anan, 2017). Specifically, ambiguity is experienced as an aversive state that people are driven to resolve (Fox & Tversky, 1995). Individuals high in NFC (Webster & Kruglanski, 1994) are driven to resolve such conflict, and in their effort to avoid ambiguity, such individuals are more likely to experience misattribution. Finally, it may also be the case that people who simply experience more affect generally (e.g., those who are high in neuroticism, emotional reactivity, or who experience stronger moods), in having more affect “ready” for attributional processing, increase their odds of misattribution.

The IMM makes no predictions about individual differences affecting storage of attitudes brought about through IM.

Because the attitudes resulting from IM are experienced as low-level, gut affect, people who are more likely to attend to such “gut feelings” will be more likely to attend to and express such attitudes when queried explicitly (i.e., on a direct measure). Kendrick and Olson (2012), for example, asked their respondents, “Do you find yourself to be more rational or more intuitive?” Those who identified themselves as more intuitive were more likely to express their implicitly-formed attitudes on a direct measure. There are also individual differences in people’s use of their internal experience as a good and proper source of information (Ecker & Bar-Anan, 2017; Schwarz & Clore, 1983; Slovic, Finucane, Peters, & MacGregor, 2007). For example, people high in body awareness (BA) are more likely to attribute their current affective state to whatever stimuli they happen to be evaluating (Mehling et al., 2009). Since people high in BA are no more accurate in determining the source of their affect than people low in BA, that sureness (i.e., high BA) will increase attention to gut feelings.

**Does the model specify differences between types of paired stimuli? Are there any specific categories of stimuli that are assumed to produce different EC outcomes?**

The IMM does not make specific predictions regarding differences between types of paired stimuli, per se, but it does speak to characteristics of CS-US content that encourage source confusability. Indeed, aspects of the CS-US content comprise one of the boundary conditions of IM. As we have discussed, the more evocative a US is, the less likely source confusion is to occur because such USs attract and focus attention and facilitate the “correct” attribution. Arousal evoked by a naked attractive individual is not likely to be confused with the (lack of) arousal evoked by a Pokémon character. The model assumes that as source clarity increases, IM decreases. As discussed above, we tested this assumption and found that mildly evocative USs were more effective at producing EC through IM than strongly evocative USs (Jones et al., 2009).

Another characteristic of CS-US content that may moderate EC through IM is belongingness (e.g., Hamm, Vaitl, & Lang, 1989). IM relies on the creation of a relationship between the response evoked by the US and the CS, which is more likely to occur if such a link is plausible. Consider a study that exposed thirsty participants to pairings of soda with either fearful or disgusted faces (Verwijmeren, Karremans, Stroebe, & Wigboldus, 2012). Thirsty participants showed EC effects to beverage brands paired with disgusted faced (informative, i.e., belong together) and no EC effects to beverage brands paired with fearful faces (uninformative, i.e., do not belong together). Both disgust and fearful faces are negative, and indeed fearful faces are perhaps more important from a survival perspective. Yet, to the thirsty person, the fearful face is not an appropriate source of information (i.e., not a good fit with the CS) about the potability of a soft-drink, but the disgusted face may imply useful information about the CS.

Related research suggest that the relatedness of CS-US pairings influences how they are processed (Blask, Walther, & Frings, 2017). Of course, within the IMM, we could consider notions of belongingness or relatedness factors that influence source confusability.

**Are there any important predictions, not mentioned previously, that would be central for testing the model?**

Given the responses, occasionally somewhat redundant, to the various questions that the editors posed, the model has been presented fully. In closing, we simply wish to emphasize a few points. First, the IMM is concerned specifically with acquisition, the transfer of valence from the US to the CS. It is silent with respect to many other issues, but as we have discussed, related theoretical frameworks and associated empirical evidence allow us to shed light on such matters as whether any implicit misattribution of valence to the CS becomes evident in an individual's verbal expressions and behavior. Second, the IMM highlights a specific aspect of the CS-US relation, namely the extent to which the potential exists for source confusability regarding the evaluation activated by the US. That is the key mechanistic variable through which contextual factors and individual differences might influence any EC that is observed. Finally, we wish express our firm conviction that EC theory and research, if it is to have any substantial impact on the broader scientific community, must move beyond debates regarding assertions that are presumed to apply universally to all the multiple mechanisms by which EC – the effect – can occur. Delineation of specific mechanisms and the paradigms to which they are applicable will allow for more appropriately-nuanced conclusions regarding the nature of the association acquired as a result of CS-US pairings, the role of awareness, the likelihood of extinction, etc. As such theoretical understanding develops, scientists beyond the limited community pursuing EC research will be in a better position to identify the potential relevance of

any given EC mechanism to the phenomena that interest them and also specific paradigms that might serve as useful social influence tools within that domain. We look forward to such progress.

## References

- Baeyens, F., Vanhouche, W., Crombez, G., & Eelen, P. (1998). Human evaluative flavor-flavor conditioning is not sensitive to post-acquisition US-inflation. *Psychologica Belgica*, *38*, 83-108.
- Blask, K., Walther, E., & Frings, C. (2017). Ignorance reflects preference: the influence of selective ignoring on evaluative conditioning. *Journal of Cognitive Psychology*, *8*, 939-948.
- Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in cognitive sciences*, *16*, 411-417.
- Cone, J., & Ferguson, M. J. (2015). He did what? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology*, *108*, 37-57.
- Cooper, J., Blackman, S. F., and Keller, K. T. (2016). *The Science of Attitudes*. New York: Routledge.
- Cotton, J. L. (1981). A review of research on Schachter's theory of emotion and the misattribution of arousal. *European Journal of Social Psychology*, *11*, 365-397.
- De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, *10*, 230-241.
- De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning and Behavior*, *37*, 1-20.
- Houwer, J. D. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, *8*, 342-353.
- De Houwer, J., Baeyens, F., Vansteenwegen, D., & Eelen, P. (2000). Evaluative conditioning in the picture-picture paradigm with random assignment of conditioned stimuli to unconditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*, 237-242.
- Dijksterhuis, A., & Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological science*, *1*, 95-109.
- Dutton, D. G., & Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of personality and social psychology*, *30*, 510-517.
- Ecker, Y., & Bar-Anan, Y. (2017, August 10). The Misattributing Personality: The Relationship between Affect Misattribution, Need for Closure and Body Awareness. *Unpublished manuscript*. Retrieved from [osf.io/my8f2](https://osf.io/my8f2).

- Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247-282). Hillsdale, NJ: Erlbaum.
- Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition, 25*, 603-637.
- Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition and Emotion, 25*, 639-656
- Field, A. P., & Moore, A. C. (2005). Dissociating the effects of attention and contingency awareness on evaluative conditioning effects in the visual paradigm. *Cognition & Emotion, 19*, 217-243.
- Förderer, S., & Unkelbach, C. (2012). Hating the cute kitten or loving the aggressive pit-bull: EC effects depend on CS-US relations. *Cognition & Emotion, 26*, 534-540.
- Gast, A., & Rothermund, K. (2011). What you see is what will change: Evaluative conditioning effects depend on a focus on valence. *Cognition and Emotion, 25*, 89-110.
- Gawronski, B. (2004). Theory-based bias correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias. *European Review of Social Psychology, 15*, 183-217.
- Gawronski, B., Gast, A., & De Houwer, J. (2015). Is evaluative conditioning really resistant to extinction? Evidence for changes in evaluative judgements without changes in evaluative representations. *Cognition and Emotion, 29*, 816-830.
- Gawronski, B., & Mitchell, D. G. (2014). Simultaneous conditioning of valence and arousal. *Cognition & Emotion, 28*, 577-595.
- Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189-211). New York: Guilford.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*, 21-38.
- Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of personality and social psychology, 54*(5), 733-740.
- Haddock, G., Foad, C., Windsor-Shellard, B., Dummel, S., & Adarves-Yorno, I. (2017). On the attitudinal consequences of being mindful: Links between mindfulness and attitudinal ambivalence. *Personality and Social Psychology Bulletin, 43*, 439-452.

- Hamm, A. O., Vaitl, D., & Lang, P. J. (1989). Fear conditioning, meaning, and belongingness: A selective association analysis. *Journal of abnormal psychology, 98*, 395-406.
- Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: a meta-analysis. *Psychological bulletin, 136*, 390-421.
- Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General, 141*, 539-557.
- Hütter, M., & Sweldens, S. (2013). Implicit misattribution of evaluative responses: contingency-unaware evaluative conditioning requires simultaneous stimulus presentations. *Journal of Experimental Psychology: General, 142*, 638-644.
- Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of personality and social psychology, 96*(5), 933-948.
- Jones, C. R., Olson, M. A., & Fazio, R. H. (2010). Chapter five-evaluative conditioning: The “how” question. *Advances in experimental social psychology, 43*, 205-255.
- Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3*, 1-24.
- Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual review of psychology, 31*, 457-501.
- Kendrick, R. V., & Olson, M. A. (2012). When feeling right leads to being right in the reporting of implicitly-formed attitudes, or how I learned to stop worrying and trust my gut. *Journal of Experimental Social Psychology, 48*, 1316-1321.
- Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 830-845.
- Maio, G., & Haddock, G. (2014). *The psychology of attitudes and attitude change*. Sage.
- Maio, G. R., & Olson, J. M. (Eds.). (1999). *Why we evaluate: Functions of attitudes*. Psychology Press.
- Mehling, W. E., Gopisetty, V., Daubenmier, J., Price, C. J., Hecht, F. M., & Stewart, A. (2009). Body awareness: construct and self-report measures. *PLoS one, 4* e5614.
- Mierop, A., Hütter, M., & Corneille, O. (2016). Resource availability and explicit memory largely determine evaluative conditioning effects in a paradigm claimed to be conducive to implicit attitude acquisition. *Social Psychological and Personality Science, 8*, 758-767.

- Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*, 183-198.
- Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, *64*, 723-739.
- Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*, *12*, 413-417.
- Olson, M. A., & Fazio, R. H. (2002). Implicit acquisition and manifestation of classically conditioned attitudes. *Social Cognition*, *20*, 89-103.
- Olson, M. A., & Fazio, R. H. (2006). Reducing automatically-activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32*, 421-433.
- Olson, M. A., Kendrick, R. V., & Fazio, R. H. (2009). Implicit learning of evaluative vs. non-evaluative covariations: The role of dimension accessibility. *Journal of Experimental Social Psychology*, *45*, 398-403.
- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: affect misattribution as implicit measurement. *Journal of personality and social psychology*, *89*, 277-293.
- Pryor, J. B., & Kriss, M. (1977). The cognitive dynamics of salience in the attribution process. *Journal of Personality and Social Psychology*, *35*, 49-55.
- Pryor, J. B., Reeder, G. D., & Monroe, A. E. (2012). The infection of bad company: stigma by association. *Journal of Personality and Social Psychology*, *102*, 224-241.
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of experimental psychology: General*, *118*, 219-235.
- Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, *9*, 45-48.
- Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*, 364-382.
- Roskos-Ewoldsen, D. R., & Fazio, R. H. (1992). On the orienting value of attitudes: Attitude accessibility as a determinant of an object's attraction of visual attention. *Journal of Personality and Social Psychology*, *63*, 198-211.

- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in experimental social psychology*, *10*, 173-220.
- Ross, M., & Fletcher, G. J. (1985). Attribution and social perception. *Handbook of social psychology*, *2*, 73-122.
- Ruys, K. I., Aarts, H., Papiés, E. K., Oikawa, M., & Oikawa, H. (2012). Perceiving an exclusive cause of affect prevents misattribution. *Consciousness and Cognition*, *21*, 1009-1015.
- Rydell, R. J., & Jones, C. R. (2009). Competition between unconditioned stimuli in attitude formation: Valence asymmetry versus spatiotemporal contiguity. *Social Cognition*, *27*, 905-916.
- Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of personality and social psychology*, *91*, 995-1008.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 1926-1928.
- Sanbonmatsu, D. M., Posavac, S. S., Vanous, S., Ho, E. A., & Fazio, R. H. (2007). The deautomatization of accessible attitudes. *Journal of Experimental Social Psychology*, *43*, 365-378.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of personality and social psychology*, *45*, 513-523.
- Sinclair, R. C., & Mark, M. M. (1992). The influence of mood state on judgment and action: Effects on persuasion, categorization, social justice, person perception, and judgmental accuracy. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 165-193). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sklar, A. Y., Levy, N., Goldstein, A., Mandel, R., Maril, A., & Hassin, R. R. (2012). Reading and doing arithmetic nonconsciously. *Proceedings of the National Academy of Sciences*, *109*, 19614-19619.
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European journal of operational research*, *177*, 1333-1352.
- Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology*, *97*, 404-420.

- Thomas, W. I., & Znaniecki, F. (1918). *The Polish peasant in Europe and America: Monograph of an immigrant group* (Vol. 2). Chicago: University of Chicago Press.
- Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Treisman, A. M., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*, 107–141.
- Verwijmeren, T., Karremans, J. C., Stroebe, W., & Wigboldus, D. H. (2012). Goal relevance moderates evaluative conditioning effects. *Learning and Motivation*, *43*, 107-115.
- Walther, E. (2002). Guilty by mere association: evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, *82*, 919-934.
- Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the back door: The US-revaluation effect. *Cognition and Emotion*, *23*, 889-917.
- Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*, 1049-1062.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, *92*, 548-573.
- Weinreich, A., & Gollwitzer, A. (2016). Automaticity and affective responses in valence transfer: Insights from the crossmodal auditory-visual paradigm. *Psychology of Music*, *44*, 1304-1317.
- Winkielman, P., & Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology*, *81*, 989–1000.