On the Prioritized Processing of Threat in a Dual Implicit Process Model of Evaluation

David S. March, Lowell Gaertner & Michael A. Olson


To link to this article: https://doi.org/10.1080/1047840X.2018.1435680

Published online: 21 Mar 2018.
ABSTRACT
Dual-process models of attitudes distinguish between implicit and explicit processes in which the valence (i.e., positivity or negativity) of a stimulus influences judgments and behavior toward the stimulus. Developing parallel to the dual-process literature has been a threat detection literature suggesting that the mind is preferentially attuned to threats to immediate bodily harm. That literature reveals early privileged responses (e.g., shorter latency of detection, stronger reflexive reactions, and faster and stronger physiological responses) to threatening stimuli relative to negative, neutral, and positive stimuli. By integrating those literatures, we develop the dual implicit process model that postulates two functionally distinct and serially linked automatic processes in which an implicit threat process precedes (and potentially influences) an implicit valence process (positive vs. negative), which precedes (and potentially influences) explicit processes. In addition to explicating the nature of the model, we examine insights it offers various research areas, and conclude by identifying open questions regarding the model.

Imagine two experiences: (a) You were once terrorized by a knife-wielding funhouse clown and now automatically recoil whenever you see one; (b) you have a strong automatic dislike of mimes. Existing dual-process models of evaluation (e.g., Fazio & Olson, 2014; Gawronski & Bodenhausen, 2006) would conceptualize both as implicit negative evaluations. We argue that the two reactions are the product of distinct evaluative processes, both of which can be activated quickly and without intention yet have unique causes and consequences, and likely different analogues in the brain. We propose, in particular, that your reaction to the clown is the result of an implicit threat evaluation and a threat detection literature that developed parallel to (but in isolation of) the former. We then consider limitations of dual-process approaches given the threat literature, and introduce the dual implicit process model (DIPM), which integrates key aspects of both literatures. Finally, we examine insights that the DIPM offers various research areas, and we conclude with open questions about the model.

Dual-Process Models of Evaluation
Dual-process models propose that evaluation occurs through both automatic (i.e., implicit, spontaneous) and controlled (i.e., explicit, deliberate) processes. Automaticity implies that a process functions with at least one of the four characteristics of being unintentional, effortless, uncontrollable, or beyond awareness ( Bargh, 1994; Greenwald & Banaji, 1995). In terms of latency, automatic processes are relatively fast and controlled processes are relatively slow. Such a dual-process framework has been applied quite broadly to account for a variety of psychological phenomena including attitude–behavior correspondence ( Fazio, 1990), attribution ( Trope, 1986), decision making ( Sloman, 1996), impression formation ( Brewer, 1988; Fiske & Neuberg, 1990), persuasion ( Chaiken, 1980; Petty & Cacioppo, 1986), and prejudice ( Devine, 1989; Gaertner & Dovidio, 1986; Pettigrew & Meertens, 1995). A number of dual-process models of evaluation have been developed that differ in regard to operating principles (i.e., mechanisms by which stimuli are processed) and operating conditions (i.e., circumstances in which the mechanisms function). We describe the current state of several dual-process models with a focus on the aspects of each that are relevant to our argument that threat processing is distinct from other forms of automatic evaluative processing.

The Motivation and Opportunity as Determinants Model
The Motivation and Opportunity as DETERminants (MODE) model ( Fazio, 1990; Fazio & Olson, 2014) proposes that attitudes can be expressed in judgments and behavior through spontaneous and/or deliberative processes. From the perspective of the MODE model, the starting point for evaluative responding is whether an attitude of sufficient strength is activated automatically upon perception of an attitude object. Spontaneous processing occurs automatically, without intention, and can influence downstream judgments and behavior. Deliberative processing occurs downstream, is effortful, and may attempt to steer judgments and behaviors away from the implications of spontaneously activated evaluations given...
ample motivation and opportunity to do so. But, if the sponta-
neously activated evaluation is sufficiently strong, it can bias
initial stages of processing that have downstream consequences
on deliberative perceptions, judgments, behaviors—in terms of
guiding attention toward the object (Roskos-Ewoldsen & Fazio,
1992) and construal of it (Smith, Fazio, & Ciejka, 1996). Motiva-
tion and opportunity together determine whether the expressed
judgment or behavior mainly reflects spontaneous and/or
deliberate processing. Motivation refers to the desire to engage
in deliberative processing, whether in a general pursuit of accu-
ricacy, or in pursuit of a specific conclusion (e., to avoid preju-
dice). Opportunity refers to the capacity in a given situation to
engage in deliberative processing; some situations, like those
requiring fast reactions, limit the capacity to deliberatively con-
roll responses. Similarly, fatigue, intoxication, and other factors
that limit cognitive capacity also limit opportunity. In the
absence of motivation or opportunity, the spontaneous evalu-
ation of the object in question is the main determinant of judg-
ments and behavior. With both motivation and opportunity
present, evaluative judgments and behaviors are more likely to
reflect deliberately held motives rather than automatically acti-
vated attitudes.

For instance, imagine a chocolate lover encountering a
chocolate bar. A strong positive preexisting attitude toward
chocolate would lead one to automatically categorize a
chocolate bar as something delicious to consume. If the
chocolate lover, however, were on a diet (motivated to
avoid empty calories) and fully rested and sober (able to
template calories), the chocolate lover would likely abstain. On the other hand, if the dieting chocolate lover
were hungry, intoxicated, or stressed, (i.e., a compromised
capacity to think), the spontaneously positive evaluation
would likely lead to consumption of the chocolate.

For the present purposes, it is important to note that the
MODE model considers automaticity of the attitude along a
continuum defined only by an object-evaluation association,
that is, the association in memory between an object and its
summary evaluation. It does not consider other factors that
might determine spontaneity of activation. Regardless of the
nature of the object, attitudes characterized by stronger object-
evaluation associations are more likely to have such properties
of automaticity as spontaneous activation of the evaluation in
response to perceiving the object, attitude-consistent construal
of the object, and spontaneous approach/avoidance behavior
toward the object.

The Associative-Propositional Evaluation Model

The associative-propositional evaluation (APE) model
(Gawronski & Bodenhausen, 2006; 2007) is focused on operat-
ing principles (i.e., characteristics of the mechanism operating)
in contrast to the MODE model’s focus on operating conditions
(i.e., when a mechanism is operating). The APE model argues
that evaluation manifests through associative and propositional
processes, with the former generally being automatic and the
latter being controlled (Gawronski & Bodenhausen, 2009,
2014). The associative process links objects and events through
contiguity (i.e., environmental or spatiotemporal co-occur-
rence; e.g., evaluative conditioning; Jones, Olson, & Fazio,
2010). The activation of associative processes occurs via feature
similarity (e.g., “that looks like chocolate”) and is experienced
as automatic affect. Propositional reasoning is more effortful
and includes truth/falsity tags by weighing the validity of the
activated associations using both preexisting information and
contextual cues.

Because associative and propositional processes glean inform-
ation from different sources, they may not always agree.
Whatever “momentarily considered” propositions are activated
in response to the associative input determines whether those
associations are accepted or rejected. Such acceptance or
rejection depends on principles of cognitive consistency. For
example, if the proposition is congruent with the activated
association, the associative evaluation will likely influence
evaluative judgments. But, if propositional reasoning diverges
from the associative evaluation, the association is rejected as a
basis for evaluative judgments (and possibly behavior). Because
propositional (i.e., correction) processing occurs last, it is the
operation most vulnerable to constraints on attention, time, or
cognitive resources (analogous to the “opportunity” factor of
the MODE model).

Continuing the chocolate example, chocolate is often por-
trayed as delicious, and repeated co-occurrence of chocolate
and delicious would yield an associatively formed positive eval-
uation of chocolate. Yet a chocolate-loving dieter’s positive
automatic associative response to chocolate is inconsistent with
the momentarily considered propositional belief that chocolate
is high in calories. Similar to MODE model logic, whether
propositional reasoning overrules the associative link between
chocolate and delicious is influenced by operating conditions.
If one has both motivation (e.g., to lose weight) and opportu-
nity, the propositional concern with weight gain (e.g., “I believe
that eating that delicious chocolate would be inconsistent with
my weight loss goals”) would override the positive associations
activated in response to the chocolate. Lacking either (e.g., if
starving or intoxicated) would likely lead to judgments reflect-
ning the initial association.

Like the MODE model, the APE model does not make dis-
tinctions about the nature of the evaluative associations that
produce automatic evaluative responses. All that matters are
their associative strength.

The Quad Model

The quad model distinguishes between four processes of
evaluation that involve the automatic versus controlled dis-
tinction: activation, guessing, discrimination, and overcom-
ing bias (Conrey, Sherman, Gawronski, Hugenberg, &
Groom, 2005). The activation (AC) process activates exist-
ing evaluative associations in response to objects encoun-
tered in the environment. The guessing (G) process drives
responses in the absence of other information, often result-
ing in response biases (i.e., a default response tendency in a
given context, such as to respond with approach rather than
avoidance). The discrimination (D) process uses explicit information from previous experience or the envi-
nment to determine possible outcomes, and the overcom-
ing bias (OB) process monitors for the appropriate response
and attempts to suppress automatically activated evaluative
associations or response tendencies. These processes are not purely automatic or controlled. Although AC and G usually function automatically, G can function more deliberatively. Conversely, although D and OB usually function deliberatively, they can acquire automatic features.

For instance, imagine that a chocolate lover receives what appears to be chocolate but it is actually the terrible-tasting chocolate substitute carob. A preexisting positive automatic evaluation of chocolate is activated (AC) because the chocolate lover’s default response toward anything resembling chocolate is to assume (G) that it is indeed chocolate. But, having been tricked into eating carob in the recent past, the perceiver is motivated to overcome the initial bias (OB) to ensure that they discriminate (D) what the candy bar actually is, delicious chocolate or terrible carob, before eating it.

Like the preceding models, the QUAD model is agnostic regarding what sorts of objects might become associated with automatic evaluation. It does not make distinctions about the nature of the associations within the AC and G processes.

**Other Relevant Models**

Dual-attitudes models (e.g., Wilson, Lindsey, & Schooler, 2000) propose that attitudes can form both implicitly and explicitly toward the same object and coexist as distinct mental entities. This implies the possibility of evaluative dissociation, with the implicit attitude and explicit attitudes having opposite valence.

Dual-system models locate automatic and controlled processes, or what these models refer to as reflexive versus reflective (Lieberman, Gaunt, Gilbert, & Trope, 2002) or intuitive versus rule based (Sloman, 1996, 2014), in distinct mental systems with unique neural substrates. As with dual-attitudes models, these dual-systems approaches suggest that the automatic and controlled processes can yield distinct attitudes of opposite valence (McConnell & Rydell, 2014). Downstream reflective (rule-based) processing may inhibit the reflexive (intuitive) system similarly to the MODE model.

Resuming the chocolate example, upon an initial enjoyable exposure to chocolate the automatic/intuitive/reflexive system may form a positive attitude toward chocolate. After learning about the high caloric content of chocolate, the controlled/ rule-based/reflexive system forms a second negative evaluation. Dual-attitudes and dual-systems models agree that the positive implicit attitude and negative explicit attitude now coexist, but dual-attitudes models assume one system is responsible for both evaluations, whereas dual-systems models place each evaluation within the confines of a unique system. Whether the automatic positive evaluation of chocolate is constrained by the controlled system is again dependent on the MODE model’s operating conditions (i.e., motivation and opportunity).

In summary, a number of dual-process models of evaluation have been proposed. They differ in detail but share the common distinction between automatic and controlled evaluative processing. And they do not make functional distinctions between different kinds of automatic or implicit evaluative processing. As we discuss in subsequent sections, this is a limitation that the literature on threat processing brings to light.

**Threat Processing**

Threat is a broadly employed concept in psychological research, such as threat to identity (Steele, 1997), ingroup status and resources (Stephan & Stephan, 2000), and self-esteem (Bushman, Smart, & Boden, 1996). Anchored in an evolutionary framework, the threat- (i.e., fear-) processing literature has focused exclusively on threats to immediate bodily harm, which is our focus here. Ancestors who reacted more quickly when confronted with such threats were more likely to survive than were their slower conspecifics (Blanchette, 2006), which is likely why humans overestimate the threat relevance of stimuli in ambiguous situations (Haselton & Buss, 2000; Nesse, 2005). A psychology of threat perception ostensibly evolved as an adaptive mechanism for the rapid avoidance of physical danger (e.g., Schaller & Neuberg, 2012).

Ohman has made the case for such an adaption in the form of a hypothetical “fear module” responsible for processing and initiating reaction to threat (Ohman & Mineka, 2001, 2003). Such processing ostensibly utilizes neural circuitry that evolved prior to the cortices, enabling it to function semi-independent of (i.e., in parallel to) non-threat-relevant mental processes (Calder, Lawrence, & Young, 2001). Consistent with such a possibility is what LeDoux (1996, 2012) referred to as the “low road,” a subcortical pathway to the amygdala capable of detecting threat and activating associated responses without explicit processing (in contrast to the “high road,” which is slower but provides more processed, i.e., cortical, information). Such a low road allows for rapid responses after only superficial processing of the stimuli, whereas the high road undertakes more comprehensive processing before instigating a response (Morris, Öhman, & Dolan, 1999).

Although the amygdala is involved in the general processing of affective and motivationally relevant information (including novel and extremely positive stimuli; Cunningham & Brosch, 2012), it is particularly attuned to the initial processing of threatening information (Campeau et al., 1991; Cunningham, Packer, Kesek, & Van Bavel, 2009; Lewis, Critchley, Rotshtein, & Dolan, 2007; Pascoe & Kapp, 1985). The amygdala may thus initiate responses to stimuli and activate associated processes before neocortical structures have received, interpreted, and responded to the same information. Neuroanatomical research of the human brain supports such a “low road” capable of detecting and evaluating emotionally salient information and initiating responses without explicit awareness (Garrido, Barnes, Sahani, & Dolan, 2012; Garvert, Friston, Dolan, & Garrido, 2014; Whalen et al., 2004; cf. Pessoa & Adolphs, 2010).

As we review, research is consistent with the possibility that humans inherited an ability to preferentially process threats to immediate survival—an effect that has been dubbed the threat-superiority effect (Blanchette, 2006; E. Fox & Damjanovic, 2006). Stimuli such as lions, snakes, sharks, and weapons pose an imminent threat to survival and necessitate quick detection and avoidance. Clearly, such threatening stimuli are negative. But not all negative stimuli are threatening. Important to note, the existing research indicates that the mind preferentially
processes threatening stimuli, not simply negative stimuli—this is a point to which we return. Furthermore, the preferential response to threatening stimuli relative to negative, neutral, and positive stimuli occurs in terms of very early responses, such as shorter latency of detection, stronger reflexive reactions, and faster and stronger physiological responses, all of which underlie the presumed adaptive value of quickly detecting and reacting to survival threats.

Consistent with the possibility that threatening stimuli are preferentially processed is research indicating that people more quickly detect threatening stimuli (e.g., snake, spider, gun, knife) than innocuous stimuli (e.g., flowers, mushrooms; Blanchette, 2006; Brosch & Sharma, 2005; E. Fox, Griggs, & Mouchlianitis, 2007; Öhman, Flykt, & Esteves, 2001) and are quicker to detect angry faces than happy or sad faces (Fox et al., 2000; Öhman et al., 2001; Öhman, Lundqvist, & Esteves, 2001). Eye-tracking research has shown that initial attention is more frequently drawn to threatening images than to positive or neutral images (Hermans, Vansteenwegen, & Eelen, 1999; Rinck & Becker, 2006). Research using continuous flash suppression (which prevents awareness of stimuli for multiple seconds) suggests that misattribution of affects occurs for angry but not happy faces (Almeida, Pajas, Mahon, Nakayama, & Carazza, 2013) and threat-related faces emerge from suppression and into consciousness more quickly than do neutral or happy faces (Yang, Zald, & Blake, 2007).

Similarly, research suggests that both supraliminally and subliminally presented threatening stimuli elicit unique physiological responses (Knight, Waters, & Bandettini, 2009; Morris, Öhman, & Dolan, 1998; Whalen et al., 1998). For example, subliminally presented angry faces increase skin conductance responses while happy faces do not (Esteves, Dimberg, & Öhman, 1994), subliminally presented fearful faces increase amygdala activity relative to happy faces (Whalen et al., 1998), and classical conditioning occurs to subliminal angry but not happy faces (Esteves, Parra, Dimberg, & Öhman, 1994). Also, participants conditioned to supraliminal angry faces continue to show evidence of conditioning (i.e., increased skin conductance) after the aversive shock is removed (i.e., during extinction), but show no lingering conditioning to happy or neutral faces (Dimberg & Öhman, 1996). Supraliminally presented directly threatening stimuli elicit stronger and faster amygdala responses than do other types of negative stimuli (Kveraga et al., 2014) and supraliminally presented threatening stimuli elicit earlier cortical activity than do other types of negative and positive stimuli (Costa et al., 2014). The amygdala also shows more activity to subliminally presented fearful than neutral faces independent of fusiform facial area activity, suggesting that such activation is independent of attention (Vuilleumier, Armony, Driver, & Dolan, 2001)

Although nondiagnostic of a threat versus negative difference, similar patterns are reported in studies that collapsed across threatening stimuli (e.g., guns) and nonthreatening-negative stimuli (e.g., attributes such as rude). People, for example, are better able to identify the presence and categorize the valence of subliminally presented threatening/negative words than neutral or positive words (Dijkstra Reagh & Knight, 2013) and event-related brain potentials are larger during evaluative categorization of supraliminally presented threatening/negative stimuli than positive or neutral stimuli (Ito, Larsen, Smith, & Cacioppo, 1998). When attention is distracted from the emotional content of stimuli, amygdala activity is heightened in response to threatening/negative stimuli relative to positive and neutral stimuli (Straube, Pohlack, Mentzel, & Milner, 2008).

Our own research provides direct evidence that threatening stimuli, not simply negative stimuli, elicit preferential responses (March, Gaertner, & Olson, 2017). To distinguish threatening stimuli and nonthreatening-negative stimuli, we pilot tested images to obtain four stimulus categories: threatening, nonthreatening-negative, positive, and neutral. The nonthreatening-negative category consisted of stimuli that are unpleasant but do not pose an immediate survival threat (e.g., rotten teeth, excrement, sick or dead animals, maggots). Important to note, both the threatening and nonthreatening-negative stimuli were evaluated as negative but only the former were evaluated as highly threatening. All four categories were equivalent in red value and luminance. We subsequently employed those stimuli in three paradigms that assess very early responses: visual search, eye-tracking, and startle-eyeblink. Consistent with the argument that threat is preferentially processed, participants (a) were faster to detect a threatening than nonthreatening-negative image when each was embedded among positive or neutral images; (b) oriented their initial gaze more frequently toward threatening than nonthreatening-negative, positive, or neutral images; and (c) evidenced larger startle-eyeblinks to threatening than to nonthreatening-negative, positive, or neutral images.

Hence the threat-processing literature yields robust evidence of a threat-superiority effect that manifests as faster and stronger perceptual, physiological, and behavioral reactions to stimuli that pose an immediate survival threat. Important to note, this preferential processing occurs in regard to early reactions and does not imply that it occurs at a longer or more protracted time course. Indeed, the functional utility of the effect is ostensibly rooted in the survival value of avoiding imminent danger. Notably, this early preferential processing of threat carries an interesting and important implication for dual-process models of evaluation.

Limitation of Dual Process

Dual-process models are valuable for understanding the differences between and interplay of implicit versus explicit evaluation and have established those processes as being serial in nature. Implicit processing is associative, intuitive, and reflexive, and as such it is fast and efficient. Explicit processing is controlled, propositional, deliberative, and reflective, and as such it is slow, effortful, and downstream of implicit processing. In terms of dual-process models, the fast and early responses of the threat-superiority effect constitute an implicit response. Existing dual-process models, however, cannot explain or account for threat superiority. From the perspective of these models, the automaticity of an implicit response is thought to occur equally for positively and negatively evaluated stimuli. The MODE model, for example, conceptualizes automaticity of
evaluative responses solely in terms of the strength of the object-evaluation association and does not qualify that tenant in terms of valence or threat potential. Similarly, the APE model discusses activated associations between objects and affect without specifying further the nature of those associations. In short, these models do not articulate a speed or strength difference between evaluations of different valences. The threat literature, on the other hand, indicates a qualitative distinction in implicit processing that existing dual-process models do not capture. Threatening stimuli are evaluated negatively, as are nonthreatening-negative stimuli, but threatening stimuli systematically elicit faster and stronger responses than do other negative or positive stimuli. Here the mind is queued not just to valence but particularly to threat.

When juxtaposed with dual-process models, the threat literature implies the possibility of two serially linked implicit processes. One implicit process, evaluative threat processing, is attuned to survival threats such that objects are evaluated implicitly as to whether they pose an immediate survival threat. The other implicit processes, evaluative valence processing, is attuned to valence such that objects are evaluated implicitly as to whether they are good or bad. In what follows, we integrate those processes into a DIPM and demonstrate the conceptual value of distinguishing implicit threat processing and implicit valence processing.

The Dual Implicit Process Model

The DIPM postulates two functionally distinct and serially linked automatic evaluative processes: the first implicit process (i1) is solely oriented toward threats to bodily harm. This process precedes and potentially influences the subsequent implicit process (i2) that encompasses the full evaluative continuum (negative to positive). As we define it, i2 aligns well with notions of automatic and implicit evaluative and explicit processes in the dual-process models we reviewed earlier. Both i1 and i2 precede and potentially influence subsequent explicit (“e”) processing (see Figure 1). Based on findings that threatening stimuli are preferentially processed via the low-road (e.g., LeDoux, 1996), the DIPM purports that i1 processing occurs first to activate threat responses, given proper input. The slower operating high-road processes nonthreatening stimuli via i2 to provide more detailed evaluative responses. Given that i2 and e functioning are consistent with extant dual-process models, we focus primarily on explicating i1. The following sections detail the inputs to i1, outputs of i1, and influences on the magnitudes of the outputs, and then puts it all together in the full temporal sequence of evaluative processing.

Inputs to i1

Öhman (Öhman et al., 2001; Öhman & Mineka, 2001, 2003) proposed the idea of threat sensitivity in terms of a neural “fear module,” which implies the possibility that threat evaluation is queued uniquely to the phylogenetic stimuli (e.g., snakes, spiders) on which it evolved. To be functional, however, a threat evaluation process should learn and incorporate new threats (Gould & Vrba, 1982). Indeed, fear in humans can be learned both directly through firsthand experience and indirectly through social learning, with the amygdala playing a central role in both instances (Ohlsson & Phelps, 2007). Consistent with the possibility of a flexible system, threat-superiority research indicates that ontogenetic threats (e.g., modern weapons) and phylogenetic threats (e.g., animals, fire) have the same functional effect of being preferentially processed (Blanchette, 2006; Brosch & Sharma, 2005; E. Fox et al., 2007). In our own research (March et al., 2017), for example, participants (a) were faster to detect both ontogenetic and phylogenetic threats than negative stimuli; (b) oriented their initial gaze more frequently to both ontogenetic and phylogenetic threats than to negative, positive, or neutral stimuli; and (c) evidenced larger startle-eyeblinks to both ontogenetic and phylogenetic threats than to negative, positive, and neutral stimuli. Hence, i1 is supported by a flexible system that learns and incorporates new objects as potential threats.

Outputs of i1

Implicit process i1 yields two interrelated outcomes: (a) immediate physiological reactions to protect/prepare the body from threat and (b) organized downstream information processing queued to the threat. Both outcomes function to protect against harm and ensure safety (e.g., via vigilance, active avoidance mechanisms; Bolles & Fanselow, 1980; Pratto & John, 1991; Wentura, Rothermund, & Bak, 2000).

Physiological outputs. The physiological outputs of i1 are those typically associated with responding to threat (LeDoux, 2014), including protective reflexes (e.g., freezing, flailing, defensive fighting), autonomic arousal (e.g., pupil dilation, heart rate increase, sweating), and neurological activity (e.g., amygdala activation, secretion of epinephrine by the adrenal glands). Neurological actions prepare the body to respond to threat by instigating overt behavioral movements and autonomic arousal meant to aid the body’s response to threat (Löw, Weymar, & Hamm, 2015). Once initiated, such responses inevitably run to completion with little opportunity for interruption (i.e., a fixed action pattern; Lorenz, 1965).
Organized downstream processing. Activated i1 organizes downstream information processing for a vigilant focus on the threatening stimulus (e.g., Pratto & John, 1991). Such organization allocates attentional resources toward gathering further information about the threat and is enabled via neuronal pathways between the amygdala, supplementary structures, and cortices as explicated by LeDoux (1996), Ohman (2005) and others (e.g., A. S. Fox, Oler, Tromp, Fudge, & Kalin, 2015). This allows for the possibility that i1 can negatively bias the evaluative process of i2 and e in a manner consistent with the activated threat. Imagine, for example, three people: Donna dislikes cats, Lisa likes cats, and Nancy is neutral toward cats. If each were to encounter a cat in the absence of i1 activation, Donna’s automatic i2 evaluation would be negative, Lisa’s would be positive, and Nancy’s would be neutral. In contrast, if each were to encounter a hissing cat with exposed fangs and arched back, the now-activated i1 would negatively bias i2 for all three persons (relative to their own prior reaction): Donna would have a faster and more negative evaluation, Lisa would have a slower and less positive evaluation, and Nancy would have a relatively fast and negative evaluation. Hence, i1 activation influences the evaluative process of i2 and alters it from what it would have been absent i1 activation.

There is also the possibility that i1 directly affects e (i.e., independent of i2), whereby the activated threat is made apparent to e which then processes it directly. For example, (neutral) Nancy’s i1 response to the hissing cat promotes active attention to the cat with deliberate avoidance (e.g., Nancy might say to herself, “I should avoid that nasty cat”). Alternatively, i1 may indirectly affect e via i2. For example, Nancy’s i1 response to the hissing cat promotes a negative implicit evaluation of the cat (i.e., i2), which in turn could promote a negative explicit evaluation of the cat and deliberate avoidance.

Magnitude of Outcomes Produced by i1

The magnitude of i1’s outcomes is determined by the threat’s perceived imminence and potency. Threat imminence refers to the proximal distance of the threat in time and space, and, all else being equal, the magnitude of the i1 response increases with imminence (Fanselow, 1994; Kveraga et al., 2014; Löw, Lang, Smith, & Bradley, 2008; Löw et al., 2015; Mobbs et al., 2007). For example, a threat at a greater distance would likely elicit less sympathetic arousal and attentional resources than would a threat at a closer distance (Löw et al., 2008). Threat potency refers to the presumed capacity of the threat to inflict bodily harm (e.g., Lundqvist, Esteves, & Öhman, 1999). That is, some threats (e.g., a roaring lion) are perceived to have greater potency to harm than other threats (e.g., hissing cat) and, all else being equal, the magnitude of the i1 response increases with potency (Blanchard, Hynd, Minke, Tiffanie, & Blanchard, 2001).

As depicted in Figure 2, the resulting magnitude of the i1 response is conditioned simultaneously on imminence and potency. When the threat is of high imminence and potency, the magnitude of the i1 response is maximized. For example, coming face-to-face with a bear in the woods would activate i1 and yield strong physiological responses (e.g., release of epinephrine, elevated heartrate) and highly organized downstream cognitive processing of the bear to maximize a protective behavioral response (fight or flight). In contrast, viewing a photograph of a bear (or encountering the bear from a distance), for which perceived potency remains high but perceived imminence is reduced, would activate i1 with a weaker response (physiology, downstream cognition, behavior) than in the face-to-face encounter.

Finally, an interesting possibility to consider is the case of an ambiguous threat, such as a creakingloorboard at midnight or a creepy-crawling movement across one’s back. Properties of error management (Haselton & Buss, 2000) would lead to the overperception of threat imminence and potency, which would yield a stronger i1 response than would have occurred with accurate perception of the creaking as due to the wind rather than an intruder or the creepy crawling as due to a tree branch rather than a spider.

Putting It All Together

Evaluation absent i1 threat activation follows the evaluative process explicated by current dual process models. For instance, imagine standing in a park when in the distance, you notice your friend approaching. At this point, an evaluation of your friend is activated. i2 activates a (presumably) summary positive association, and likewise explicitly you are happy to see your friend. This chain of evaluative events does not include an i1 activation, and hence no i1 influence. The conditions laid out by current dual process models (e.g., MODE, APE) stipulate that given the motivation and opportunity to control one’s responses, the explicit positive attitude toward your friend is that likely to be expressed. Alternatively, when a threat is present, the DIPM proposes an alternative process from that proposed by current dual process models. For example, imagine waiting in that same park, but now instead of approaching from the front, your friend sneaks up behind you and scares you. Before consciously (e.g., e) identifying the threat, i1 is activated, involving immediate behavioral and physiological responses preparing you to respond to the threat but also a negative biasing effect downstream. The i2 evaluation is now less positive than it would have been absent an i1 event, as is your explicit evaluation.
It may appear to the reader that we are merely tacking on threat processing to dual-process models, and in a sense we are. However, as we argue next, it is likely to be active in a number of domains, including those often considered within the context of attitudes research (e.g., prejudice), as well as in domains into which attitudes researchers are typically more reticent to venture (e.g., phobia, intimate partner violence), with implications for evaluative responding across them all. Indeed, we believe an important strength of the DIPM is its ability to integrate literatures with little cross-talk, highlighting similarities in evaluative processing across them, with implications for behavioral responses as well as the potential for change.

**Implications of the DIPM**

By distinguishing implicit threat processing and implicit valence processing, the DIPM provides a fuller understanding of evaluative processing and offers unique insights to various fields of study. The following sections examine some of those insights to prejudice, phobia, and intimate partner violence.

**Implications for Prejudice**

By distinguishing an implicit threat process from an implicit valence process, the DIPM suggests that implicit biases toward social groups can be functionally distinguished in regard to threat versus automatic positive/negative evaluation. Black-danger stereotypes, for example, uniquely draw attention to Black faces whereas non-threat-related but negative Black stereotypes do not (Donders, Correll, & Wittenbrink, 2008). Similarly, White participants show stronger amygdala activation to subliminally presented Black than White faces (Cunningham et al., 2004). In these examples, threat cues ostensibly drive processing in a manner different than do merely negative (i.e., nonthreatening stereotypes) or positive (own-race face for White participants).

Perhaps even more interesting is the possibility that ingroup perception can yield both ingroup favoritism and outgroup favoritism among groups associated with threat or aggression. Black participants, for example, show a shooter bias against Blacks (Kahn & Davies, 2011) but also evidence ingroup favoritism on other implicit measures (e.g., evaluative priming: Fazio, Jackson, Dunton, & Williams, 1995; Implicit Association Test: Olson, Crawford, & Devlin, 2009). Similarly, Middle Eastern participants are more likely to shoot Middle Eastern targets wearing traditional Saudi headgear than less traditional baseball hats (Schofield, Deckman, Garris, DeWall, & Denson, 2015). There has been much debate as to whether members of particular marginalized groups, despite their low status, evince ingroup favoritism or reflect society’s negative views of their group in the form of outgroup favoritism (e.g., Jost, Gaucher, & Stern, 2015; Olson et al., 2009). The DIPM offers a possible reconciliation of those views by suggesting that members of marginalized groups—particularly those associated with aggression—might simultaneously show both automatic outgroup favoritism (via i1 threat responses to own-group members) and automatic ingroup favoritism (via i2 positive responses to own-group members).

We also see implications for the DIPM in terms of prejudice measurement. Specifically, the DIPM’s implications of separate i1 and i2 responses in the domain of prejudice provide a framework for future research to uncover the extent to which a given measure of prejudice assesses threat-related responses (as we earlier implied the shooter-bias might; Correll, Park, Judd, & Wittenbrink, 2002) versus i2-related evaluative responses (as the Implicit Association Test and priming measures that employ only valence-relevant attribute items might; Fazio & Olson, 2003). This leads to the possibility of better prediction of discriminatory behavior (Guglielmi, 1999). As threat responses and “mere” approach/avoidance responses likely look different behaviorally, the i1/i2 distinction provides a finer-grained analysis of the sorts of behaviors that different measures of prejudice are likely to predict. For example, a measure more attuned to threat responses might better predict fearful emotions and avoidance of the target, whereas a measure more attuned to valence might better predict anger or disgust and either approach or other information-seeking behaviors.

Finally, in line with contemporary theories of the functional bases of prejudice (e.g., Schaller & Neuberg, 2012; Smith, 1993), the DIPM suggests a functional distinction between more physical-threat-based and more evaluative-based prejudices (perhaps, e.g., in terms of violations of traditional values). Different interventions might be developed to address the specific underlying bases of the prejudice based on this distinction. However, as discussed next, threat responses may be more difficult to unlearn, suggesting that threat-based prejudices may be more difficult to reduce.

**Implications for Phobia**

Phobias are the result of a dysfunctional predictive relationship that occurs when the level of threat associated with a stimulus becomes catastrophic (Grillon, Pellowski, Merikangas, & Davis, 1997). Some treatments appear able to reduce the magnitude of fear responses, often extinguishing them entirely (Hermans, Craske, Mineka, & Lovibond, 2006). Yet, a return of fear (e.g., renewal, reinstatement, spontaneous recovery) often occurs when the ostensibly extinguished fear response suddenly returns. The DIPM suggests that phobia treatments appear to target i2 and e, with less influence on i1. In particular, the spontaneous return of fear suggests that the phobic target remains associated with threat via i1 despite becoming more favorably evaluated via i2 and e (Hermans et al., 2006). The temporary reduction of fear responses is likely a highly controlled process through explicit behavior modification supported by more favorable i2 and e evaluations that attempt to suppress i1’s outputs.

The regulating ability of i2 and e is suggested by research showing that the more favorable peoples’ automatic and explicit evaluations became of the fear inducing stimulus (i.e., the effectiveness of treatment), the better was their ability to control behavioral manifestations of fear (Huijding & de Jong, 2007; Vasey, Harbaugh, Buffington, Jones, & Fazio, 2012). Further emphasizing that what is being effected is the regulation of fear, and not the unlearning of the association between the target and threat, is research showing that spider phobia treatment had a positive impact on evaluations of valence (good/bad) but had no impact on threat (danger/safety) evaluation (Teachman & Woody, 2003). Similarly, after undergoing positive training that biases responses to emotional
situations, participants evaluations of emotional situations became more positive, but such training had no impact on their fearful responses to threatening stimuli (Teachman & Addison, 2008). In lieu of therapy-induced change to i1’s evaluation, reduction of fear response may reflect an increased ability to (at least temporarily) regulate fear responses. Return of fear implies that i1 maintains at least some threat association to the phobic object even when phobic responses can be successfully suppressed.

This research serves to raise an important issue about whether it is possible to unlearn threat associations over extended time. Imagine, for example, a novice snake-handler with a deadly cobra. On the first few encounters, the novice would likely experience a racing heart, sweaty palms, and palpable fear. Over time, however, absent fear-reinforcing events, the practiced snake-handler would likely experience less activation of i1. Of interest, snakes and spiders elicit negative implicit evaluations from snake neophytes and spider neophytes and positive implicit evaluations from snake experts and spider experts (Ehgartner, Rinck, & Becker, 2006; Purkis & Lipp, 2007). Such a positive (i2) evaluation (and presumably positive explicit evaluation) might contribute to the unlearning of an i1 input over time. Yet there remains an open question about whether any evaluative association can be unlearned (Ebbinghaus, 1885/1964; Petty, Tormala, Brinol, & Jarvis, 2006), and it would certainly be a survival benefit for threat associations to be particularily inlexible to change. Indeed, recall research showing that conditioned angry faces evoke fear responses long after the paired shock is removed but that happy and neutral faces ceased evoking fear responses almost immediately (Dimberg & Öhman, 1996). This implies that even if it is possible to unlearn an association, those tied to survival value may linger much longer.

The amygdala is vital to conditioned fear (i.e., phobias; LaBar, Gatenby, Gore, LeDoux, & Phelps, 1998), meaning that successfully unlearning phobias may involve altering the amygdala’s ingrained threat associations. Therefore, phobia treatments aimed at making attitudes toward the object more positive (via i2 and e) may affect i1 only indirectly as a result of changes in i2 and e evaluations. But important to note, changes in i2 and e are not diagnostic of a reduction in the i1 phobic evaluation. Phobias may be less susceptible to certain types of cognitive therapy. Phobia treatments that alter the i1 evaluation may therefore be more effective changing the long-term outlook.

Implications for Intimate Partner Violence

Victims of intimate partner violence often stay in abusive relationships for extended periods, and even after leaving often return to the abusive partner (Anderson et al., 2003; Schutte, Malouff, & Doyle, 1988). Victims often convey positive feelings and attitudes toward their abuser, frequently even denying the occurrence of abuse, and going so far as to defend their abuser (Wallace, 2007). Several theories have been put forth that focus on victim characteristics and logistical factors behind stay/leave decisions (e.g., Rusbult & Martz, 1995; Strube, 1988). The incongruence of victims’ positive self-reported attitudes and the objective negativity of abuse is often explained though traumatic bonding theories (i.e., Stockholm Syndrome) as a sort of cognitive dissonance. Yet much is still unknown about the processes that underlie victims’ stay/leave decisions (Bell & Naugle, 2005; Strube, 1988). In addition to the many factors already identified as predictors of stay/leave behavior, the DIPM would suggest that victims can hold positive i2 evaluations and concurrently associate their abuser with i1 threat. Acts of warmth or kindness intermittent with acts of abuse essentially create opposing i1 and i2/e evaluations. Decisions to stay may be partially explained by the abuse victim’s strong positive i2 and e evaluation, effectively controlling/overcoming their i1 fear of their partner.

Indirect evidence comes from research showing that the amount of self-reported emotional attachment, feelings of love, loyalty, and the level of commitment toward the abusive partner all predict likelihood of remaining in an abusive relationship (Hayes & Jeffries, 2013). That is, the level of positive i2 and explicit evaluation toward the abuser predicted victim stay/leave behaviors. Of interest, research has shown that increased emotional abuse (ostensibly affecting i2) is more predictive of a victim’s decision to leave than is increased physical abuse (ostensibly affecting i1; Gortner, Berns, Jacobson, & Gottman, 1997). This indicates that perhaps a negative change in i2 or e is the determining factor leading to a change in behavior. That is, positive i2 and e evaluations allowed the victim to control his or her i1 fear, but a weakened i2 or e may lessen that control and allow an increase in protective leave behaviors. More direct evidence for i1/i2/e roles in stay/leave decisions awaits the more precise measurement of implicit constructs called for by the DIPM.

Open Questions

There remain a number of open conceptual questions regarding the DIPM that necessitate further discussion. We examine several such issues in the current section.

Summation of Successive i1 Events

Existing research suggests the possibility that the downstream output of i1 is stronger when there are successive i1 events occurring in short duration than when there is only a single i1 event. That is, perhaps successive i1 events have a summative

---

**Figure 3.** Magnitude of the first implicit process (i1) response across time.
effect such that preceding i1 activation strengthens subsequent i1 activation (see Figure 3). Indirect evidence of this summative effect comes from the potentiating effect of ambient darkness (vs. ambient light) on both the startle-reflex (Grillon et al., 1997) and White participants’ expression of threat-relevant prejudice against Black targets (Schaller, Park, & Mueller, 2003). In each instance, ambient darkness can be considered the initial i1 trigger, which strengthened (relevant to ambient light) the subsequent startle-response to an unexpected noise blast (second i1) and activation of threat-relevant (e.g., hostile) but not threat-irrelevant (e.g., lazy) stereotype ratings of Black targets (second i1; assuming Blacks are perceived as threatening; see the Implication for Prejudice section). Stated otherwise, participants experiencing two threats (darkness and noise-blast or images of Black persons) evidenced larger threat responses than did participants experiencing a single threat response (light and noise-blast or images of Black persons). This suggests the summative nature of successive i1 activation.

### Dual Implicit Processing in a Simple versus Complex World

In their now classic publication, Murphy and Zajonc (1993) demonstrated that both angry and smiling faces presented at 4 ms yielded valance-consistent misattribution to a subsequent supraliminally presented neutral stimulus. Some readers might construe this finding as being inconsistent with the DIPM: If smiling faces trigger positive misattribution at only a 4-ms exposure, how could there be room for earlier processing of a threatening stimulus? In our view, this poses no problem for the DIPM. Murphy and Zajonc created a simple world in which participants were presented on any given trial with a single stimulus (angry or smiling face) without competition from other stimuli. In a complex world involving multiple co-occurring stimuli, the DIPM suggests that threatening stimuli are processed preferentially, and, in the absence of threatening stimuli, nonthreatening stimuli are processed directly via i2. So, Murphy and Zajonc is quite consistent with the DIPM.

### Is i1 Less Susceptible Than i2 to Downregulation?

There is evidence that explicit goals (e) can affect the activation of implicit valence (i2). For example, Stewart and Payne (2008) demonstrated that an implementation intention to activate counterstereotypic thoughts in response to a Black face reduced the automatic activation of prejudice. Thus, e appears capable of downregulating i2. However, consistent with our preceding discussion of the phobia research, i1 may be less susceptible to downregulation. The imbalanced connections between the amygdala and the cortices (i.e., more afferent than efferent) implies that during the initial processing of a threatening stimulus, i1 informs i2 and e more than the reverse. This suggests that it might be difficult for any higher order processes (i.e., e) to impact i1. For example, explicit knowledge that you are watching a horror movie does not stifle the threat response when the axe murderer jumps out of the closet.

On the other hand, perhaps i2 and e have some potential to downregulate i1. The model posits that when encountering a threat, i1 activates and evokes the cascade of outcomes described previously. But i2 and e process the same stimulus, and their evaluations might “turn down” the i1 downstream biasing effect on information processing despite the continued physiological arousal. Imagine walking through the woods and hearing rustling sounds behind you. Initially, i1 evokes behavioral and downstream outcomes that direct automatic attention to the rustling. Upon further deliberate inspection, the rustling is realized to be the pattering of a baby rabbit. The identification of the rabbit triggers an automatic positive (i2) evaluation. Although your heart is still pounding and your palms are sweating, perhaps i1 no longer negatively biases information processing.

Indirect support of such downregulation comes from fMRI evidence of differential amygdala activity to Black versus White faces (with the assumption that Black faces constitute a threat to many White participants). Although supraliminal presentation yielded equivalent fMRI amygdala activation to Black and White faces (Cunningham et al., 2004; Hart et al., 2000), (a) amygdala habituation was slower (i.e., activity remained heightened longer) to Black than White faces (Hart et al., 2000) and (b) subliminal presentation yielded stronger amygdala activation to Black than White faces (Cunningham et al., 2004). These results imply that the i1 response to Black faces dominated at very short (i.e., subliminal) exposures that granted no opportunity for downregulation by i2 or e. But, when images were presented supraliminally, though i1 evinced a downstream influence (i.e., slower habituation), i2 and e downregulation of i1 was possible.

Cunningham et al. (2004) suggested that increased activity in the dorsolateral prefrontal cortex and anterior cingulate cortex (areas implicated in the control of regulatory responses) congruent with decreased amygdala activity reflect the suppression of automatic amygdala activation when threats are consciously perceived. We supplement this interpretation by suggesting that these results reflect the modulation of i1 over the time course of processing. Interpreting these results is limited by the poor temporal resolution of fMRI, which builds a summary image over multiple seconds. Were this same experiment able to take millisecond-precise measurements of amygdala activity, it may have revealed that initial activity to Black faces was equal across sub- and supraliminal conditions and that only over time (i.e., after i2 and explicit processing) was amygdala activity dampened.

### Individual Differences in i1 Processing

LeDoux (2015) argued that fear and anxiety are two manifestations of the same phenomena—threat. The only difference is that fear is a reaction to a specific stimulus, whereas anxiety occurs toward something more diffuse. Research is starting to illuminate how individual differences may predispose certain people toward both fear and anxiety. If so, this may imply that some people have heightened i1 activity. A specific serotonin transporter gene polymorphism (5-HTTLPR) has been associated with heightened sensitivity and reactivity to threat (Cheon, Livingston, Hong, & Chiao, 2014). People with one copy of the short allele display greater amygdala activity toward threatening stimuli (Munafo, Brown, & Hariri, 2008), which manifests as heightened anxiety, vigilance, and fear conditioning (Canli & Lesch, 2007). In addition, individuals with posttraumatic stress disorder show biased attention and quicker reaction to threatening stimuli (Block & Liberson, 2016).
Indirect evidence that individual differences in i1 responses lead to heightened fear-related outcomes comes from research showing that chronic anxiety may facilitate threat associations, potentiate experiences of threat, and predispose people to overperceive threat. For example, White, Ratcliff, Vasey, and McKoon (2010) showed that high- versus low-anxious individuals have a processing advantage for threatening words on a lexical decision task. Pitman and Orr (1986) adversely conditioned high- and low-anxious individuals to angry and neutral faces and found that angry faces resisted extinction only among high-anxious participants, indicating that anxiety played a role in buffering extinction. Bishop, Duncan, and Lawrence (2004) found that among low-anxious participants, amygdala activity was reduced when instructed to attend to a house than a simultaneously appearing threatening face but that high-anxious participants showed greater overall amygdala activity and no differences when attending to the house versus the threatening face.

Cheon et al. (2014) proposed that those with an S-allele “may appraise or react to cues of outgroup threat differently” such that the 5-HTTLPR genotype may predispose individuals to experience more negative intergroup contact or perceptions of the environment, which may ultimately shape intergroup bias” (p. 1269). Perhaps individual differences in (amygdala-mediated) i1 activation are likely to predispose certain persons to develop specific phobias and experience more chronic anxiety but may also make them more likely to endorse certain group stereotype, particularly those associated with threat.

**Conclusion**

We propose that automatic evaluation is made up of (at least) two distinct processes. First is an initial implicit processing of whether a stimulus poses a survival threat (i1). This process is sensitive to threats engrained from the evolutionary past, such as snakes and heights, as well as idiosyncratically learned threats unique to individual experience and social history, such as guns and social groups. Second is a subsequent implicit processing (i2) of the full evaluative spectrum (i.e., positive and negative), which reflects extant dual-process models of evaluation. The DIPM’s integration of early threat-processing into implicit evaluation more fully captures the dynamics that that underlie social cognition.

**References**


