

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

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

David Scott March
May 2019

Copyright © 2019 by David S. March
All rights reserved.

ACKNOWLEDGEMENTS

Well, Lowell and Michael, we fucking did it. I still have no reason why you accepted me in to this program, but I will forever be grateful that you looked past my past and gave me an opportunity to shape my future. Its here that I've truly found my calling, and that's in no small part due to your mentorship, guidance, friendship, and selflessness. It's not a stretch to say that you've given me more than most grad students dream of getting from their mentors. Without your commitment to this endeavor and your unflapping belief in its worth, I wouldn't have achieved this dream I thought impossible. You've been the best mentors a fella could ask for, so for lack of better words, thank you.

Mom, you've done so much for me without ever asking anything in return. Your sacrifice and determination have directly led me to where I am today. You've set an example of perseverance and generosity that I can only hope to emulate. I'm so proud to call you my mom and am grateful for everything you've done and continue to do for me. I love you so much!

And to my friends, family, A*Rae, Tom the cat, office mates, plants, and my aeropress, I love you all and appreciate you so!

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, I 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 what follows, I review existing dual-process models of evaluation and a threat detection literature that developed parallel to (but in isolation of) the former. I then consider weaknesses of dual-process approach given the threat literature, but also limitations in the threat literature that preclude making strong claims about unique threat processing. I then describe three studies that overcome those limitations and introduce the Dual Implicit Process Model (DIPM), which integrates key aspects of both literatures. I subsequently offer two studies that apply the DIPM's underlying dynamics to explore the basic mechanisms of implicit social cognition that often lead to prejudice. Finally, I close by discussing how the DIP may apply to certain other psychological phenomenon and discuss some open conceptual questions about the model.

TABLE OF CONTENTS

Chapter 1 Introduction.....	1
Dual-process Models of Evaluation.....	2
The Motivation and Opportunity as Determinants (MODE) Model	3
The Associative-Propositional Evaluation (APE) Model	5
The Quad Model	6
Other Relevant Models	7
Threat Processing.....	9
Limitation of the Dual Process Approach	13
Limitations in Threat Research.....	14
Chapter 2 Evidence for Threat Processing as a Distinct Evaluative Process	16
Pilot Study	16
Study 1	18
Method	19
Results.....	20
Discussion	21
Study 2	23
Method	23
Results.....	24
Discussion	26
Study 3	27
Method	29
Results.....	30
Discussion	32
Chapter 2 General Discussion	33
Is the Threat Effect Driven by a Facet of Nonthreatening-Negativity?	33
Ontogeny versus Phylogeny	35
Morbid Fascination	36
Implications of Threat Sensitivity for Evaluative Processing.....	38
Chapter 3 The Dual Implicit Process Model	40
Inputs to i1	42
Outputs of i1	42
Magnitude of Outcomes Produced by i1	44
Putting it all Together	46
Chapter 4 Prejudice from the Perspective of the DUal Implicit Process Model.....	48
Study 4	51
Methods.....	53
Results.....	54
Discussion	57
Study 5	59
Pilot Study.....	63
Method	64
Data Preparation.....	66

Results.....	73
Discussion.....	82
Implications of the DIPM to Prejudice.....	86
Chapter 5 General Discussion, Open Conceptual Questions, and Implications for	
Other Lines of Research.....	90
Implications of the DIPM.....	90
Implications for Phobia.....	90
Implications for Intimate Partner Violence.....	93
Implications for Suicide.....	94
Open Questions.....	96
Summation of Successive i1 Events.....	96
Dual Implicit Processing in a Simple vs. Complex World.....	97
Is i1 less Susceptible than i2 to Downregulation?.....	98
Individual Differences in i1 Processing.....	100
Conclusion.....	101
References.....	103
Appendix.....	126
Stimuli from Studies 1-4.....	127
Mouse-Tracking Metrics.....	136
Facial Stimuli from Study 5.....	150
Vita.....	158

LIST OF TABLES

Table 2.1. Mean ratings of good, bad, and threatening in final stimuli sets.	18
Table A.1. Mouse-tracking metrics of each Target x Distractor pairing by race of the face	149

LIST OF FIGURES

Figure 2.1. Mean visual search detection time.	22
Figure 2.2. Percent of trials a stimulus in each pairing was gazed at first.....	26
Figure 2.3. Mean standardized eye-blink amplitude as a function of image type	32
Figure 3.1. Diagram representation of the DIPM	41
Figure 3.2. Threat potency and imminence determine magnitude of i1 response	45
Figure 4.1 Mean reaction times to respond to each Prime x Target pairing	56
Figure 4.2. Sample trial from Study 5.....	66
Figure 4.3. Visual representation of a trial response.	68
Figure 4.4. (a) Actual estimated difference of Euclidean Distances. (b) Typical bacterial growth curve.	70
Figure 4.5. Maximum deviation location.....	72
Figure 4.6. Euclidean distance difference scores plotted across time for angry faces (and area-of-focus on lag time) as a function of Race and Distractor.	75
Figure 4.7. Euclidean distance difference scores plotted across time for sad faces (and area-of-focus on lag time) as a function of Race and Distractor.	77
Figure 4.8. Euclidean distance difference scores plotted across time for happy faces (and area-of-focus on lag time) as a function of Race and Distractor.	79
Figure 4.9. Euclidean distance difference scores plotted across time for neutral faces (and area-of-focus on lag time) as a function of Race and Distractor.	81
Figure 4.10. Area-of-focus on lag times of each race in every level of Target by Distractor. Note: White and Asian lines often overlap.	84
Figure 5.1. Magnitude of i1 response across time.	97
Figure A.1 .Raw mouse-tracking paths. Note: All paths have been horizontally re-mapped to the left. X-axis: 0 = center of screen, 1 = left edge of screen. Y-axis: 0 = bottom of screen, 1.5 = top edge of screen. Asian and White lines often overlap.....	148

CHAPTER 1

INTRODUCTION¹

Imagine two experiences: (1) you were once terrorized by a knife-wielding funhouse clown and now automatically recoil whenever you see one; (2) you have a strong automatic dislike of mimes. Existing dual-process models of evaluation (e.g., Fazio & Olson, 2014; Gawronski & Bodenhausen, 2006, 2011) conceptualize both as implicit negative evaluations and highlight the automatic properties of strong evaluation. I 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. I propose, in particular, that your reaction to the clown is the result of an implicit threat evaluation that is functionally distinguishable from an implicit negative evaluation responsible for your reaction to the mime. In what follows, I review existing dual-process models of evaluation and a threat detection literature that developed parallel to (but in isolation of) the former. I then consider weaknesses of dual-process approach given the threat literature, but also limitations in the threat literature that preclude making strong claims about unique threat processing. I then describe three studies that overcome those limitations and introduce the Dual Implicit Process Model (DIPM), which integrates key aspects of both literatures. I subsequently offer two studies that apply the DIPM's underlying dynamics to explore the basic mechanisms of implicit social cognition that often lead to prejudice. Finally, I close

¹ Much of Chapter 1 was previously published in March, D., Gaertner, L., & Olson M.A. (2018). On the prioritized processing of threat in a dual implicit process model of evaluation. *Psychological Inquiry*, 29, 1-13.

by discussing how the DIP may apply to certain other psychological phenomenon and discuss some open conceptual questions about the model.

Dual-process Models of Evaluation

Dual-process models propose that evaluation occurs thorough 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 (Gilbert, Pelham, & Krull, 1988; 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). I describe the current state of several dual-process models with a focus on the aspects of each that are relevant to my argument that threat processing is distinct from other forms of automatic evaluative processing.

The Motivation and Opportunity as Determinants (MODE) 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 spontaneously 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, & Cejka, 1996). Motivation 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 accuracy, or in pursuit of a specific conclusion (e.g., to avoid prejudice). 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 deliberately control responses. Similarly, fatigue, intoxication, and other factors that limit cognitive capacity also limit opportunity. In the absence of motivation or opportunity, the spontaneous evaluation of the object in

question is the main determinant of judgments and behavior. With both motivation and opportunity present, evaluative judgments and behaviors are more likely to reflect deliberately held motives rather than automatically-activated 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 contemplate 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 (APE) Model

The Associative-Propositional Evaluation (APE) model (Gawronski & Bodenhausen, 2006; 2007) is focused on operating *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-occurrence; 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/falsehood tags by weighing the validity of the activated associations using both preexisting information and contextual cues.

Because associative and propositional processes glean information 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 portrayed as delicious, and repeated co-occurrence of chocolate and delicious would yield an associatively formed positive evaluation 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 opportunity, 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 reflecting the initial association.

Like the MODE model, the APE model does not make distinctions 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 vs. controlled distinction: Activation, Guessing, Discrimination, and Overcoming Bias (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005). The Activation (AC) process activates existing evaluative associations in response to objects

encountered in the environment. The Guessing (G) process drives responses in the absence of other information, often resulting 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 environment to determine possible outcomes, and the Overcoming 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. While AC and G usually function automatically, G can function more deliberately. Conversely, while D and OB usually function deliberately, they can acquire automatic features.

For instance, imagine 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) in order to ensure that they discriminate (D) what the candy bar actually is, delicious chocolate or terrible carob, before eating it.

Like the above 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 co-exist 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 vs. reflective (Lieberman, Gaunt, Gilbert, & Trope, 2002) or intuitive vs. 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/reflective 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, while 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 I 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 (Baumeister, 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 my 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).

Öhman makes the case for such an adaption in the form of a hypothetical “fear module” responsible for processing and initiating reaction to threat (Öhman & Mineka, 2001, 2003). Such processing ostensibly utilizes neural circuitry that evolved prior to the cortices, enabling it to function semi-independently 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) refers 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, while 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 (Campeu 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 (Garvert, Friston, Dolan, & Garrido, 2014; Garrido, Barnes, Sahani, & Dolan, 2012; Whalen et al., 2004; cf. Pessoa & Adolphs, 2010).

As I will 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; 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. Importantly, the existing research indicates that the mind preferentially processes threatening stimuli, not simply negative

stimuli – this is a point to which I will 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; Fox, Griggs, & Mouchlianitis, 2007; Öhman, Flykt, & Esteves, 2001) and are quicker to detect angry faces than happy or sad faces (Eastwood, Smilek, & Merikle, 2001; Fox et al., 2007; 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 affect occurs for angry but not happy faces (Almeida, Pajtas, Mahon, Nakayama, & Caramazza, 2013) and threat-relevant 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 et al., 1999; 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., 2015) 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 such activation is independent of attention (Vuilleumier, Armony, Driver, & Dolan, 2001)

Although non-diagnostic 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 (Dijksterhuis & Aarts, 2003). Subliminally presented threatening/negative but not positive stimuli potentiate startle-eyeblinks relative to neutral stimuli (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, & Miltner, 2008).

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. Importantly, 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 the Dual Process Approach

Recall how earlier I discussed several dual-process models that each describe separate implicit and explicit processes. These perspectives 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, just 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 not queued just to valence, it is particularly queued 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 the model I develop below, I do articulate such a distinction in an integrated model.

Limitations in Threat Research

Although extant research implies that threat receives preferential responding, methodological limitations prevent a firm conclusion and allow the possibility that negativity, not threat per se, is the trigger. Some studies precluded a direct test by

assessing reactions to threatening stimuli without additionally assessing reactions to nonthreatening-negative stimuli (e.g., Esteves et al., 1994; Öhman, Flykt, et al., 2001). Other studies, as noted above, confounded threatening and nonthreatening-negative stimuli by combining them in a single stimulus set (e.g., Dijksterhuis & Aarts, 2003; Ito, Cacioppo, & Lang, 1998; Reagh & Knight, 2013). One study directly compared threatening stimuli (i.e., angry face) versus nonthreatening-negative stimuli (i.e., sad face; Öhman, et al., 2001), but as the authors acknowledged, the threatening stimuli were rated as more negative than the nonthreatening-negative stimuli, leaving the possibility that responses were driven by negativity, not threat. Only Kveraga et al. (2015) and Costa et al. (2014) operationally distinguished threatening from nonthreatening-negative stimuli. Though they found differential brain activity, they did not assess different attentional and behavioral responses to threatening versus nonthreatening-negative stimuli, which should occur if sensitivity to threat functions for survival. In the next section, I provide such a test whereby I examine differential behavioral and physiological responses to threat and nonthreatening negative stimuli.

CHAPTER 2²

EVIDENCE FOR THREAT PROCESSING AS A DISTINCT EVALUATIVE PROCESS

In the current chapter, I provide a missing critical test as to whether threatening versus merely all negative stimuli elicit preferential responses. I established with pilot testing distinct sets of threatening, nonthreatening-negative, positive, and neutral stimuli and conducted three studies. Each study used a different paradigm to test whether the mind responds preferentially to threatening stimuli. Study 1 used visual search to test whether threatening stimuli are detected more quickly than nonthreatening-negative stimuli. Study 2 used eye-tracking to test whether initial attention is biased to threatening stimuli than other stimuli. Study 3 used startle-eyeblick to test whether threatening-stimuli elicit a stronger reflexive response than do other stimuli.

Pilot Study

I conducted a pilot study to obtain stimuli that are experienced as threatening, nonthreatening-negative, positive, or neutral, respectively. I collected 400 images from public sources on the Internet, the International Affective Picture System (Lang, Bradley, & Cuthbert, 2008), the Bank of Standardized Stimuli (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010), and images provided to me from Kveraga et al. (2015). I scaled all images to 500 x 500-pixels.

² Much of Chapter 2 was previously published in March, D., Gaertner, L., & Olson M.A. (2017). In harm's way: On preferential response to threatening stimuli. *Personality and Social Psychology Bulletin*, 11, 1519-1529.

One hundred and forty-nine undergraduates participated for partial credit in introductory psychology. Seated in computer cubicles, they rated 400 images (presented in a random order) on one of three randomly assigned dimensions of how good ($n = 50$), bad ($n = 51$), or threatening ($n = 48$) they deemed each image (1 = “Not at All” to 7 = “Extremely”). I computed each image’s mean rating of good, bad, and threatening, and, based on those ratings, assigned each image to one of four categories: positive, neutral, nonthreatening-negative, or threatening. *Positive category* images ($n = 94$) had bad and threat ratings less than 2 and good ratings greater than 5. *Neutral category* images ($n = 92$) had bad and threat ratings less than 2 and good ratings less than 5. *Nonthreatening-negative category* images ($n = 77$) had good ratings less than 3, bad ratings greater than 3, and threat ratings less than 4. *Threat category* images ($n = 92$) had good ratings less than 3, bad ratings greater than 3, and threat ratings greater than 4 (see Table 2.1).

I eliminated 45 images that could not be categorized and eliminated categorized images that were (a) rendered ambiguous when scaled to 300 x 300-pixels (which was necessary for Study 1), (b) natively too bright or dark to equate luminance across sets, or (c), could shift categories based on context (e.g., a plant could shift from neutral to positive if co-occurring with other positive stimuli). This yielded a final set in which the four categories were equated on luminance and red value and contained 40 images each.

Table 2.1. Mean ratings of good, bad, and threatening in final stimuli sets.

Image Set ^a	Rating Type: Mean (SD) ^b		
	Good	Bad	Threatening
Positive	6.14 (0.76)	1.07 (0.18)	1.09 (0.11)
Negative	1.53 (0.47)	4.78 (1.14)	3.18 (1.31)
Threatening	2.02 (0.86)	4.40 (1.54)	5.73 (0.99)

^aEach set contains 40 images.

^bRating type varied between-subjects (scale: “1 = *Not at All*” to “7 = *Extremely*”).

The non-threatening-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). Importantly, both the threatening and non-threatening-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. I subsequently employed these stimuli in the next three studies (see Appendix for all images).

Study 1

I employed a visual search paradigm (e.g., Öhman Flykt, et al., 2001; Öhman, Lundqvist, et al., 2001) to test whether threatening stimuli are detected faster than nonthreatening-negative stimuli. I compared the speed with which participants could detect a threatening stimulus embedded among all positive or neutral stimuli versus a nonthreatening-negative stimulus embedded among the same array of positive or neutral stimuli.

Method

One hundred and seven undergraduates (69 female, 38 male) participated for partial credit in introductory psychology. Participants sat in individual computer cubicles containing a 48cm high speed, high-resolution monitor and computer. Instructions explained that the study examined attention and rapid responding and that the participant would be presented with several trials consisting of a 3 x 3 grid of eight 300 x 300-pixel images surrounding a fixation point “X” in the center cell. Participants were informed that the grid would consist of pictures that were good, bad, or neutral, and that on some trials all eight images would be of the same type (i.e., all good, all bad, or all neutral) and on other trials seven of the images would be of one type and one image would be of a different type (see Supplemental Figure 1 for an example). Their task was to indicate whether the images were all of the “same” type by pressing the Z-key or whether one of the images was of a “different” type by pressing the /-key. Each trial was heralded by a 1000ms center-screen fixation point (“X”), after which the grid appeared and remained until participants responded. A variable 2000-6000ms blank screen separated each trial.

Participants completed 384 trials divided into four blocks of 96 trials, with a one-minute rest between blocks. Each block consisted of 48 congruent trials and 48 incongruent trials. Congruent trials were balanced across the four category sets. *Critical* incongruent trials within a block consisted of six trials each of one threatening image embedded among seven positive, one threatening image embedded among seven neutral, one nonthreatening-negative image embedded among seven positive, and one nonthreatening-negative image embedded among seven neutral. To ensure that the

presence of threatening or nonthreatening-negative images did not serve as a cue that the stimuli were “different,” each block also included *non-critical* incongruent trials that consisted of six trials each of one positive image embedded among seven threatening, one neutral image embedded among seven threatening, one positive image embedded among seven nonthreatening-negative, and one neutral image embedded among seven nonthreatening-negative. Hence, positive, neutral, nonthreatening-negative, and threatening images were equally likely to occur on congruent and incongruent trials and the order of congruent and incongruent trials was randomized. The image that populated each space in the matrix was fully randomized and all images from each respective category were presented before any image reoccurred. No effects were moderated by block; hence this variable is not discussed further. Upon completion of the task, participants were debriefed, thanked, and sent on their way.

Results

I excluded the responses of 15 participants due to a power outage that prematurely ended the session ($n = 2$), a fire alarm during the session ($n = 4$), or excessive error rates ($> 25\%$) on incongruent trials ($n = 9$), which yielded a sample of 92 participants (55 female, 37 male). I retained reaction times (RTs) to incongruent trials, excluded RTs for incorrect responses (9.3%), natural-log transformed RTs to adjust for excessive positive skew, and excluded transformed RTs that were 3SD above (0.68%) or below (0.43%) the mean RT. I report the results of inferential tests based on transformed data, and report descriptive statistics based on raw reaction times. I computed for each participant mean

RTs to the threatening and nonthreatening-negative targets embedded among the positive and neutral distractors.

I submitted RTs to a 2 (Target: threatening, nonthreatening-negative) x 2 (Distractor: positive, neutral) repeated measures ANOVA. A main effect of distractor, $F(1,92) = 60.24, p = .0001$, indicated that participants were faster to detect a discrepancy embedded among neutral ($M = 2060.40\text{ms}, SD = 560.34$) than positive distractors ($M = 2226.93, SD = 587.86$). More importantly, a main effect of target, $F(1,92) = 61.28, p = .0001, \eta^2 = .39$, indicated that participants were faster to detect threatening ($M = 2072.55\text{ms}, SD = 575.25$) than nonthreatening-negative stimuli ($M = 2214.84\text{ms}, SD = 576.56$). This pattern was consistent across distractors as indicated by the absence of a Target x Distractor interaction, $F(1,92) = .06, p = .8019$ (see Figure 2.1).

Discussion

Using empirically validated stimuli, the visual search paradigm indicated that people more quickly detect a discrepant image when it is threatening than nonthreatening-negative. Although these data are consistent with the possibility that threat is found more quickly (Itti, Koch, & Niebur, 1998), the visual search paradigm has two limitations. Because the task requires participants to detect the discrepant stimulus and press a key signaling stimulus detection, it is possible that participants were equally likely to detect the threatening and nonthreatening-negative stimuli but were slower to disengage from the nonthreatening-negative stimuli and press the key (West, Anderson, & Pratt, 2009). Also, the processing of threatening and nonthreatening-negative stimuli was never placed in direct competition because those stimuli did not co-occur on a given

trial. Consequently, I employ a paradigm in Study 2 that overcomes both limitations to test whether threatening stimuli preferentially capture reflexive attention.

I presented participants with pairings of all four stimulus-types (e.g., a threatening image paired with a nonthreatening-negative image) and used an eye-tracker to assess the stimulus in each pair at which participants first gazed (Rayner, 1978). Because the orienting of attention in the visual field is influenced by processing goals (West et al., 2009), I created critical trials that lacked any explicit goal other than attending to the screen. If threat preferentially captures attention, initial gaze should be drawn more frequently to threatening stimuli.

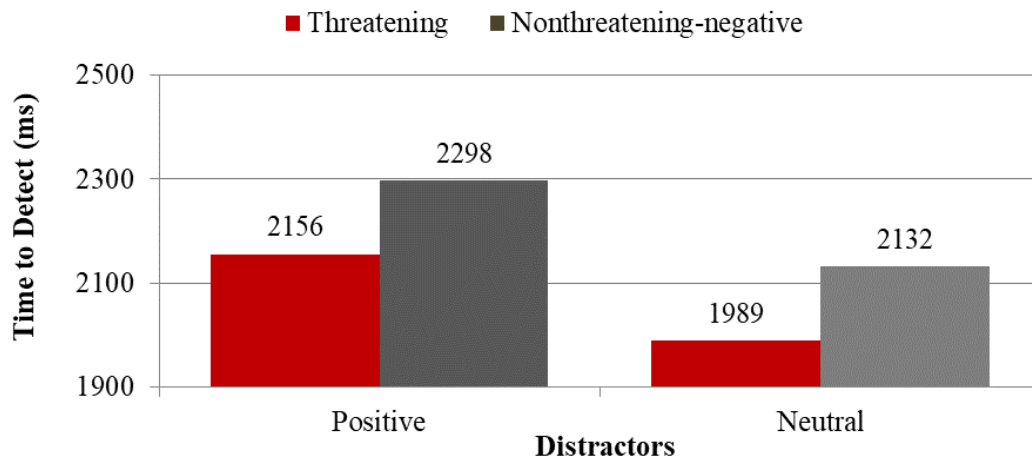


Figure 2.1. Mean visual search detection time.

Study 2

Method

Eighty-nine undergraduates (39 females, 50 males) participated for partial credit in introductory psychology. Participants sat in a computer cubicle ~60cm from a 60cm high speed, high-resolution monitor. Eye movements were recorded by a Gazepoint GP3 tracker (Gazepoint Research Inc., Vancouver, Canada) mounted below the monitor sampling at 60Hz and captured using iMotions Biometric Research Platform (iMotions Inc., Boston, MA). A 9-point calibration was administered to ensure accurate eye-tracking (12 participants who failed to track were dismissed without completing the remaining procedure).

Instructions informed participants that the study examined the perception of motion and they would complete two tasks. For the first task, they would see a grey ball appear in the middle of the screen, move around, return to the middle, and be replaced with two still images, one to the left and one to the right of where the ball was. Participants were instructed to follow the ball with their eyes until it disappeared and simply look at the screen when the images appeared. For the second task (which never occurred), participants were told the ball would move around in front of the still images – this offered an explanation as to why there were image pairs without providing an explicit goal as to what to do with the images other than look at the screen.

Participants then completed 96 trials. Each trial began with a centrally located 2cm grey ball that moved randomly around the screen for 4 s before returning to the center (ensuring eye-gaze was centrally fixated) and disappearing, at which point a pair

of 500 x 500-pixel images appeared for 4s. There were six types of image pairs from my four stimulus categories. Specifically, there were 16 trials each of threat with nonthreatening-negative, threat with positive, threat with neutral, nonthreatening-negative with positive, nonthreatening-negative with neutral, and positive with neutral. Images were left-right counterbalanced such that each stimulus type within a pair appeared an equal number of times on the left and right side of the screen. The image that populated the left or right side of the screen was fully randomized and all images from each respective category were presented before any image reoccurred. Participants subsequently completed the disgust sensitivity scale ($\alpha = .86$; Olatunji et al., 2007), which did not moderate initial eye gaze, and were thanked and debriefed.

Results

Of the 77 participants whose eyes tracked, fourteen yielded unusable data due to software malfunction ($n = 10$) or experimenter error ($n = 4$), which yielded a sample of 63 (29 females, 34 males) participants with a total of 6,048 trials. I excluded 1,486 (24.57%) unusable trials because (a) less than 80% of the trial tracked ($n = 768$, 12.70%), (b) the trial began with the participant gazing where one of the two stimuli would appear rather than centrally fixating ($n = 700$, 11.57%), or (c) the participant gazed at neither stimulus during the trial ($n = 18$, 0.30%); thereby yielding 4,562 usable trials.

Latency to the stimulus in a given pair to which participants first gazed did not vary across the six pair-types (median = 461ms). To test for an attentional bias toward the stimulus in each pair to which participants first gazed, I conducted a multilevel logistic regression using PROC GLIMMIX of SAS with a random intercept to control for the

nested trial-by-trial gazes within participants. Initial attention was more likely to be drawn to threatening stimuli than to any other stimulus type (see Figure 2.2). Participants first gazed at the threatening stimulus on 62% of trials when paired with a nonthreatening-negative stimulus, $b = 0.47$, 95% CI [0.32, 0.62], $t(4527) = 6.05$, $p = .0001$, 63% of trials when paired with a positive stimulus, $b = 0.55$, 95% CI [0.40, 0.70], $t(4527) = 7.02$, $p = .0001$, and 68% of trials when paired with a neutral stimulus, $b = 0.74$, 95% CI [0.58, 0.90], $t(4527) = 9.16$, $p = .0001$.

For pairs that did not involve a threatening stimulus, initial attention was more likely to be drawn to nonthreatening-negative than to positive or neutral stimuli. In particular, participants first gazed at the nonthreatening-negative stimulus on 54% of trials when paired with a positive stimulus, $b = 0.17$, 95% CI [0.02, 0.31], $t(4527) = 2.18$, $p = .0292$ and 62% of trials when paired with a neutral stimulus, $b = 0.50$, 95% CI [0.35, 0.65], $t(4527) = 6.35$, $p = .0001$. Finally, initial attention was more likely to be drawn to positive than neutral stimuli with participants first gazing at the positive stimulus on 60% of trials when paired with a neutral stimulus, $b = 0.41$, 95% CI [0.26, 0.56], $t(4527) = 5.24$, $p = .0001$.

Furthermore, the tendency to first gaze at the threatening stimulus paired with a neutral or positive stimulus was stronger than the corresponding tendency to first gaze at the nonthreatening-negative stimulus. In particular, the odds of first gaze were 1.28 higher for the threatening than nonthreatening-negative stimulus when each was paired with a neutral stimulus, $t(4527) = 2.26$, $p = .0238$, and 1.47 higher for the threatening than nonthreatening-negative stimulus when each was paired with a positive stimulus,

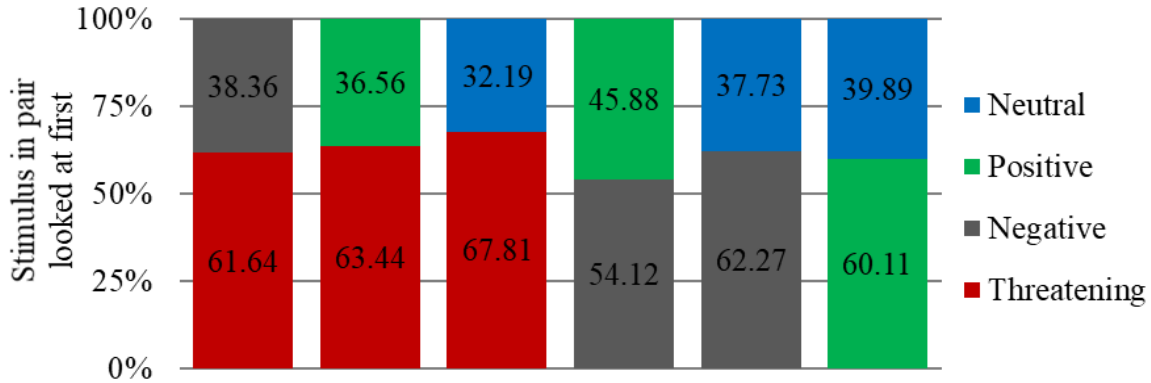


Figure 2.2. Percent of trials a stimulus in each pairing was gazed at first

$t(4527) = 3.70, p = .0002$. Similarly, the odds of a first gaze were 1.40 higher for the threatening than positive stimulus when each was paired with a neutral stimulus, $t(4527) = 3.11, p = .0019$.

Discussion

Using empirically validated stimuli, the eye-tracking paradigm indicated that initial attention (i.e., first gaze) was most strongly drawn to threatening stimuli. Participants were more likely to gaze first at a threatening stimulus when it was paired with either a nonthreatening-negative, positive, or neutral stimulus. Furthermore, when paired with a positive or neutral stimulus, threatening stimuli drew first gaze more frequently than did a corresponding nonthreatening-negative stimulus and the head to head pairing of threat and nonthreatening-negative directly revealed the greater attention capturing power of threat than that of nonthreatening-negativity. These patterns conceptually replicate and extend the findings from Study 1 and together suggest that threat elicits preferential response. In Study 3, I turn to a physiological paradigm to test whether threatening stimuli elicit uncontrolled responses more strongly than do other

stimuli. In particular, I examine a response that directly captures the presumed reflexive nature of threat processing, namely, the startle-eyeblick (Guglielmi, 1999).

Study 3

The startle-eyeblick paradigm utilizes a noise-blast (i.e., startle probe) to induce a blink during stimulus processing. Facial electromyography (fEMG) measures blink amplitude by recording electrical potential generated by the orbicularis oculi muscle responsible for closing the eye (Grillon, Pellowski, Marikangas, & Davis, 1997; Lang et al., 1990). The eyeblink reflex is characterized by rapid contraction of the orbicularis oculi, causing a blink 30-50ms after onset of a startle probe. Projections from the amygdala directly to the nucleus reticularis pontis caudalis (sensorimotor interface, located in caudal pons) influence potentiation of the startle-eyeblick reflex (Davis, 1992; Hitchcock & Davis, 1991). Evaluative information from the amygdala is carried to these brain regions, which in turn modulates startle responses (Rosen et al., 1991). As an index of amygdala activation (i.e., underlying autonomic activation of unique evaluations), the startle response distinguishes reactions to different classes of stimuli (Amodio, Harmon-Jones, & Devine, 2003; Robinson & Vrana, 2000). This paradigm is especially appropriate to my current goal given the amygdalar role in processing information related to motivationally relevant stimuli (Davis, 1992).

The startle method has been used to discriminate responses to various types of stimuli (for review see Bradley, Cuthbert, & Lang, 1999). Eyeblick responses are a combination of *affective* modulation and *attentional* interest effects (Filion, et al., 1998; Vanman, Ryan, Pedersen, & Ito, 2013). Affective modulation reflects biphasic evaluation

leading to either negative (aversive) action dispositions and startle potentiation, or positive (appetitive) action disposition and no startle potentiation. Alternatively, the attentional resource framework holds that startle amplitude decreases as relative interest increases. However, these effects often co-occur; for instance, startle responses to negative objects can involve both an interest and aversive response. Negative stimuli, being evaluatively congruent with the aversive noise-blast, enhance startle responses (Lang et al., 1990), effectively overwhelming inhibition from interest. People, for example, manifest larger startle-eyeblinks when viewing negative stimuli relative to positive or neutral stimuli at long lead intervals (i.e., when the startle probe is administered >1s after prime onset; Amodio et al., 2003; March & Graham, 2015; Vrana, Spence, & Lang, 1988; for review see Filion, Dawson, & Schell, 1998), even though interest in these stimuli is high. (Though it should be noted that these negative stimuli confounded threat and negativity.) Conversely, positive stimuli are evaluatively incongruent with the subsequent probe, and because only attention is influencing startle responses absent aversion, the startle reflex is inhibited (Filion, et al., 1998; Vanman et al., 2013). People therefore exhibit weaker startle-eyeblinks when viewing positive relative to negative or neutral stimuli at long lead intervals (Dillon & LaBar, 2005).

Although it has been suggested that startle-eyeblink discriminates between positive and negative stimuli, as discussed previously, research has not systematically disentangled the impact of threatening versus nonthreatening-negative stimuli. Given results from Studies 1 and 2, I expect threatening stimuli to enhance startle-eyeblinks relative to nonthreatening-negative, positive, or neutral stimuli. When threat is removed

from negativity, given the attentional effects just mentioned, it is unclear what impact nonthreatening-negative stimuli will have on startle-eyeblinks relative to positive or neutral stimuli.

Method

One hundred and fifty-five undergraduates (116 females, 37 males, 2 unspecified) participated for partial credit in introductory psychology. Participants were seated in a cubicle ~75cm from a 60cm monitor, affixed with stereo headphones, and screened for normal or corrected-to-normal vision, hearing, and acoustic sensitivity. Skin was lightly abraded and cleaned with alcohol to ensure proper impedance. 4mm Ag-AgCl electrodes were placed ~20mm apart over the orbicularis oculi muscle under the left eye, with a forehead ground (Blumenthal et al., 2005). EMG data were acquired with a BioPac MP36 amplifier and AcqKnowledge 4.1 software (Biopac, Goleta, CA) at a rate of 2000Hz, amplified with a gain of 5000, and notch (60Hz) and band-pass filtered (HP = 10Hz, LP = 500Hz) online. Additional stop (57-63Hz) and band-pass (HP = 28Hz, LP = 500Hz) filters were applied offline. Raw EMG data were rectified, fully integrated, and averaged over 20 samples with the root mean square.

Instructions explained that participants would view various images and occasionally hear a loud noise, with their task being to look at the images. Before the critical task, participants sampled the noise-blast – a 50ms binaural burst of 1000Hz, 100dB white noise (headphones were calibrated daily with a decibel meter). Participants subsequently completed 112 trials. Each trial began with a 1000ms presentation of a centrally located fixation “X” followed by a 6000ms presentation of a centrally located

500 x 500-pixel image, which was followed by an 8000-12000ms between-trial blank screen. 28 trials apiece were positive, neutral, nonthreatening-negative, or threatening. The image that appeared was fully randomized and no images reoccurred. On 32 critical trials (8 of each image type) the noise-probe sounded 2000-4000ms after image onset. Probes also sounded during the blank-screen between 16 trials. Presentation order of image types and occurrence of probes were fully randomized. This presentation pattern was utilized to mitigate the influence of control (i.e., predictability) and maximize the likelihood of engaging affective processes (Amodio et al., 2003; Robinson & Vrana, 2000). Participants were subsequently debriefed and thanked.

Results

Thirty-one participants provided unusable data: 19 were non-responders (i.e., did not blink in response to the noise), 8 cringed excessively thereby impeding assessment of eyeblink amplitude, and 4 requested to end the study early, which yielded a sample of 124 participants (94 females, 28 males, 2 unspecified) with a total of 3965 critical trials (one session terminated unexpectedly after the participant was exposed to 29 of the 32 critical trials).

Startle-eyeblink amplitude to a given trial was calculated by subtracting the mean fEMG amplitude across the 50ms baseline-period preceding the probe from the maximum amplitude achieved during the 200ms period following probe onset. Occasional trials are unusable (Blumenthal et al., 2005) and I excluded 574 trials across participants (14.47% of all trials) due to the absence of a blink ($n = 257$), blink during baseline ($n = 161$), or excessive orbicularis oculi movement during the trial ($n = 156$).

Exclusion was unrelated to image type, $\chi^2(3) = 1.99, p = .5726$. To control for substantial between-person variation in baseline and blink fEMG levels, startle-eyeblick amplitudes were standardized within person (Blumenthal et al., 2005). After standardization, I excluded amplitudes that varied by at least 2.5 standard deviations from the person-mean ($n = 70$) and exclusion was unrelated to image type, $\chi^2(3) = 3.18, p = .3642$. Examination of the 3321 usable trials revealed 8 participants who had responses on 50% or fewer of critical trials. Conclusions (based on direction of effects and p-values) are the same with or without those 8 participants and I report results that include those participants.

I computed for each participant mean startle-eyeblick amplitude to each of the four stimuli types (positive, neutral, nonthreatening-negative, threatening) and submitted them to a repeated-measures ANOVA. A significant stimulus effect, $F(3, 121) = 10.83, p = .0001$, indicated that amplitudes varied as a function of stimulus type. Consistent with predictions, threatening stimuli elicited a stronger startle response than did any other stimulus type (see Figure 2.3). In particular, startle-eyeblick amplitude was larger to threatening stimuli than to nonthreatening-negative, $F(1, 123) = 29.86, p = .0001, \eta^2 = .20$, neutral, $F(1, 123) = 9.72, p = .0023, \eta^2 = .07$, or positive, $F(1, 123) = 21.24, p = .0001, \eta^2 = .15$, stimuli. Furthermore, startle-eyeblick amplitude to nonthreatening-negative stimuli was smaller than it was to neutral stimuli, $F(1, 123) = 4.05, p = .0465, \eta^2 = .03$, and equivalent to positive stimuli, $F(1, 123) = 0.67, p = .4130, \eta^2 = .01$. There was no difference in eyeblick amplitude to neutral and positive stimuli, $F(1, 123) = 2.36, p = .1270, \eta^2 = .02$,

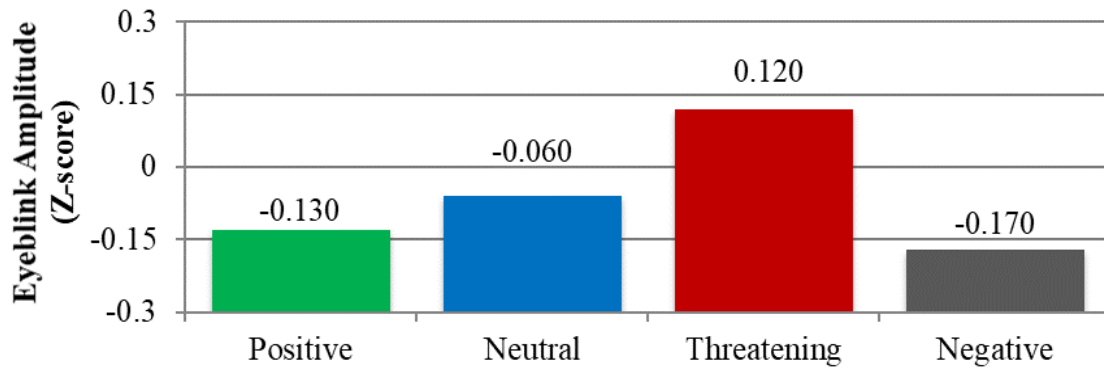


Figure 2.3. Mean standardized eye-blink amplitude as a function of image type

Discussion

Using empirically validated stimuli, the startle paradigm indicated that threatening stimuli elicited a stronger reflexive startle response than did nonthreatening-negative, positive, and neutral stimuli. These data converge with those of Studies 1 and 2 and point to a sensitivity to threat that is distinct from nonthreatening-negativity. Also noteworthy is that when threat was removed from negativity the nonthreatening-negative stimuli produced a *weaker* startle-eyeblick than did neutral stimuli. This weaker response makes sense given the argument that aversive affect and prolonged attention differentially influence startle with aversion enhancing startle (Lang et al., 1990) and attention inhibiting startle (Filion, et al., 1998; Vanman et al., 2013). Negative stimuli lacking in threat might evoke morbid fascination (Kveraga et al., 2015; Oosterwijk et al., 2015) or what could be construed as a temporary approach motivation (i.e. increased attention) to determine whether avoidance is necessary (Rimé et al., 2005; Rubenking & Lang, 2014; Turner & Silvia, 2006). The interested reader should see the below section on morbid fascination for further discussion and exploratory analysis of this possibility.

Chapter 2 General Discussion

I examined the possibility that the mind is particularly sensitive to immediate threats to bodily harm. To rectify limitations of past research, I pilot-tested stimuli to obtain images that are threatening, nonthreatening-negative, positive, or neutral and employed three paradigms assessing differential responses to those stimuli 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. These data indicate that threat elicits preferential responses in terms of fast-detection, initial-attention, and reflexive responding. The mind's apparent sensitivity to threatening stimuli has an important implication for social psychological approaches to evaluative processing. Before elaborating on that implication, however, I first consider issues regarding my stimulus categories.

Is the Threat Effect Driven by a Facet of Nonthreatening-Negativity?

My thesis is that because survival requires rapid response to threats to immediate physical harm, the mind should have evolved a preferential sensitivity to such threats (e.g., Öhman & Mineka, 2001, 2003). Consequently, I created stimulus categories that differentiated threatening stimuli from non-threatening negative stimuli (as well as positive and neutral stimuli). The curious reader might question whether it is threat per

se, or some other facet of the stimuli sets causing the noted effects. One obvious question is to whether my nonthreatening-negative stimuli could be further subdivided and whether any of those divisions yield different response patterns in regard to threat. To explore this possibility, I further categorized the nonthreatening-negative stimuli into systematic groupings. Visual inspection (see Appendix) revealed two categories: (1) dead/injured animals and (2) repulsive objects (e.g., maggots, excrement, vomit, decayed teeth). This categorization maps onto Mikels et al. (2005) analysis with dead/injured animals eliciting self-reported sadness and disgust, and repulsive objects eliciting disgust.

I reanalyzed each study to determine whether the preferential response to threatening stimuli was unique to one category of nonthreatening-negative stimuli. In Study 1, participants were faster to detect the threatening stimuli ($M = 2072.55\text{ms}$) than either dead animals ($M = 2211.49\text{ms}$), $F(1, 92) = 40.40$, $p < .0001$, or repulsive object ($M = 2236.45\text{ms}$), $F(1, 92) = 44.28$, $p < .0001$, and the latter two did not differ, $F(1, 92) = .03$, $p = .8730$. In Study 2, participants first gazed at the threatening stimulus on 59% of trials when paired with a dead animal, $t(4524) = 3.35$, $p = .0008$, and on 64% of the trials when paired with a repulsive object, $t(4524) = 5.37$, $p < .0001$, and those odds did not differ, $t(4524) = -1.45$, $p = .1461$, $OR = .81$. In Study 3, startle-eyeblick amplitude was stronger to threats ($M = .12$) than either dead animals ($M = -.20$), $F(1, 118) = 33.22$, $p < .0001$, or repulsive objects ($M = -.18$), $F(1, 118) = 18.39$, $p < .0001$, and the latter two did not differ, $F(1, 118) = .21$, $p = .6447$. Hence, the preferential response (fast detection, initial-attention, reflexive response) to threatening stimuli was not driven by a particular

facet of my nonthreatening-negative stimuli and occurred in regard to both categories of nonthreatening-negative stimuli.

For the sake of clarity, it should be emphasized that a sensitivity to threatening stimuli should manifest in terms of early (i.e., initial or fast) responses. Such an early response is functional for the detection and avoidance of immediate harm, which is why I measured such responses. If I was to examine slower, more deliberate, or delayed responses, it is possible that such responses to particular facets of nonthreatening negativity might trump those to threatening stimuli. Again, my thesis pertains to early responses and that is where I see evidence for sensitivity to threat.

Ontogeny versus Phylogeny

Öhman offered an argument for threat sensitivity in terms of a neural “fear module” (Öhman, Flykt, & Esteves, 2001; Öhman & Mineka, 2001, 2003). Such a module raises the possibility that sensitivity is conditioned uniquely or more strongly to the phylogenetic stimuli (e.g., snakes, spiders) on which it evolved than to ontogenetic threats of modern day (e.g., weapons; Blanchette, 2006; Brosch & Sharma, 2005; Fox et al., 2007). To explore this possibility, I recoded the threatening stimuli as being either ontogenetic (e.g., guns, weapons) or phylogenetic (e.g., animals, fire), and found the same functional effect for each. In Study 1, participants were faster to detect the ontogenetic ($M = 1979\text{ms}$) than phylogenetic threat ($M = 2172\text{ms}$), $F(1, 92) = 47.20$, $p = .0001$. Nonetheless, participants were faster to detect either threat type than the nonthreatening-negative image ($M = 2215\text{ms}$), $F_{\text{ontogenetic}}(1, 92) = 100.43$, $p = .0001$ and $F_{\text{phylogenetic}}(1, 92) = 5.23$, $p = .0246$. In Study 2, initial attention was more likely drawn to

both ontogenetic and phylogenetic threats than any other stimulus type. Specifically, (a) when paired with a nonthreatening-negative stimulus, participants first gazed at the ontogenetic threat on 63% of trials, $t(4524) = 5.03$, $p = .0001$, and the phylogenetic threat on 60% of trials, $t(4524) = 3.68$, $p = .0002$, and their odds did not differ, $t(4524) = 0.83$, $p = .4166$, $OR = 1.13$; (b) when paired with a neutral stimulus, participants first gazed at the ontogenetic threat on 65% of trials, $t(4524) = 5.89$, $p = .0001$, and the phylogenetic threat on 71% of trials, $t(4524) = 7.36$, $p = .0001$, and their odds did not differ, $t(4524) = -1.57$, $p = .1161$, $OR = 0.78$; and (c) when paired with a positive stimulus, participants first gazed at the ontogenetic threat on 61% of trials, $t(4524) = 4.24$, $p = .0001$, and the phylogenetic threat on 66% of trials, $t(4524) = 5.97$, $p = .0001$, and their odds did not differ, $t(4524) = -1.64$, $p = .1016$, $OR = 0.78$. Lastly, in Study 3, startle-eyeblink amplitude did not differ between ontogenetic ($M = 0.16$) versus phylogenetic ($M = 0.08$) threats, $F(1, 119) = 1.14$, $p = .2883$, and they each produced stronger startle-eyeblink amplitudes than did the positive, neutral, or nonthreatening-negative images, $F_s(1, 119) > 3.75$, $p_s < 0.055$. These data support the idea of a flexible system that learns and incorporates an expanding repertoire of what constitutes threat.

Morbid Fascination

People experience morbid fascination toward certain negative stimuli (Oosterwijk et al., 2015; Rimé et al., 2005; Turner & Silvia, 2006) whereby they pay more attention to disgusting content (Rubenking & Lang, 2014). A morbid fascination account of my negative results suggests that startle responses to negative stimuli were attenuated relative to neutral because negative stimuli both failed to evoke avoidance behaviors and

engendered information seeking behaviors (i.e. increased attentional interest). Absent a startle enhancing threat component, interest value toward negative stimuli diminished startle responses relative to neutral (Filion, et al., 1998; Vanman et al., 2013). If such is the case, my eye-tracking data should reveal participants spend more time looking at negative paired with neutral stimuli. I analyzed gaze duration data with the expectation that morbid fascination would be more evident among persons low (but not high) in disgust sensitivity. That is, even though disgust sensitivity did not moderate the stimulus in each pair to which participants first gazed in Study 2, persons high in disgust sensitivity might intentionally avoid prolonged looking at negative stimuli and hence not evidence morbid fascination.

To test this possibility, I computed for each participant the mean time spent looking at each stimulus in the negative-neutral pairing and regressed time onto a factorial crossing of mean centered disgust and stimulus (negative vs. neutral), with the latter as a within-subject variable. Consistent with morbid fascination, there was a Stimulus x Disgust interaction, $F(1, 61) = 10.16, p = .0023$, indicating that persons low in disgust sensitivity spent more time looking at the negative stimulus ($M = 2020.18\text{ms}, SE = 79.26$) than neutral stimulus ($M = 1170.63\text{ms}, SE = 82.80$), $F(1, 61) = 34.00, p = .0001$, and persons high in disgust sensitivity looked equally at the negative stimulus ($M = 1527.99\text{ms}, SE = 79.26$) and neutral stimulus ($M = 1335.73\text{ms}, SE = 82.80$), $F(1, 61) = 1.74, p = .1919$. Explained otherwise, disgust sensitivity reduced time spent looking at the negative stimulus, $b = -431.74\text{ms}, t(61) = -4.39, p = .0001$ and was unrelated to time spent looking at the neutral stimulus, $b = 144.83, t(61) = 1.41, p = .1640$. Repeating these

analyses for the negative-positive pairing and the negative-threatening pairing revealed no Stimulus x Disgust interaction for the negative-positive pairing, $F(1, 61) = 2.67, p = .1072$, nor the negative-threatening pairing, $F(1, 61) = 0.05, p = .8192$, and no tendency to differentially look at the negative ($M = 1581.50\text{ms}, SE = 62.40$) vs. positive stimulus ($M = 1476.21\text{ms}, SE = 69.84$), $F(1, 61) = 0.75, p = .3900$, nor the negative ($M = 1454.45\text{ms}, SE = 55.58$) vs. threatening stimulus ($M = 1565.69\text{ms}, SE = 51.14$), $F(1, 61) = 1.46, p = .2321$.

Morbid fascination may be functional in that it compels attention toward ambiguously negative stimuli; the gory, disgusting, or dead object may signify that a threat is nearby, or it may simply be the remnant of a past but no longer salient threat. If ambiguity exists, examining the scene to gather information is necessary to plan future action. Morbid fascination is therefore one example of an atypical reaction to negative stimuli that challenges traditional conceptualizations assuming negativity prompts avoidance.

Implications of Threat Sensitivity for Evaluative Processing

As I intimated in the introduction, current dual process models do not delineate a speed or strength difference between positive and negative implicit evaluations. An implicit response to a positive stimulus is assumed equivalent to an implicit response to a negative stimulus. The previous studies, however, suggest that such is not so and makes plausible a theoretical delineation between implicit evaluative threat processing and implicit evaluative valence processing. If evaluation is a dynamic process that unfolds over time (Cunningham, Zelazo, Packer, & Van Bavel, 2007), threat evaluation may be

temporally distinct and prior to valence processing. This implies the possibility of dual implicit processes in which an implicit threat process precedes (and potentially influences) a subsequent implicit valence process (positive vs. negative) which precedes (and potentially influences) explicit processes. In what follows, I integrate those processes into a unified Dual Implicit Process Model and demonstrate the conceptual value of distinguishing implicit threat processing and implicit valence processing.

CHAPTER 3³

THE DUAL IMPLICIT PROCESS MODEL

The *Dual Implicit Process Model* (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 I define it, i2 aligns well with notions of automatic and implicit evaluative responses in the dual-process models I reviewed earlier. Both i1 and i2 precede and potentially influence subsequent explicit (“e”) processing (see Figure 3.1).

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 non-threatening stimuli via i2 to provide more detailed evaluative responses. Given that i2 and e functioning are consistent with extant dual-process models, I focus primarily on explicating i1. The following sections detail the inputs to i1, outputs of i1, influences on the magnitudes of the outputs, and then puts it all together in the full temporal sequence of evaluative processing.

³ Much of Chapter 3 was previously published in March, D., Gaertner, L., & Olson M.A. (2018). On the prioritized processing of threat in a dual implicit process model of evaluation. *Psychological Inquiry*, 29, 1-13.

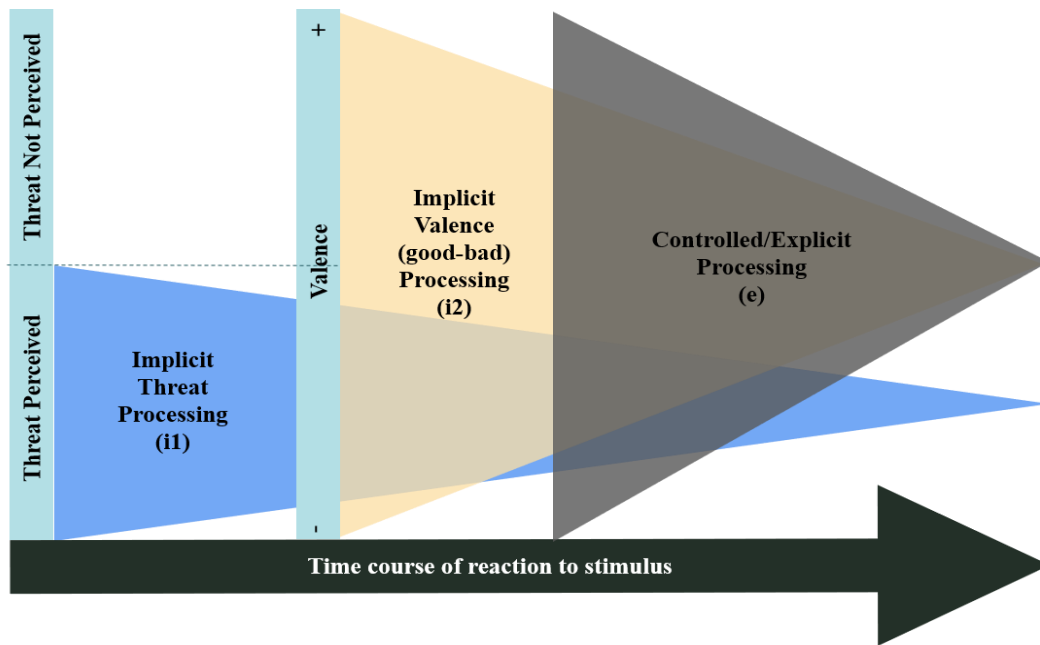


Figure 3.1. Diagram representation of the DIPM

Inputs to i1

Öhman (Öhman, Flykt, & Esteves, 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 first-hand experience and indirectly through social learning, with the amygdala playing a central role in both instances (Olsson & 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; Fox et al., 2007). In Studies 1-3, 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

i1 yields two interrelated outcomes: (1) immediate physiological reactions to protect/prepare the body from threat and (2) 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; Intura, 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 Öhman (2005), LeDoux (1996), and others (e.g., 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 effects 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., 2015; 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, Minemoto, & Blanchard, 2001). As depicted in Figure 3.2, 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.

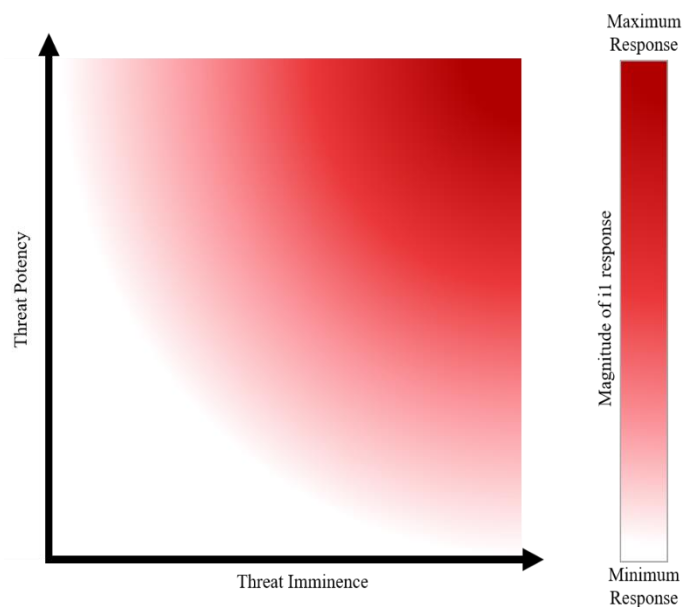


Figure 3.2. Threat potency and imminence determine magnitude of i1 response.

Finally, an interesting possibility to consider is the case of an ambiguous threat, such as a creaking floorboard at midnight, a creepy-crawling movement across one's back, or even simply being in darkness or any situation that compromises perceptual systems involved in determining object location. Properties of error management (Haselton & Buss, 2000) would lead to the over-perception of threat imminence and potency (i.e., err on the side of more imminent, and perhaps more potent), 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 series of processes from those 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, i_1 is activated, involving immediate behavioral and physiological responses preparing you to respond to the threat, but also a negative biasing effect downstream. The i_2 evaluation is now less positive than it would have been absent an i_1 event, as is your explicit evaluation.

It may appear to the reader that I am merely tacking on threat processing to dual-process models, and in a sense, I am. However, i_1 is likely to be active in many 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, I 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.

In the next chapter, I discuss the DIPM's implications for prejudice, and review two studies that explore the underlying attitudinal representation leading to prejudice toward Black Americans.

CHAPTER 4

PREJUDICE FROM THE PERSPECTIVE OF THE DUAL IMPLICIT PROCESS MODEL

The DIPM proposes that the mind preferentially processes threatening stimuli over negative and positive stimuli. The DIPM attributes such threat processing to *i1* and proposes a serial process whereby *i1* assesses the threatening stimuli and initiates responses separate from *i2*, which assesses negative or positive stimuli. In the current chapter, I describe how such a distinction may elucidate certain types of prejudice, and present two studies that test whether the *i1/i2* distinction within implicit evaluation clarifies unique prejudices toward Black Americans.

By distinguishing an implicit threat process from an implicit valence process, the DIPM suggests that implicit biases toward social groups can be functionally distinguished regarding threat versus automatic positive/negative evaluation. This shares conceptual overlap with nuanced approaches to prejudice suggesting that different intergroup contexts prompt different vulnerabilities (Schaller & Neuberg, 2012) or appraisals (Smith, 1993) that evoke different emotional and behavioral responses. This distinction implies an underlying mechanism that, in part, gives rise to the distinct evaluative and attentional reactions evoked by different stereotyped groups. From an evolutionary perspective, certain prejudices are components of risk management systems. Distinct types of risk (e.g., risk of violence, risk of contamination, ego risk, risk to economic prosperity) should promote functionally specific adaptive strategies to mitigate the risk (Cottrell & Neuberg, 2005; Mackie, Maitner, & Smith, 2009; Neuberg, Kenrick,

& Schaller, 2010; Schaller & Neuberg, 2012). For example, individuals that imply a health risk (e.g., sick individuals, homosexuals) evoke disgust and avoidance, while individuals that imply risks to group hegemony (e.g., immigrants) elicit anger. Most relevant to the current discussion, groups that imply threats to physical safety (e.g., Black-Americans, Mexican-Americans) elicit fear, at least among majority groups (Schaller & Neuberg, 2012). Groups associated with threat should therefore elicit different attentional, behavioral, and evaluative outcomes than nonthreatening groups. Such groups would activate *i1* and be processed as a survival threat in addition to the valenced processing of whether the group is liked/disliked or good/bad. Indeed, stereotypes of Black Americans often associate Black individuals with violence, criminality, and aggression, concepts that evoke fear of physical threat (i.e., *i1*; Cottrell and Neuberg, 2005). The functional specificity unique to a fear response is most likely to be expressed in threat related behaviors and responses associated with responding to perceived survival threats (Schaller & Neuberg, 2012).

Circumstantial evidence supports the idea that threat has a unique role in prejudice toward Blacks. In all these examples, threat cues ostensibly drive processing in a manner different than do merely negative or positive cues. As the DIPM proposes, an initial threat response should facilitate the processing of a subsequent threat. For example, in a shooter task in which participants are instructed to “shoot” only armed Black or White men, White Americans are more likely to mistakenly shoot unarmed Black than White men (Correll, Park, Judd, & Wittenbrink, 2002). The decision to “shoot” is ostensibly a self-protective reaction toward a threatening target. Relatedly,

priming participants with Black versus White faces specifically facilitates responding to weapons and not tools and leads to increased misidentifications of tools as weapons (Payne, 2001), indicating that Black faces share a threat association with weapons. Consistent with the idea that Black faces themselves are more threatening than White faces, White participants show stronger amygdala activation to both supraliminally (Phelps et al., 2000) and subliminally presented Black than White faces (Cunningham et al., 2004). The amygdala is particularly attuned to the initial processing of threatening information and larger activation in this scenario is thought to reflect an increased threat response.

The startle eyeblink is an amygdala-mediated defensive function meant to protect the individual from harm. As my previous research indicated (Study 3 above), people show increased startle eyeblinks toward threatening relative to neutral relative to positive/negative stimuli. Research using differently raced face stimuli has found that Black faces yield stronger startle-eyeblink than do White or Asian faces (Amodio, Harmon-Jones, & Devine, 2003), again indicating that Black faces uniquely evoke a threat response. Convergent evidence for this position also comes from research looking at attentional capture. As my previous work (Study 2 above) showed, initial attention is drawn to threatening stimuli more than to positive, negative, or neutral stimuli. Relatedly, attention is captured more quickly and held longer by Black than White faces and, tellingly, the degree to which Black faces biased attention is directly related to their perceived threat relevance (Donders, Correll, & Wittenbrink, 2008). Specifically, Black-danger stereotypes uniquely draw attention to Black faces whereas non-threat-related but

negative Black-stereotypes do not. Relatedly, in a study in which White participants rated Blacks and Whites on several trait dimensions (e.g., hostile vs. ignorant), being in a well-lit vs. dim room affects the attribution of hostility but not ignorance (Schaller, Park, & Mueller, 2003). Darkness ostensibly signaled threat imminence and affected the threat-relevance trait (hostile) but not the equally negative but non-threat relevant trait (ignorant). And lastly, subliminally priming Black versus White faces led White participants engage in a more hostile manner (as rated by an outside observer) toward another White interaction partner (Chen & Bargh 1997).

These results imply that a unique association between Blacks and threat may drive prejudice toward Blacks independent of evaluative valence. Consequently, in line with theories of intergroup emotion, conceptualizing prejudice in general valenced terms (i.e., positive/negative) may lack the necessary specificity to capture the unique causes and outcomes of distinct stereotypes. The following two studies explore the differential associations of race with valence and threat.

Study 4

The current study uses an evaluative priming paradigm to test the associations of Black and White with positivity, negativity, and threat. With evaluative priming, positively vs. negatively valenced words or images precede the presentation of a target (either a word or image) and participants are tasked with indicating whether the target is “good” or “bad”. The good/bad decision is faster when the valence of the prime matches the valence of the target than when prime and target valence mismatch. In such a paradigm, when Black and White faces (or names) are used as primes, White participants

typically have an easier time identifying positive target words preceded by White than Black faces and, conversely, negative words preceded by Black than White faces (Fazio, Jackson, Dunton, & Williams, 1995; Ito, Willadsen-Jense, Kaye, & Park, 2011). Also, Black primes facilitate the identification of negative Black-stereotypic words, while White primes facilitate the identification of positive White-stereotypic words (Wittenbrink, Judd, & Park, 2001). While these measures insinuate an association between Black and negativity, the flexibility of words as targets (more so than pictures) implies that negative words could apply to both threatening and merely negative attitude objects. For example, the word “horrible” aptly describes many negative things (e.g., guns, starving children, and dead animals). Yet, it is unclear what is activated when participants are tasked with responding to categorically vague negative terms. That is, the “bad” target likely confounds negative and threatening stimuli: a cockroach is awful and terrible, but so is a person pointing a gun. Therefore, it remains a question of whether these results are the consequence of a general Black/negative association or a specific Black/threat association.

It might not be the case that White prejudice toward Blacks is driven by a general affective “negativity”. Instead, White-Americans may predominantly associate Black-Americans with danger, which may manifest as a specific association between Black and threat, and not between Black and negativity. If portrayed in a typical priming paradigm, the DIPM’s threat versus valence distinction would predict that (1) Black more than White will facilitate the identification of threatening things, and (2) White more than Black will facilitate the identification of positive things. To the extent to which Black is

associated with threat and negativity, Black would facilitate both threat and negativity. Study 4 uses a priming task to measure such differential associations. I used a picture-picture evaluative priming task (Hermans, De Houwer, & Eelen, 1994). In particular, (the previously piloted tested) pictures of positive, negative, and threatening targets were primed by pictures of a Black or White face with the dependent variable being the speed with which participants indicated whether the target was “good” or “bad.” To the extent to which White participants more strongly associate White than Black with positivity they should more quickly respond “good” to a positive picture primed with a White than a Black face. The same logic applies to differential associations of White than Black with threat and negativity, respectively.

Methods

Eighty-One White undergraduates (51 female, 29 male) participated for partial credit in an introductory psychology course. Participants sat in individual cubicles containing a 48cm high-speed, high-resolution monitor and computer. Instructions explained that pairs of pictures would be presented quickly and sequentially with the first picture being a face and the second picture being an object and that they should indicate as quickly and accurately as possible whether the object (i.e., second picture) is bad or good by pressing the “Z” or “/” key, respectively. Participants subsequently completed 256 trials. Each trial began with a 500ms presentation of a centrally located mosaic image that functioned both as a fixation and a pre-mask, which was replaced by a 200ms presentation of a Black or White face of neutral expression, which was replaced by a 200ms presentation of a positive, negative, or threatening target, which was replaced by a

100ms presentation of the mosaic (serving as a post-mask). Finally, participants were prompted to indicate whether the target was bad or good. A 1500ms blank screen separated each trial. Participants then completed a demographics questionnaire and were debriefed and thanked.

I used 30 Black and 30 White neutral faces from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). I cropped each face to 500x500-pixels by confining the image to just above the eyebrows to just below the lips, and just outside each eye. I used 30 positive, 30 negative and 30 threatening targets from the previously pilot image sets. Six trials types resulted from the 2 prime and 3 target categories. Specifically, 64 trials of a Black or White face were followed by a positive target (128 total) and 32 trials of a Black or White face were followed by a negative or threatening target (128 total). To ensure that a “good” or “bad” response was correct on equal numbers of trials, there were twice as many positive as negative or threatening targets, respectively. Trial order was completely randomized, and every image was presented before it reoccurred.

Results

The 81 participants yielded 20,736 trials. When examining error rates, I noted that four target stimuli (1 positive, 3 negative) garnered an inordinate frequency of incorrect responses (averaged 20.61% incorrect vs. an average of 2.8% incorrect for all other stimuli). After visual examination of these stimuli, it became apparent that they were difficult to discern at the 200ms presentation duration. Therefore, I excluded the 875 (4.2%) trials involving those 4 target stimuli, yielding 19,861 trials (results below are not affected by this exclusion). I subsequently excluded the remaining incorrect trials ($n =$

415, 2%) and any trials ($n = 610$, 2.9%) with excessively slow responses (>3 interquartile ranges above the third quantile; Tukey, 1977; Wentura & Degner, 2010), yielding 18,836 usable trials. Finally, I excluded three participants who had less than 70% of their initial data remaining (results below are not affected by this exclusion). This yielded a sample of 78 participants with a total of 18,511 trials. I report the results of inferential tests based on natural-log transformed data, and report descriptive statistics based on raw reaction times.

To test whether Black or White faces differentially facilitated the identification of positive, negative, or threatening targets, I submitted the response time judgements to a $2(\text{Prime: Black, White}) \times 3(\text{Target: Positive, Negative, Threatening})$ repeated measures ANOVA. Consistent with the possibility derived from the DIP model that White participants differentially associate Black and White with positivity, negativity, or threat was the significant Prime Race $\times$ Target interaction effect, $F(2,77) = 17.41$, $p < .0001$. As depicted in Figure 4.1, White participants responded (a) faster to positive targets preceded by a White ($M = 665\text{ms}$) than Black ($M = 681\text{ms}$) face, $F(1,78) = 41.41$, $p < .0001$, $d = .72$, (b) faster to threatening targets preceded by a Black ($M = 670\text{ms}$) than White ($M = 681\text{ms}$) face, $F(1,78) = 7.14$, $p = .0092$, $d = -.30$, and (c) equally fast to negative targets preceded by either a Black ($M = 691$) or White ($M = 693$) face, $F(1,78) = .36$, $p = .5492$, $d = -.067$. These data suggest that White-Americans' prejudice toward Black-Americans is driven by both a positive association with White and a threat association with Black, but not a mere negative evaluation of Blacks. There was no Prime Race effect, $F(1,78) = 1.38$, $p = .2939$, though there was a Target effect, $F(2,77) = 15.67$,

$p < .0001$, indicating that negative targets ($M = 692$) were responded to more slowly than were positive ($M = 673$), $F(1,78) = 20.66$, $p < .0001$, or threatening targets ($M = 675$), $F(1,78) = 25.23$, $p < .0001$.

One might suggest that the previous association of Black and threat is simply a replication of the tendency to associate Black with weapons (Payne, 2001). Keep in mind that my threat stimuli consist of images of weapons and threatening animals. If the prior results are merely a weapons effect, then Black pictures should facilitate response to guns but not threatening. Such was not the case. Black (versus white) facilitated responding to both guns (Black $M = 648$ ms vs. White $M = 656$ ms), $F(1, 78) = 3.34$, $p = .0716$, and animals (Black $M = 685$ ms vs. White $M = 697$ ms), $F(1, 78) = 4.52$, $p = .0368$, and such facilitation did not differ (i.e., Black vs. White x Gun vs. Animal), $F(1, 78) = .01$, $p = .9397$. Stated otherwise, Black facilitated responses to threatening targets not just guns. The pattern is broader than a mere weapons association.

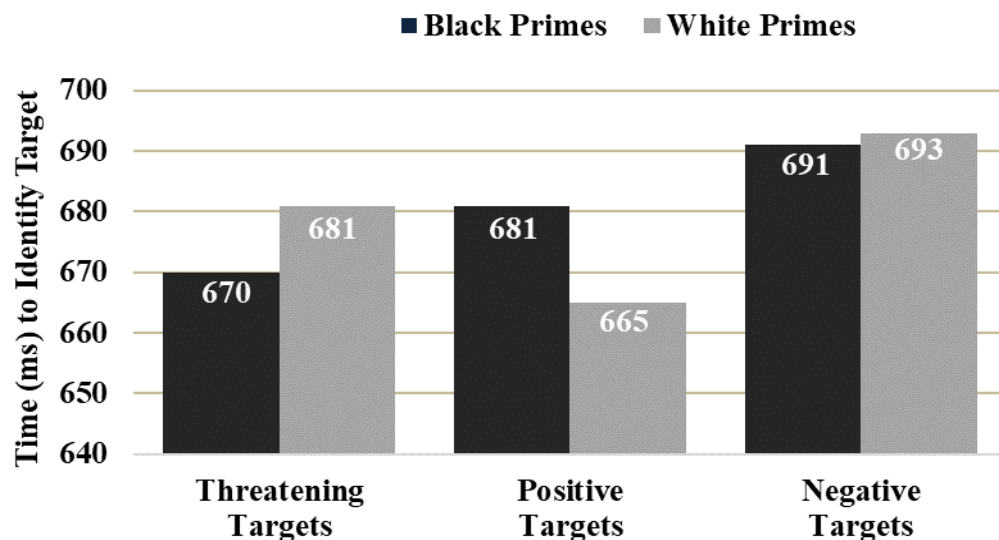


Figure 4.1 Mean reaction times to respond to each Prime x Target pairing

Discussion

In line with the DIPM threat vs. valence distinction, these data suggest that White American's associate Blacks with threat and Whites with positivity. Interestingly, no differential association as a function of race emerged for merely negative targets. These results imply that certain prejudices, particularly those towards groups whose stereotypes contain elements that connote physical threat, may arise from a fear-based reaction, and not simply an unspecified negative reaction. The lack of a Black/Negative association found here contrasts previous research that have found such an effect (e.g., Fazio et al., 1995). As detailed in the introduction to the current study, those studies used Black and White primes and subsequent word targets. It is possible that those negative words could also have been captured by threat (e.g., horrible; i.e., a threatening stimulus could facilitate "horrible" as could a negative stimulus). By using easily classifiable pictures, the current study isolates the category of the target to something more specific than valence and more so directly tests the association between Black/White and such categories.

A methodological limitation of the current study provides the possibility for an alternative explanation. Because I only used pictures of Blacks and Whites as primes, it is possible that the threat-association is not linked specifically to Blacks, and, instead, is an effect that occurs in response to all outgroups. To assess this possibility, I conducted a second study using the same basic procedure but with two changes: (1) I primed the target images with Black, White, *and* Asian men and (2) I replaced the original gun pictures with new gun pictures (that I pilot tested). The original gun pictures depicted

blurred Black and White men holding the guns in ecologically realistic looking scenarios (importantly the previously reported results did not differ as a function of the race of the gun holder). In the new pictures (which I took myself), the race of the person holding the gun is not visible (gloves and long sleeves obscure the skin).

The results from this conceptual replication again found stronger a positive/White than positive/Black association. I also found a stronger positive/Asian than positive/Black association, and no difference between Asian and White positive associations. This implies that the positive association is not merely lacking when comparing White to any outgroup but is specifically lacking from Black relative to White or Asian. Yet, I failed to find any group differences in associations toward threat or negative targets. That is, Asian, Black, and White primes did not differentially facilitate the identification of negative or threatening targets.

There are three possible explanations for this failure to replicate: (1) the effect from Study 4 may be measurement noise, weak, or unreliable, (2) the addition of a third prime category may have altered how participants approached the task or affected their race-target associations, (3) though I was able to create several combinations of shirt and glove colors with different types of guns, the images lost their ecological validity, instead appearing more sterile and like each other than the previous gun images. This may have led to quick habituation toward these images and a weakened threat-response. I am currently replicating the unaltered original paradigm. In study 5, I again explore associations among Asian, Black, and White with threat, negativity, and positivity but do

so using a methodology that improves upon evaluative priming, and reaction time measures in general.

Study 5

A limitation of traditional reaction time measures of evaluation (e.g., the priming measure used above) is that they only capture the output of the evaluation process. That is, they record an overall reaction time to click a key yoked to the identity/category of some class of object relative to some other class of object. Such responses give no information about the process that led to the click of that key, only how long it took to engage in such a behavior. This obscures the process(es) that take(s) place over the time-course of evaluation toward that outcome. As I subsequently discuss, to determine if the response to threat indeed occurs first (and does not just result in the faster click of a key), an *online* measure of response is necessary.

Mouse-tracking is an example of an online measure that dynamically captures responding in real-time as it is occurring. Mouse-tracking involves recording and analyzing the on-screen pathway that people take en route to choices. Within a mouse-tracking study, participants are presented with images and are tasked with moving the mouse to click one of two labels that correspond to the images. Participants respond by moving the mouse from a fixed “Start” position located at the bottom-center of the screen to click on one of two options positioned respectively in the top left and right corners of the screen.

For example, imagine a mouse tracking study where the task is to categorize people as “Black” or “White”; once a participant clicks “Start” they are presented with an

image (e.g., a White man) and would proceed to move the mouse toward and click on the label (i.e., “White”) that correctly identifies the presented person. Recorded is the (a) x- and y-coordinates of motion, (b) latency of motion, and (c) selected option. Across trials, aggregated latencies and trajectories are computed and provide empirical insight into response competition (i.e., when a stimulus activates more than one response such that both responses compete for dominance). Response competition is a product of how much each stimulus is associated with the respective labels, the target label and the distractor label. For instance, in a recent study, participants classified images of racially typical or racially ambiguous (e.g., light-skinned African American) “Black” and “White” faces while their mouse trajectories were recorded (Freeman & Ambady, 2011). As the amount of racial ambiguity increased (i.e., as skin-tone lightened among Blacks or darkened among Whites), hand trajectories showed related increases in attraction toward the incorrect race-category (i.e., increased mouse movement toward Black for dark-skinned Whites, and toward White for light-skinned Blacks). This indicates that response competition increased as ambiguity increased. Mouse tracking is unique in that it allows the researcher to map and measure the impact of such response competition over time.

By mapping these real-time dynamics, I can assess how multiple processes temporally interact and how unique information sources differentially influence evaluation across the decision-making process. For example, if using mouse-tracking to classify combinations of angry or happy, Black or White faces (e.g., angry Black, angry White, happy Black, happy White) based on their expressions (e.g., identifying happy faces as “cheerful” vs angry faces as “dangerous”), two processes are interactively

affecting categorization: (1) the race of the face (Black vs White), and (2) the expression on the face (angry vs happy). Both the race and expression of a face function as separate influences of processing which may manifest as response competitors. In a mouse-tracking paradigm, these separate characteristics of a stimuli may compete for responses if the expression on the face is incongruent with the stereotype of the race (e.g., anger on a White face, happiness on a Black face) or, alternatively, may synergize responses if the expression on the face is congruent with stereotypes of the race (e.g., anger on a Black face, happiness on a White face). I will subsequently describe an analysis technique wherein I analyzed responses across the time-series to estimate *when* in the time-course the processing of unique characteristics (e.g., facial expression) of the stimuli began to affect evaluation. By looking at when mouse movement paths diverged from each other and the size of that divergence, I estimated when (i.e., early vs late in the time-course) and how much (i.e., in terms of magnitude difference) response competition affected evaluative response.

I present participants on each trial with a picture of an Asian, Black, or White man displaying an angry, sad, happy, or neutral facial expression, and the participant must move the mouse to choose one of two labels that describes the man (dangerous, depressed, cheerful, or calm). For example, when the face is angry the correct target label is dangerous and across blocks that label is paired with each of the other labels (depressed, cheerful, or calm). Likewise, when the face is neutral the correct target label is calm and across blocks that label is paired with each of the other labels (dangerous, depressed, cheerful).

The time-course of the participant's decision can provide insight as to whether White American's prejudice is driven by unique associations between Black and threat, or whether such a threat association is a more general outgroup effect (i.e., extends to Blacks and Asians), and whether non-threatening negativity is also involved. In particular, if White Americans associate Blacks with threat, then the presentation of a Black face should produce differential response competition between the selection of dangerous vs calm, cheerful, or depressed. If the Black face is angry, for example, the latter labels should distract less from the movement to "dangerous" than the extent to which "dangerous" distracts from the movement to the other labels when the Black face is not angry. If such patterns are a more general outgroup effect, then they should similarly occur for Asian faces.

In the current study I mapped the *time-course* of White participants processing of angry, happy, neutral, and sad Asian, Black, and White faces. Through this design, I examined how people's *early* reactions to differently raced faces differed as a function of negative, neutral, positive, and threat. Though the subsequently described analysis I estimated when those early reactions began to differ. By including Asian faces, I tested the unique influence of Black beyond that of simply any out-group face. By including two types of negative face along with a positive and neutral face, I tested the unique influence of threat beyond that of mere valence.

Pilot Study

No high-quality face databases could be located that contained all of the necessary facial expressions, so I created my own. I initially gathered 20 neutral faces of each race (Asian, Black, and White) from the Chicago Face Database (Ma et al., 2015). To create four expression categories for each face (angry, happy, neutral, and sad), I created angry, happy, and sad templates within a face morphing program (i.e., FaceGen). the same template to each face to ensure that a given expression displayed roughly equal intensity across the faces (e.g., all angry faces were equally angry). This application process involved several steps. One at a time, I imported 20 neutral faces from each race (60 total) into FaceGen. Importing each face involved designating 11-points on the model image that correspond to 11-points on a race-matched template 3D head prebuilt within FaceGen. The program then overlays the model's skin over a fungible 3D template head. First, the just-imported neutral face was exported to ensure that it matched the look and feel of each of the emotionally morphed faces (in terms of the digitalization resulting from importing/exporting from the program). Each emotion template was then applied, one at a time, at which point each newly morphed emotional face was exported. This process resulted in 60 angry faces (20 Asian, 20 Black, and 20 White), 60 happy face, 60 neutral faces, and 60 sad faces (240 total). These images were subsequently uniformly cropped to 450 x 650-pixels.

I subsequently piloted the cropped 240 images (80 Asian, 80 Black, and 80 White) and collected 165 between-subjects ratings of how angry, happy, or sad each face looked. I excluded 5 participants who responded faster than 500ms on >20% of their

trials, resulting in 160 participants providing 38,053 ratings. I deleted individual ratings that were faster than 500ms ($N = 666$, or 1.75%) or slower than 10000ms ($N = 279$, or .73%), resulting in 37,108 usable ratings. Based on visual examination, I excluded 12 models (each of their 4 faces; 2 Asian, 3 Black, & 7 White) from subsequent analyses due to face-morphing that caused them to appear abnormal (e.g., double nose, teeth bared, severe eye occlusion). I first calculated a mean score of each rating (e.g., angry, happy, sad) for each face (i.e. every face had 3 mean ratings). I then created Z-scores for each face within its emotional expression X rating type grouping using that group's mean and standard deviation. For example, separate Z-scores were created within the anger images for each rating of anger, happiness, and sadness, rendering three Z-scores for each face. I subsequently deleted 6 models (2 Asian, 1 White, and 3 Black) whose mean rating of any one of their three emotional faces fell greater than 2SD below the mean of any of their three group means (note, no face had a rating greater than 2SD above its group mean). I then excluded three models (3 Asian) whose neutral faces were rated greater than 2SD above the neutral group mean on any rating, leaving 14 Asian, 14 Black, and 12 White models. Lastly, I visually excluded models until each race group contained 10 models. These faces were used in the subsequently described main study (see Appendix for all images).

Method

One-hundred and twenty undergraduates participated for partial credit in an introductory psychology course. Participants sat in individual rooms containing a 48cm high-speed, high-resolution monitor and computer. Since there is no “neutral” expression

label, I chose not to use the terms “happy”, “angry”, or “sad” to identify each respective expression. I instead instructed participants that I was “trying to find pictures of faces that can be quickly identified as either: Cheerful, Dangerous, Depressed, or Calm. So that they knew how to identify each face, each expression was first described: “Cheerful faces look happy, friendly, joyful. Dangerous faces look angry, scary, threatening. Depressed faces look sad, gloomy, unhappy. Calm faces look emotionless, neutral, flat.” Each block presented only two facial expressions (e.g., angry vs. neutral) and their accompanying labels (e.g., “Dangerous” vs. “Calm”; see Figure 4.2). Participants were reminded of the description for the appropriate labels just prior to each block. They were told that for every trial they would see an image on the screen, and their task was to use the mouse to as quickly and accurately as possible classify the face by clicking the correct label. Each trial began with the appearance of two labels and a “Start” button. Once participants click the start button, a single face image appeared. The image disappeared once participants click within the boundaries of one of the two labels or 2000ms have passed. To ensure online processing, participants received a warning if they failed to begin moving the mouse before 400ms. Before the critical blocks, participants undertook one block of 10 practice trials where they classified pictures of fruits and vegetables as “fruit” or “vegetable”.

Participants experienced six blocks of trials. Each block contained 60 trials presenting a pair of facial expressions and their associated labels (e.g., “Dangerous” vs. “Calm”). For the first 30 trials, one label was located on the left and other on the right side of the screen, and those labels subsequently switch sides on the second set of 30

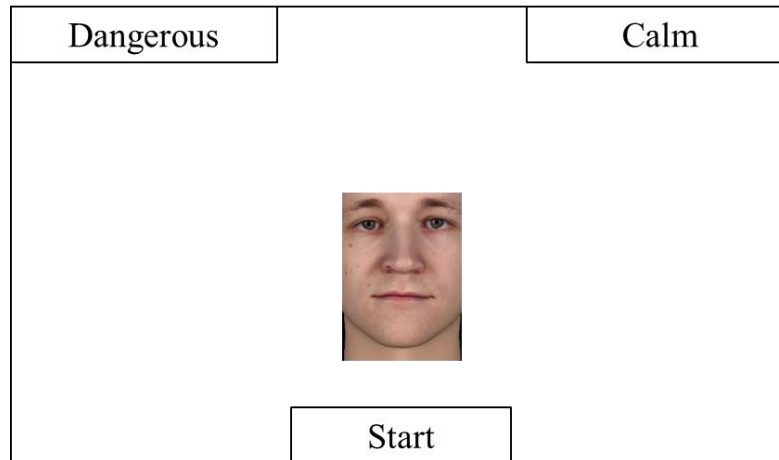


Figure 4.2. Sample trial from Study 5.

trials (with the starting position, left vs. right, randomized across participants). On each trial, participants viewed a face (either Asian, Black, or White) displaying one of the two expressions corresponding to the labels for that block. Participants were tasked with moving the mouse to the label that reflects the expression displayed on the face. That is, one label reflects the target expression (i.e., the one displayed on the face) while another label functioned as a distractor (i.e., not displayed on the face). Across the six blocks, each expression equally served as a target and a distractor (angry-sad, angry-neutral, happy-sad, happy-neutral, or sad-neutral) totaling 360 trials. Across blocks, expressions occurred equally within and across each race. The order of the six blocks was randomly assigned. Following the six blocks, participants completed a basic demographics questionnaire (race, gender, age), and were debriefed and thanked.

Data Preparation

The data were imported into the mouse-tracking analyzer program, which (a) equated mouse paths for trials in which the target-expression label was on the right versus

the left by horizontally re-mapping all right-side trials to the left, and (b) separated each trial into 20ms bins corresponding to the refresh rate of the mouse (e.g., a 1000ms trial would yield 50 bins and a 2000ms trial would yield 100 bins). Data were then imported into SAS for subsequent analyses. The 120 participants provided a total of 41,907 trials. I subsequently deleted 770 (1.8%) incorrect trials and 421 (1%) trials on which participants timed out (i.e., took longer than 2000ms). As mouse-tracking is an online measure, it is important that participants do not wait until they have decided before moving the mouse toward their response. Therefore, it is typical to delete trials where participants fail to quickly initiate movement. I subsequently deleted 2,549 (6%) trials on which participants first movement was greater than 300ms. Lastly, I deleted 1,414 (3.3%) trials that were completed abnormally fast (< 600 ms). This yielded 36,753 usable trials. Note that patterns of the below results do not change as a function of these exclusions.

Each 20ms time bin contains the X- and Y-coordinate position of the mouse across the duration of the trial (e.g., see the dashed line in Figure 4.3) The mouse tracking software also provides for each trial metrics that describe the shape and total time of the response. Specifically, each trial has an associated (1) reaction time (RT), which is the duration from clicking the start button to clicking a response, (2) area under the curve (AUC), which is the total area between a hypothetical straight line running from the start-button to response and the actual path taken from start to response, (3) the maximum deviation (MD), which quantifies the maximum distance from the hypothetical straight line to the actual response path, and (4) the maximum deviation time (MD Time), which is the time in the trial at which maximum deviation occurred.

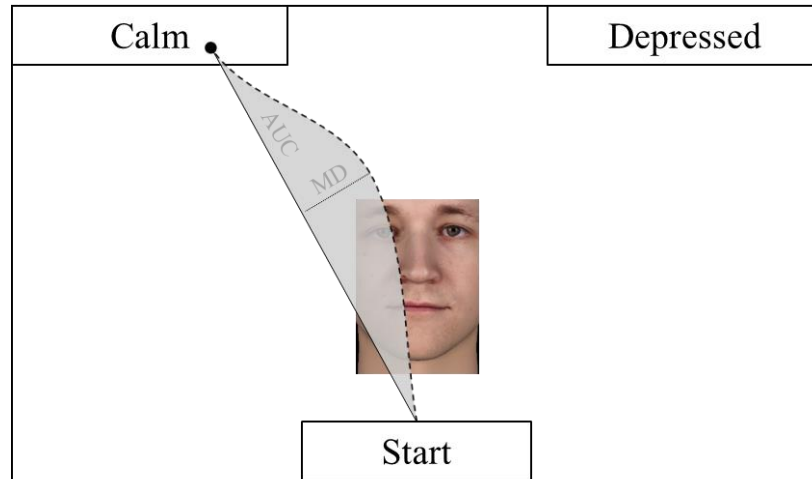


Figure 4.3. Visual representation of a trial response.

Each metric provides insight into a particular aspect of the mouse path. However, none of the default metrics provide hypothesis relevant information regarding the point in time at which participants began moving *relatively closer* to the target versus distractor label. Using Figure 4.3 as an example, notice that any pure vertical movement brings the mouse simultaneously closer to both the target and distractor labels. Likewise, simple horizontal movement is insufficient as well. For example, an early leftward movement from the start certainly brings the mouse closer to the target label (i.e., calm). But a leftward movement at the bottom of the screen is further from the target than is the same leftward movement at the top of the screen. Hence, what is needed is to calculate at each data-point the relative Euclidean distance of the mouse from the target expression vs. the distractor expression.

The Euclidean distance reflects the straight-line distance between two points in space and is given by the formula:

$$d_i = \sqrt{(x_i - x_l)^2 + (y_i - y_l)^2}$$

This formula uses the x and y coordinates to derive d , the distance of the mouse cursor at every i -th time point from the last time point, l . In addition to Euclidean distance to the selected response, it is also necessary to account for the distance to the unselected response. Recall as the mouse moves vertically along the Y-axis it is simultaneously getting closer to the selected and unselected responses at the same rate. Therefore, to calculate the relative pull along the horizontal axis toward both options, it is necessary to also calculate the Euclidean distance to the unselected response. This is done by inverting the X-values at every i -th time point in the above formula. From these two values I derived a Euclidean distance difference score by subtracting the distance to the distractor from the distance to the target response. As can be seen in the top panel of Figure 4.4, the difference score forms a sigmoidal-shaped (i.e., s-shaped) curve over time. The initial horizontal movement along the curve at or around zero is actually vertical movement bringing the mouse equally closer to both target and distractor labels. This period is followed by a positive exponential slope that reflects movement closer to the target than the distractor. The curve eventually asymptotes as the mouse-cursor reaches the selected response.

The same s-curve is found in a number of literatures, particularly those interested in modeling patterns of microbial and bacterial growth in different environments. These fields have generated several equations for estimating biologically significant parameters in the bacterial growth curve with the goal of predicting food safety and shelf life. These models are derived from earlier forms of these equations describing sigmoidal growth

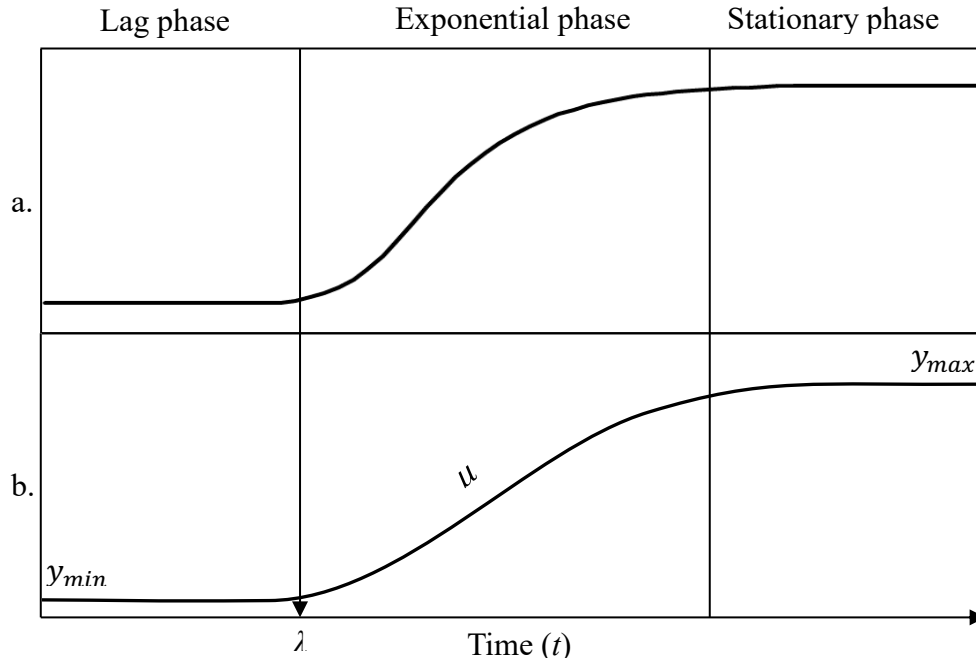


Figure 4.4. (a) Actual estimated difference of Euclidean Distances. (b) Typical bacterial growth curve.

curves simply in terms of mathematical parameters. Two widely used equations are the Gompertz (Gibson, Bratchell, & Roberts, 1988; Buchanan & Phillips 1990; Garthright, 1991) and the Baranyi (Baranyi & Roberts 1994; Baranyi, Robinson, Kaloti, & Mackey, 1995). Reparametrizing the original equations occurred by deriving biological parameters as a function of the mathematical parameters and then including them in a new equation (e.g., Borglin, et al., 2012; Zwietering, Jongenburger, Rombouts, & Riet, 1990). Several studies have examined the relative performance of these two models and have shown them both to be highly accurate (Baty, Delignette-Muller, 2004; Buchanon, Whiting, & Damert, 1997; Swinnen, Bernaerts, Dens, Geeraerd, & Van Impe, 2004), with consistency of the predicted parameters depending on the quality of the dataset (i.e., many time-points). The reparametrized Gompertz and Baranyi formulas are:

$$y_t = y_{min} + y_{max} e^{\left\{-e \left[\frac{\mu_m e}{y_{max}}(\lambda - t) + 1\right]\right\}} \quad \text{Gompertz}$$

$$y_t = y_{max} + \ln \left(\frac{-1 + e^{\mu_m \lambda} + e^{\mu_m t}}{(-1 + e^{\mu_m t}) + e^{(\mu_m \lambda + y_{max} - y_{min})}} \right) \quad \text{Baranyi}$$

In these formulas, y_{min} represents the lower asymptote of the growth curve, y_{max} is the upper asymptote, t is time, and μ_m represents the maximum growth rate ($\ln$ = natural log; e is a mathematical constant whose value is approximately 2.718). Most relevant to the current study, these reparametrized models also provide an estimate of the *lag period* or *lag phase* (λ), which in biological terms is the adjustment period when bacterial cells acclimate to a new environment and transitions from zero growth to exponential growth (see bottom panel Figure 4.4). The point at which the lag phase ends, what I herein refer to as the lag time, describes the beginning of exponential growth. When applied to the current data, lag time describes when mouse-path responses began to move relatively closer to the target than distractor. To get the precise timing of this early turn toward the selected response, I adapted this method to estimate and subsequently analyze lag time.

While it may seem that MD-time could provide a close approximation for the final turn toward the selected category, in reality the MD often occurs at a time other than the initial turn toward the selected response. For example, in Figure 4.5, the light line represents the actual path (taken from the current study), the solid black line reflects the hypothetical straight path, and the short black line at the top of the path represents the MD. For comparison, I have copied the same short black line to a point of maximum deviation closer to the initial turn. As can be seen, the actual MD occurs later in time than the initial turn toward the chosen category. The early deviation is smaller than the MD as

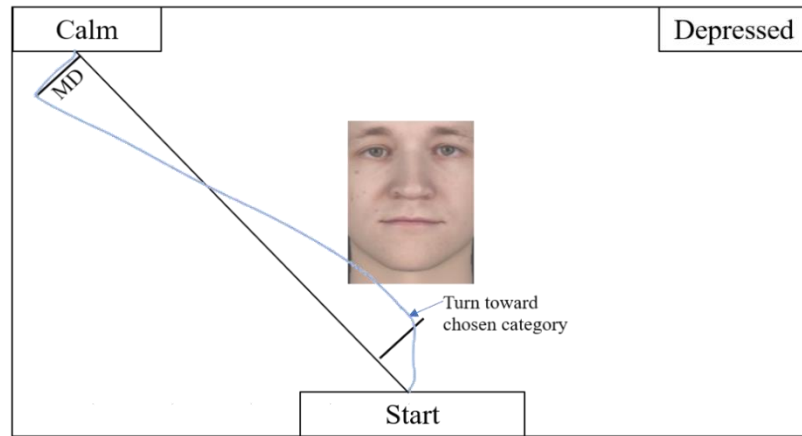


Figure 4.5. Maximum deviation location

indicated by the black line protruding beyond the response path. Therefore, MD-time will often give an incorrect account of when the path actually began to turn toward the chosen response. As will be shown below, lag time provides an accurate estimate of this hypothesis relevant information.

Each subject experienced 10 trials of each of the 12 target-label and distractor-label pairings for each of the three races (e.g., 10 trials of dangerous paired with depressed for a Black angry face). I estimated a lag time for each subject's mouse movement to each of the 36 pairings. To do so, I first created an average response path for each subject across the ten trials of a given race, target, and distractor pairing (e.g., sad Asian with depressed and cheerful labels). For each average response path, I calculated the difference in Euclidean distance to the target versus distractor at each time point.

I fit each participant's time by Euclidean difference data to the Gompertz model and Baranyi model, respectively, using SAS Proc NLIN (a non-linear regression procedure).¹ For each model, SAS estimated parameter values for y_{min} , y_{max} , μ_m , and lag time. I then averaged the estimated tag time from the Gompertz and Baranyi models for each of the 36 target-distractor pairings for each participant.²

Results

Because each target label was paired on a given trial with one of the other three labels as distractors, target and distractor were not fully crossed (e.g., dangerous was paired with cheerful, depressed, and calm, but not with itself, as in dangerous vs. dangerous). Consequently, I submitted lag times for each target label to separate 3(race) x 3(distractor-label) multi-level ANOVAs using SAS Proc Mixed. For each analysis, I used Kenwood-Rodgers degrees of freedom and log-likelihood tests to determine which random effects, beyond a random intercept, were necessary. In no instance would models converge with a random race effect and or a random Race x Distractor effect. All models included a random intercept and random effect for distractor. For ease of interpretation, I present the analyses of lag time separately within each target facial expression. See the

¹ Proc NLIN allows inputted starting values to facilitate the iterative estimation of model parameters. I generated starting values by fitting in Excel the average model (i.e., averaging across all participants). Proc NLIN allows parameter estimates to be restricted within a reasonable boundary. Lag time has to be within 0 and 2 seconds (i.e., it could not occur before a trial starts or after it ends) and y_{max} has to be between a Euclidean distance of 0 and 2 (the smallest and largest values in the units used by the MouseTracker program). Conclusions based on p-values and direction of effects are the same with and without the latter two bounds, but using the bounds lead to higher convergence rates and, hence, the retention of more data (i.e., 99.46% vs 97.04% convergence rates for Gompertz and 99.65% vs 97.23 for Baranyi).

² I averaged the predicted lag time from the Gompertz and Baranyi models only when both models converged. Of the original 4,226 trial-type means, 23 failed to converge in the bound Gompertz model and 15 failed to converge in the bound Baranyi model. The 4,199 that converged for both models had an average pseudo R-squared of 94.28% indicating excellent model fit. Conclusions based on p-values and direction of effects were the same when estimated lag-times from poorer fitting models (i.e., $R^2 < 80\%$) were excluded. The reported results include all estimates.

Appendix for supplementary analysis of other the mouse-tracking metrics (MD, AUC, MD Time).

Angry face. When viewing angry faces, there were main effects of race, $F(2,687) = 18.85, p < .0001$, and distractor, $F(2,229) = 18.88, p < .0001$, and no interaction, $F(4,687) = 1.08, p = .3650$. As depicted in Figure 4.6, the race effect indicates when the face was angry, Black faces led to shorter lag times than did White or Asian faces. In particular, persons moved *earlier* in time to dangerous when the angry face was Black ($M = 498\text{ms}$) than when it was White ($M = 527\text{ms}$), $t(687) = -4.30, p < .0001$, or Asian ($M = 538\text{ms}$), $t(687) = 5.95, p < .0001$, and lag time to angry White vs. Asian faces did not differ, $t(687) = 1.63, p = .102$. The lack of an interaction indicates that the earlier lag to dangerous for angry Black faces (vs angry White or Asian faces) did not vary as a function of the distractor label (cheerful, depressed, or calm). The distractor effect indicates that when the face was angry, cheerful had less influence on the lag time to dangerous than did depressed or calm. In particular, persons moved *earlier* in time to dangerous when it was paired with cheerful ($M = 499\text{ms}$) than when it was paired with calm ($M = 518\text{ms}$), $t(229) = -2.40, p = .017$, or depressed ($M = 547\text{ms}$), $t(230) = -6.10, p = .0001$ (see also below Figure 4.10 and Table A.1 in Appendix).

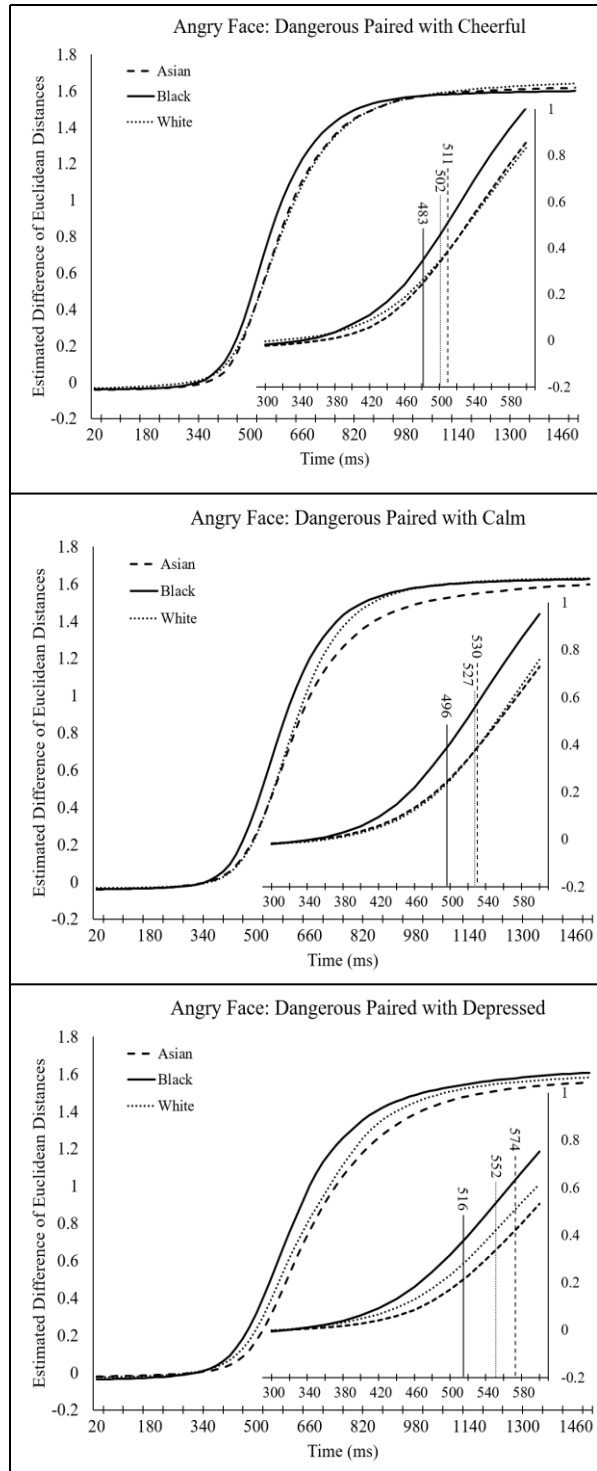


Figure 4.6. Euclidean distance difference scores plotted across time for angry faces (and area-of-focus on lag time) as a function of Race and Distractor.

Sad face. When viewing sad faces, there was no main effect of race, $F(2,691) = .07, p = .9349$, but there was a main effect of distractor, $F(2,230) = 30.71, p < .0001$, and an interaction, $F(4,691) = 7.84, p < .0001$. As depicted in Figure 4.7, the distractor effect indicates that when the face was sad, cheerful had less influence on the lag time to depressed than did dangerous and calm. In particular, persons moved *earlier* in time to depressed when it was paired with cheerful ($M = 493\text{ms}$) than when it was paired with either dangerous ($M = 539\text{ms}$), $t(230) = 5.04, p < .0001$, or calm ($M = 563\text{ms}$), $t(230) = -7.72, p < .0001$.

The interaction indicates that movement to depressed occurred *later* in time when it was paired with dangerous and the sad face was Black ($M = 573\text{ms}$) than either Asian ($M = 519\text{ms}$), $t(688) = -3.81, p = .0002$, or White ($M = 523\text{ms}$), $t(690) = 3.56, p = .0004$, and dangerous did not have a differential effect on sad Asian vs. White faces, $t(690) = -.24, p = .8117$. Conversely, movement to depressed occurred *earlier* in time when it was paired with cheerful and the sad face was Black ($M = 477$) than either Asian ($M = 498\text{ms}$), $t(692) = 2.12, p = .0344$, or White ($M = 513\text{ms}$), $t(692) = -3.14, p = .0001$, and cheerful did not have a differential effect on sad Asian vs. White faces, $t(691) = -1.02, p = .3081$. Movement to depressed did not differ across the three race faces when it was paired with calm, $ts < 1.6, ps > .11$.

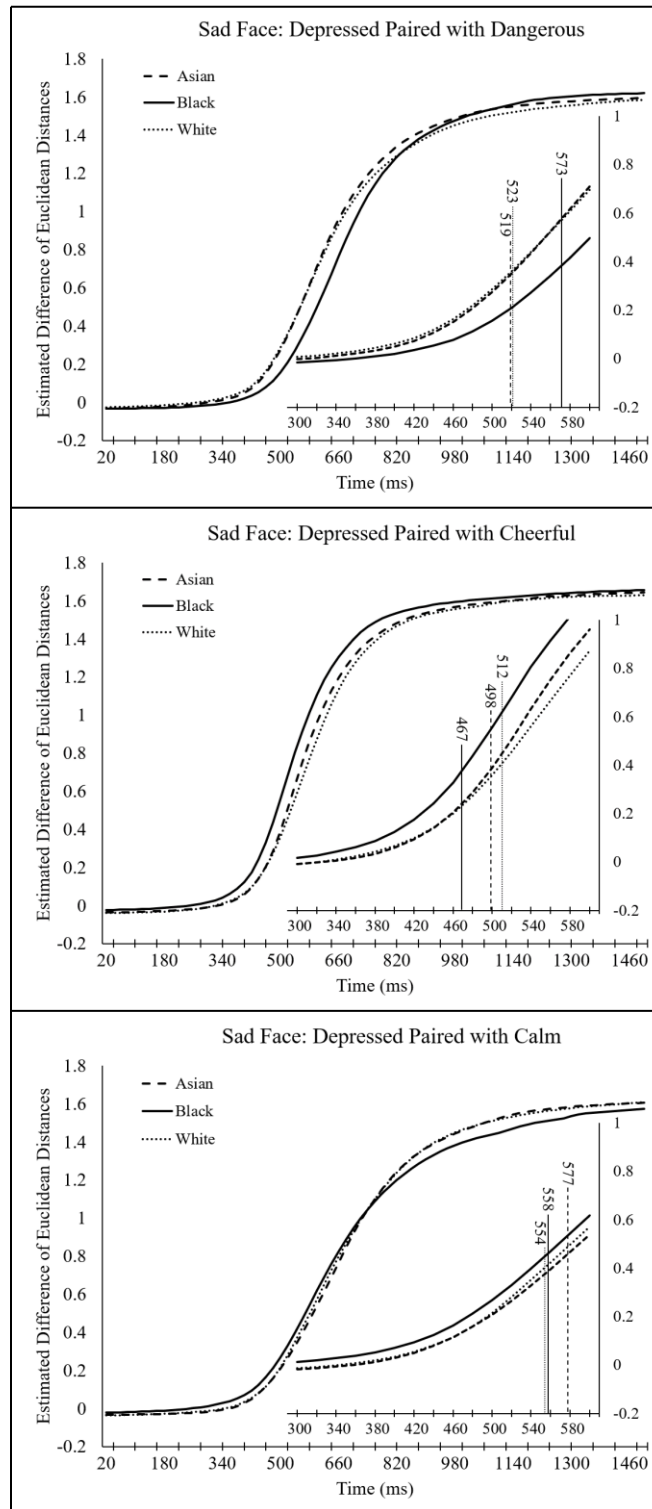


Figure 4.7. Euclidean distance difference scores plotted across time for sad faces (and area-of-focus on lag time) as a function of Race and Distractor.

Happy face. When viewing happy faces, there was no main effect of distractor, $F(2,233) = 1.99, p = .1391$, but there was a main effect of race, $F(2,694) = 18.30, p < .0001$, and an interaction, $F(4,694) = 4.78, p = .0008$. As depicted in Figure 4.8, the race effect indicates that when the face was happy, Black faces led to longer lag times than did White or Asian faces. In particular, persons moved *later* in time to cheerful when the happy face was Black ($M = 539\text{ms}$) than when it was White ($M = 498\text{ms}$), $t(694) = 5.75, p < .0001$, or Asian ($M = 506\text{ms}$), $t(694) = -4.51, p < .0001$, and lag time to happy White vs. Asian faces did not differ, $t(695) = 1.24, p = .2137$.

The interaction indicates that movement to cheerful occurred *later* in time when it was paired with dangerous and the happy face was Black ($M = 544\text{ms}$) than either Asian ($M = 490\text{ms}$), $t(693) = -4.49, p < .0001$, or White ($M = 483\text{ms}$), $t(691) = 5.04, p < .0001$, and dangerous did not have a differential effect on happy Asian vs. White faces, $t(693) = .53, p = .5933$. Likewise, movement to cheerful occurred *later* in time when it was paired with depressed and the happy face was Black ($M = 548\text{ms}$) than either Asian ($M = 498\text{ms}$), $t(697) = -4.13, p < .0001$, or White ($M = 501\text{ms}$), $t(691) = 3.91, p = .0001$, and depressed did not have a differential effect on happy Asian vs. White faces, $t(693) = -.23$, paired with calm, $ts < 1.86, ps > .06$.

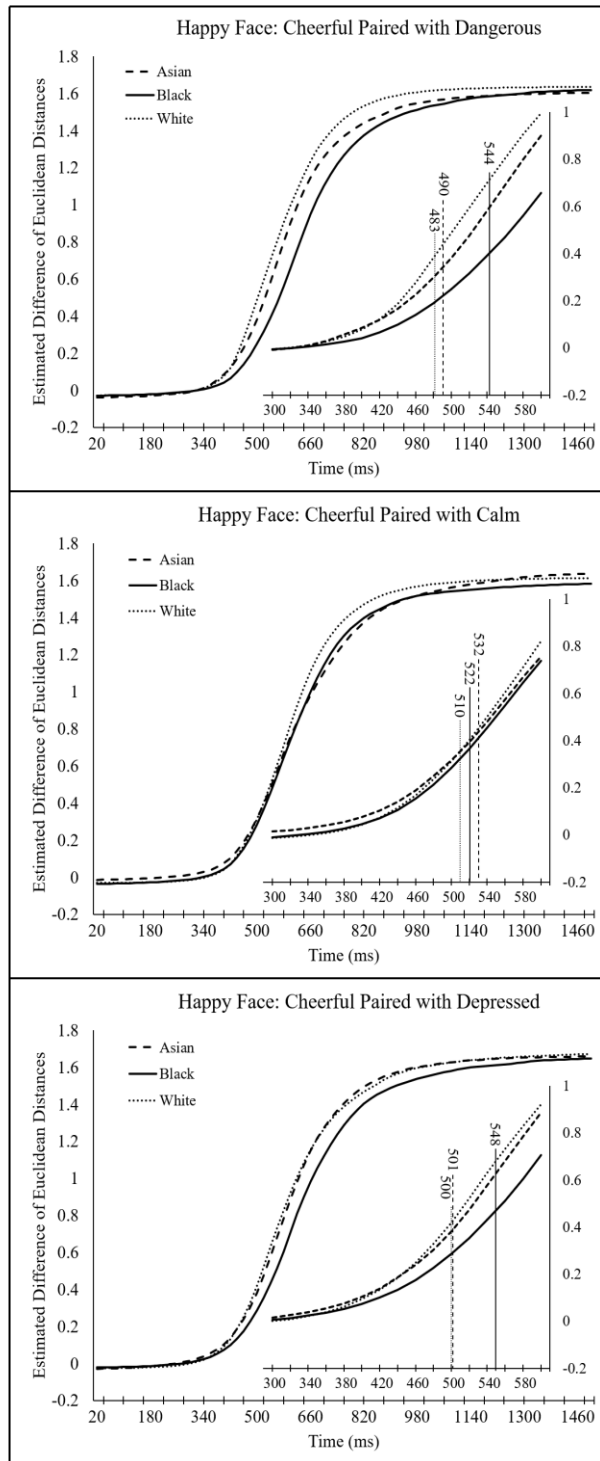


Figure 4.8. Euclidean distance difference scores plotted across time for happy faces (and area-of-focus on lag time) as a function of Race and Distractor.

Neutral face. When viewing neutral faces, there was a main effect of race, $F(2,692) = 11.40$, $p < .0001$, a main effect of distractor, $F(2,231) = 5.48$, $p = .0047$, and an interaction, $F(4,692) = 6.12$, $p < .0001$. As depicted in Figure 4.9, the race effect indicates that when the face was neutral, Black faces led to longer lag times than did White or Asian faces. In particular, persons moved *later* in time to calm when the neutral face was Black ($M = 554\text{ms}$) than either White ($M = 527\text{ms}$), $t(691) = 4.44$, $p < .0001$, or Asian ($M = 531\text{ms}$), $t(692) = -3.77$, $p = .0002$, and the lag time to neutral White vs. Asian faces did not differ, $t(693) = .66$, $p = .5125$. The distractor effect indicates that when the face was neutral, depressed had more influence on the lag time to calm than did cheerful or dangerous. Specifically, persons moved later in time to calm when it was paired with depressed ($M = 550\text{ms}$) than either cheerful ($M = 525\text{ms}$), $t(231) = -3.30$, $p = .0011$, or dangerous ($M = 536\text{ms}$), $t(260) = -1.85$, $p = .0655$, and the lag time to calm paired with cheerful vs. depressed did not differ, $t(230) = 1.45$, $p = .1483$.

The interaction indicates that movement to calm occurred *later* in time when it was paired with dangerous and the neutral face was Black ($M = 568\text{ms}$) than either Asian ($M = 524\text{ms}$), $t(690) = -4.14$, $p < .0001$, or White ($M = 516\text{ms}$), $t(690) = 4.90$, $p < .0001$, and dangerous did not have a differential effect for neutral Asian vs. White faces, $t(690) = .76$, $p = .4492$. Likewise, movement to calm occurred later in time when it was paired with depressed and the neutral face was Black ($M = 573\text{ms}$) than either Asian ($M = 549\text{ms}$), $t(692) = -2.31$, $p = .0209$, or White ($M = 530\text{ms}$), $t(695) = 4.09$, $p < .0001$, and depressed did not have a differential effect for neutral Asian vs. White faces, $t(693) =$

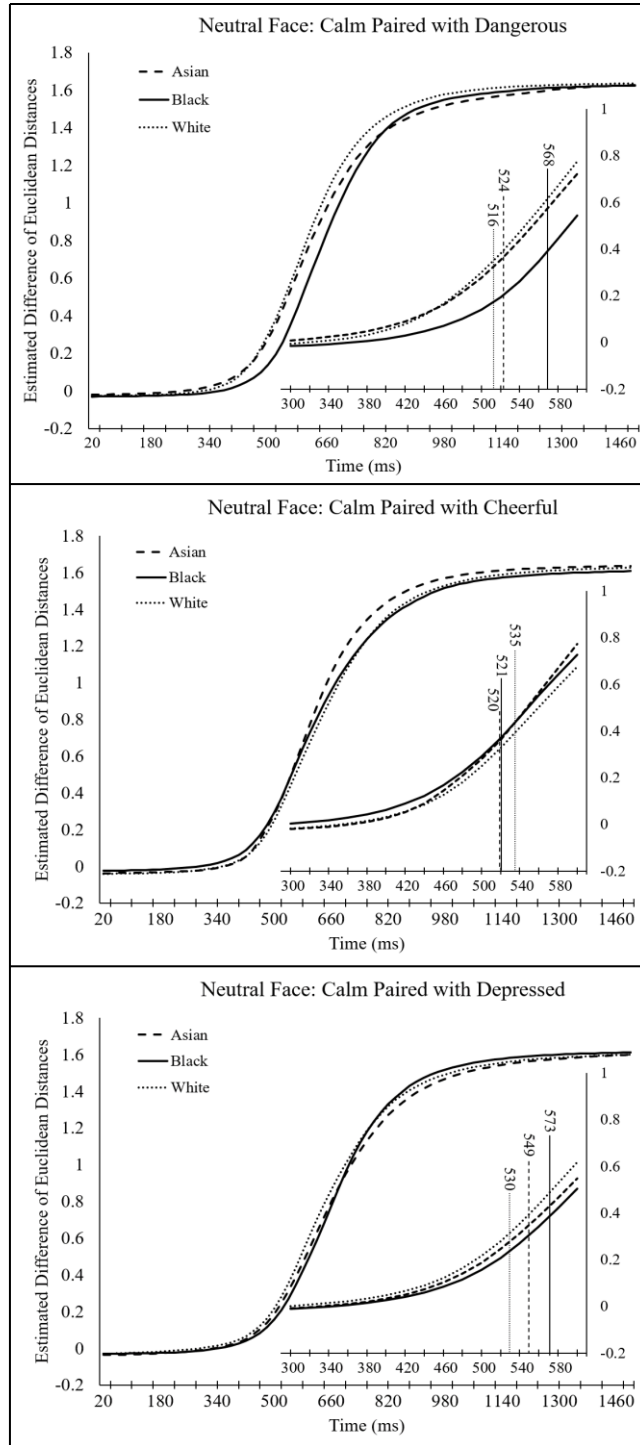


Figure 4.9. Euclidean distance difference scores plotted across time for neutral faces (and area-of-focus on lag time) as a function of Race and Distractor.

1.79, $p = .0742$. Movement to calm did not differ across the three races when it was paired with cheerful, $t_s < 1.41$, $p_s > .16$.

Discussion

In the current study, I used mouse-tracking to test whether differential associations between races (Black, White, Asian) and attributes (Dangerous, Cheerful, Depressed, Calm) affected how people evaluated these groups. Different groups have unique stereotype profiles, and these unique profiles should lead to specific evaluative associations. Most relevant to the DIP perspective, there are prevalent stereotypes associating Blacks with violence/criminality. Therefore, I expected that responses to Black faces would be specifically affected by the presence of the Dangerous (i.e., threat-relevant) label, either as a target or distractor. Alternatively, Whites and Asians are more associated with positive traits, and hence it was expected that responses to these groups would be uniquely influenced by the presence of the Cheerful (i.e., positive) label. By including Asians, I tested whether these effects relate to specific groups as more than simple in- versus out-group bias. To explore these predictions, I used mouse-tracking to see when in time responses turned toward the correct classification of these groups for each target by distractor pairing. By looking at lag times and comparing these across race and distractor pairings, I can assess both the association between each race and the target attribute and the association between each race and the distractor. In this way, as the results below outline, both earlier and later lag times convey useful information about the relative strength of such associations.

If indeed Black is associated with threat, a Dangerous target would lead to *earlier* lag times for Black than White or Asian faces, and conversely, a Dangerous distractor would lead to *later* lag times for Black than White or Asian faces. This hypothesis was consistently supported by the patterns of results presented above. Specifically, when the face was angry, regardless of the distractor (Cheerful, Calm, Depressed), participants moved to the Dangerous label *earlier* in time (i.e., shorter lag time) when responding to Black than to White or Asian faces (top row of Figure 4.10). Conversely, when Dangerous served as a distractor to non-angry faces (Happy, Sad, Neutral), people turned toward the target label *later* in time (i.e., longer lag time) when the face was Black than when it was White or Angry (first column of above Figure 4.10). In other words, when responding to threat was the goal, responses to angry Black faces were inflexible and efficient, however, responses to non-angry Black faces were influenced by the threat association. Such a primacy of threat vs. negative processing is most apparent when comparing how Dangerous distractors competed within responses to sad (i.e., negative) Black faces versus how Depressed distractors failed to compete with responses to angry Black faces. These patterns indicate both that the processing of the threat evoked by Black faces is insulated from response competition evoked by other distractors and, more so, that the Black/threat association interfered with responses to ostensibly innocuous Black faces. This supports the DIPM supposition that i1 threat processing is unique and processed efficiently in a manner relatively unaffected by response competition. Alternatively, sad expressions ostensibly processed by i2 are susceptible and affected by alternative more relevant processes. More so, responses to White and Asian faces were

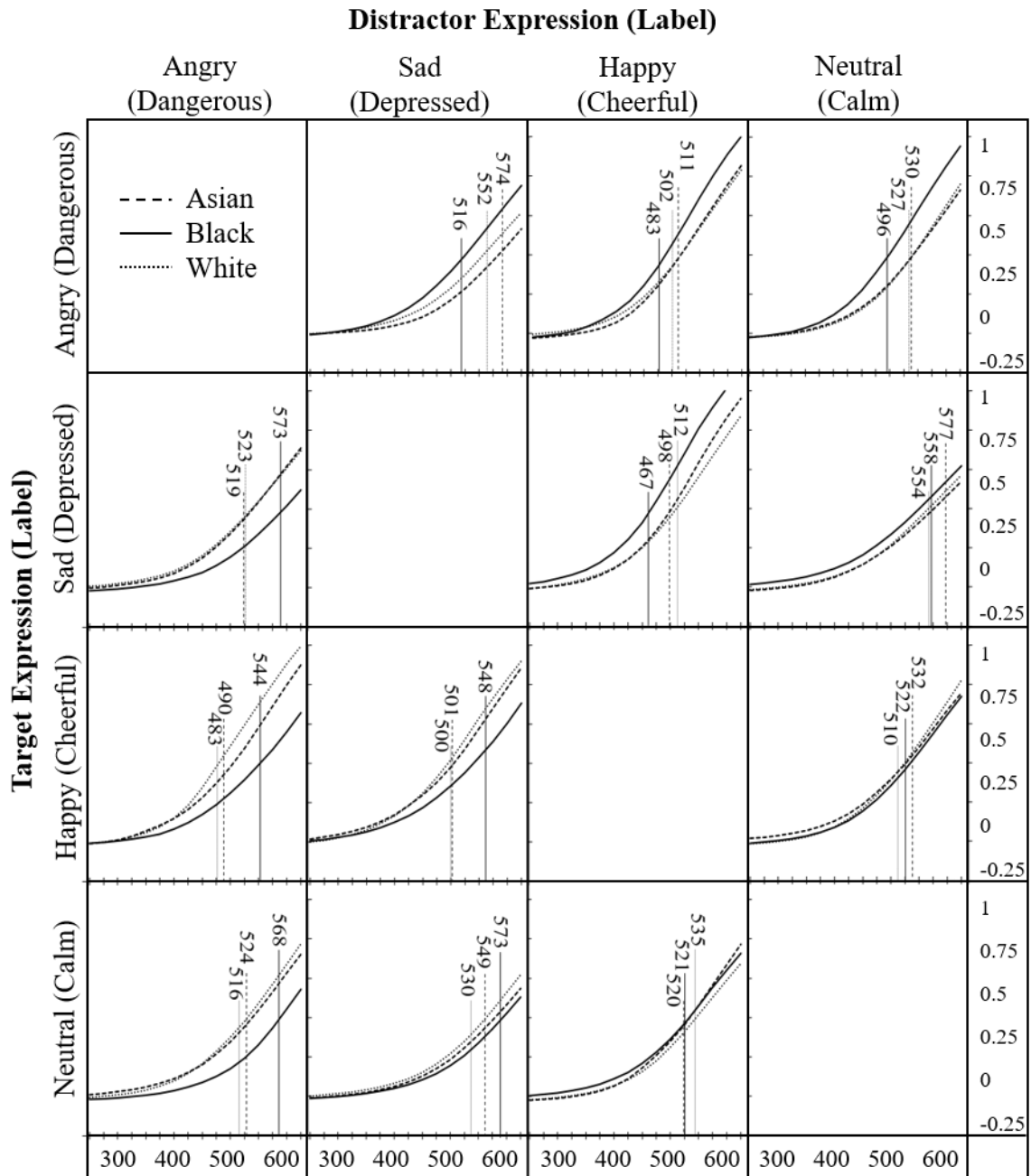


Figure 4.10. Area-of-focus on lag times of each race in every level of Target by Distractor. Note: White and Asian lines often overlap.

equal within each of these pairings, indicating that the prior patterns are not simply an outgroup effect but are unambiguously linked to Black faces.

In addition to a Black/threat association, I also posited the possibility of positive associations (i.e., cheerful/happy) with White/Asian, and perhaps a non-threatening negative association (i.e., depressed/sad) with Black. Given each groups stereotype profile, it was expected that Whites would more strongly associate White and Asian with happiness. In addition to (or because of) the Black/threat association, it is also possible that Whites associate Black with negative attributes in general. These expectations were generally supported. Specifically, when the target face was sad (second row of Figure 4.10), people turned toward Depressed *earlier* for Black than White or Asian faces when Cheerful was the distractor. Yet, when the target face was happy (third row of Figure 4.10), people turned *earlier* toward Cheerful for White and Asian than Black faces when Depressed was the distractor. Together these patterns seem to indicate (a) a stronger association between White and Asian than Black with positivity, (b) a stronger association between Black than White or Asian with sad, or (c) both. This interpretation is partially supported by people showing *later* turns toward Calm for neutral Black than White or Asian faces when Depressed was the distractor. However, people showed no differential response as a function of race to either (1) Cheerful or Depressed when Calm served as the distractor or (2) Calm when Sad served as the distractor. These latter patterns would seem to indicate indicate no differential association between Black, White, or Asian with happiness or sadness. To rectify these patterns, it may help to consider the DIPM proposal that i1 threat can manifest as i2 negativity. In that light, it

could be that non-threatening target by non-threatening distractor patterns showing differential lag time to Black faces (e.g., earlier: sad faces paired with cheerful distractor; later: happy and neutral faces paired with sad distractor) are partially driven by the i1 threat response manifesting as a congruently negative downstream process (i.e., sadness). In these blocks, there is a negatively valenced label which may serve as an outlet for such biased processing. In other pairings lacking a negative label (e.g., Happy face with Neutral distractor, Neutral face with Happy distractor), such an i1 bias on i2 has no consequence as there is no negative label for which any i1 negative can affect.

Taken as a whole these results indicate that the underlying mechanism driving response to Black faces was a Black-threat association, not a Black-negative association.

Implications of the DIPM to Prejudice

I used a priming task and a mouse-tracking task to examine White participants' differential associations of Black, White (Study 4), and Asian (Study 5) with positivity, threat, and negativity (Study 4), and neutral (Study 5). Given tendencies for in-group favoritism, I expected White individuals to more strongly associate Whites (and perhaps Asians) than Blacks with positivity. More interestingly, it might not be the case that White prejudice toward Blacks is driven by a diffuse affective “negativity”. Instead, Whites may fear Blacks, which may manifest as a specific association between Blacks and threat, and not between Black and negativity. White participants had an easier time identifying threatening targets after Black faces in a priming paradigm, and more quickly moved toward the threat identifier if the face was Black. Also, participants had a harder time identifying any Black face if the paired-expression was anger. Taken together, these

results imply a specific Black-threat association, and not simply a Black-negative association.

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; IAT: 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 et al., 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 DIP model 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).

I 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 provides a framework for future research to uncover the extent to which a given measure of prejudice assesses threat-related responses (as I earlier implied the shooter-bias might; Correll et al., 2002) vs. i2-related evaluative responses (as IAT 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 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, for example, in terms of violations of traditional values). This speaks directly to the distinctions between several theories of prejudice such that Black may be associated with bad and not threat (e.g., symbolic racism), associated with threat (e.g., Schaller & Neuberg, 2012), and/or associated with bad without specifying a bad/threat distinction (e.g., aversive racism). Additionally, this distinction implies that different interventions might be developed to address the specific underlying bases of the prejudice based on this distinction. For example, contact hypothesis (Allport, 1954) proposes that positive outgroup contact can foster positive attitude change (Pettigrew & Wright, 2011; Zhou, Page-Gould, Aaron, Moyer, & Hewstone, 2018). In line with exposure therapies aimed at reducing phobic responses (discussed below), positive exposure to feared outgroups may alter threat responses (a sort of desensitization) but not necessarily affect other negative

responses, leaving those prejudices intact. Yet, as discussed below, threat-responses may be more difficult to unlearn, suggesting that threat-based prejudices may be more difficult to reduce.

CHAPTER 5⁴

GENERAL DISCUSSION, OPEN CONCEPTUAL QUESTIONS, AND IMPLICATIONS FOR OTHER LINES OF RESEARCH

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 previous section applied the model to prejudice. The following sections examine some of those insights to phobia and intimate partner violence. I end with a discussion of some open conceptual questions about the model.

Implications of the DIPM

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 (Davies, 1997). Some treatments appear able to reduce the magnitude of fear responses, often extinguishing them entirely (Hermans, Craske, Mineka, & Lovibond, 2005). 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 i_2 and e , with less influence on i_1 . In particular, the spontaneous return of fear suggests that the phobic target remains associated with threat via i_1 despite becoming more favorably evaluated via i_2 and e (Hermans et al., 2005). The temporary reduction of fear responses is likely a highly-controlled process through explicit behavior

⁴ Much of Chapter 5 was previously published in March, D., Gaertner, L., & Olson M.A. (2018). On the prioritized processing of threat in a dual implicit process model of evaluation. *Psychological Inquiry*, 29, 1-13.

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, 2009; Vasey, Harbaugh, Buffington, Jones, & Fazio, 2012). Further emphasizing that what is being affected 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 i_1 . Interestingly, snakes and spiders elicit negative implicit evaluations from snake- and spider-neophytes and positive implicit evaluations from snake- and spider-experts (Ellwart, Rinck, & Becker, 2006; Purkis & Lipp, 2007). Such a positive (i_2) evaluation (and presumably positive explicit evaluation) might contribute to the unlearning of an i_1 input over time. Yet, there remains an open question about whether any evaluative association can be unlearned (Ebbinghaus, 1885/1964; Petty, Tormala, Briñol, & Jarvis, 2006), and it would certainly be a survival benefit for threat associations to be particularly inflexible 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 i_2 and e) may only affect i_1 indirectly as a result of changes in i_2 and e evaluations. For example, talk therapy ostensibly focuses on changing i_2 evaluations, while exposure therapy could be thought to alter i_1 responses to a threat. By changing the i_1 output (if not the i_1 association), people may “learn” to effectively control their reactions to phobic stimuli. But importantly, changes in i_2 and e are not diagnostic of a reduction in the i_1 phobic evaluation

evidenced by spontaneous recovery. Phobias may therefore be less susceptible to certain types of cognitive therapy (i.e., talk). Phobia treatments that alter the i1 response (i.e., exposure) 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 versus the objective negativity of abuse are often explained through 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 amount of self-reported emotional attachment, feelings of love, loyalty, and 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. Interestingly, research has shown that increased emotional abuse (ostensibly affecting *i2*) is more predictive of 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 their *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.

Implications for Suicide

Typical models of suicide and suicidal ideation make no formal distinction between relatively automatic/implicit vs more controlled/explicit processing (e.g., Prinstein, 2008). Yet, dual-process models have shown that incorporating this distinction can yield substantial advances in the prediction and explanation of human behavior across a variety of domains (Sherman, Gawronski, & Trope, 2014). The study of suicide and suicidal ideation is ripe for analogous gains in prediction and explanation. Specifically, interpersonal theory (Van Orden et al., 2010) has pointed to “acquired capability” (i.e., the capacity to engage in suicidal behaviors) as a critical predictor of

suicide. From a DIPM perspective, suicidal behaviors would entail overcoming seemingly well-ingrained motives to avoid threat. Yet, the DIPM indicates that threat processing is particularly efficient and pervasive, occurring prior to other forms of automatic processing. Theoretically, many experiences (e.g., combat exposure) could influence acquired capability, and are theorized to result in increased pain tolerance and decreased fear of death. This could lead to a diminishing of the automatic threat-avoidance response. This requires overcoming evolved and socially reinforced associations between threat and avoidance. From the perspective of dual-process theories, this should be difficult and largely impervious to deliberative processing (e.g., Rydell & McConnell, 2006). But a DIPM perspective would suggest that long-term, repeated “practice” via self-harm or exposure to lethal objects should have more impact on acquired capability through gradual changes in threat-avoidance associations.

The DIPM would predict that suicidal behaviors entail overcoming seemingly well-ingrained motives to avoid threat and suggests that experiences involving exposure to threat influence an acquired capability to engage in suicidal behaviors. If indeed individuals increased pain tolerance and decreased their fear of death dampens its reflexive threat response, the DIPM suggest that suicidal ideation response treatments may be more effective by targeting and even enhancing the evolved and socially reinforced associations between threat and avoidance. Indeed, if people can increase their acquired capability through gradual changes in threat-avoidance associations, treatments may be able to be developed that do the opposite, which may possibly decrease ones’ capacity for suicidal behaviors.

Open Questions

There remain a number of open conceptual questions regarding the DIPM that necessitate further discussion. I 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 effect such that preceding i1 activation strengthens subsequent i1 activation (see Figure 5.1). 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 Whites 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 section on “Implication for Prejudice”). 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.

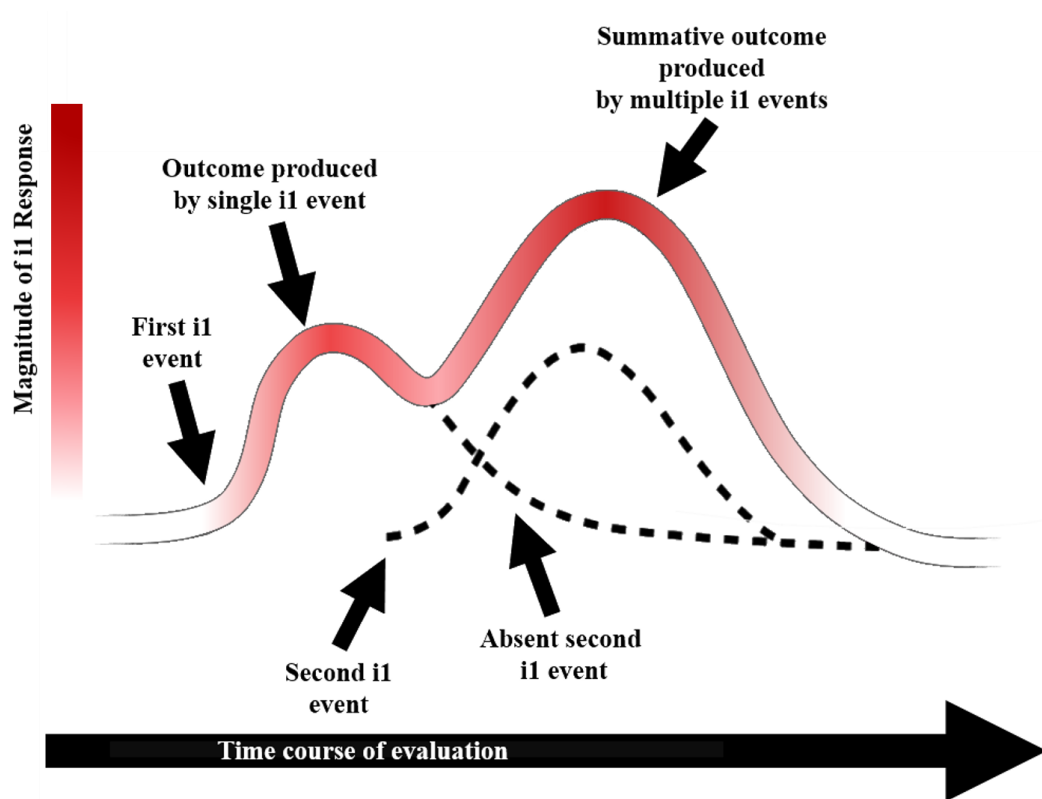


Figure 5.1. Magnitude of i1 response across time.

Dual Implicit Processing in a Simple vs. Complex World

In their now classic paper, Murphy and Zajonc (1993) demonstrated that both angry and smiling faces presented at 4ms yielded valence-consistent affect 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 affect misattribution at only a 4ms exposure, how could there be room for earlier processing of a threatening stimulus? In my 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, non-threatening stimuli are processed directly via i2. So, Murphy and Zajonc (1984) 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 counter-stereotypic thoughts in response to a Black face reduced the automatic activation of prejudice. Thus, e appears capable of downregulating i2. However, consistent with my discussion of the phobia research above, 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) suggest 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 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 amygdalar activity, it may have revealed that initial activity to Black faces was equal across sub- and supra-liminal conditions, and that only over time (i.e., after i2 and explicit processing) was amygdala activity dampened.

Individual Differences in i1 Processing

LeDoux (2015) argues 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, while 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). Additionally, individuals with PTSD show biased attention and quicker reaction to threatening stimuli (Block & Liberzon, 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 over-perceive threat. For example, White et al. (2010) showed that high- versus low-anxious

individuals have a processing advantage for threatening words on a lexical decision task. Pitman and Orr (1986) aversively 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 et al. (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

I 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 and lends insight into several psychological research domains.

、

REFERENCES

- Anderson, M. A., Gillig, P. M., Sitaker, M., McClosky, K., Malloy, K., & Grigsby, N. (2003). "Why Doesn't she just leave?": A descriptive study of victim reported impediments to her safety. *Journal of Family Violence*, 18, 151-155.
- Almeida, J., Pajtas, P.E., Mahon, B. Z., Nakayama, K., & Caramazza, A. (2013). Affect of the unconscious: Visually suppressed angry faces modulate our decisions. *Cognitive, Affective, & Behavioral Neuroscience*, 13, 94-101.
- Allport, G. W., Clark, K., & Pettigrew, T. (1954). The nature of prejudice.
- Amodio, D. M., Harmon-Jones, E., & Devine, P. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *Journal of Personality and Social Psychology*, 84, 738-753.
- Bargh, J. A. (1994). The Four Horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition*, (pp. 1-40). Hillsdale, NJ: Erlbaum.
- Baranyi, J., & Roberts, T. A. (1994). A dynamic approach to predicting bacterial growth in food. *International journal of food microbiology*, 23, 277-294.
- Baranyi, J., Robinson, T. P., Kaloti, A., & Mackey, B. M. (1995). Predicting growth of *Brochothrix thermosphacta* at changing temperature. *International journal of food microbiology*, 27, 61-75.
- Baty, F., & Delignette-Muller, M. L. (2004). Estimating the bacterial lag time: which model, which precision? *International journal of food microbiology*, 91, 261-277.

- Baumeister, R. F., Smart, L., & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review*, 103, 5-33.
- Bell, K., & Naugle, A. (2005). Understanding stay/leave decisions in violent relationships: A behavior analytic approach. *Behavior and Social Issues*, 14, 21-45.
- Bishop, S. J., Duncan, J., & Lawrence, A. D. (2004). State anxiety modulation of the amygdala response to unattended threat-related stimuli. *Journal of Neuroscience*, 24, 10364-10368.
- Blanchard, D. C., Hynd, A. L., Minke, K. A., Minemoto, T., & Blanchard, R. J. (2001). Human defensive behaviors to threat scenarios show parallels to fear- and anxiety-related defensive patterns of non-human mammals. *Neuroscience and Biobehavioral Reviews*, 25, 761-770.
- Blanchette, I. (2006). Snakes, spiders, guns, and syringes: How specific are evolutionary constraints on the detection of threatening stimuli? *The Quarterly Journal of Experimental Psychology*, 59, 1484-1504.
- Block, S. R., & Liberzon, I. (2016). Attentional processes in posttraumatic stress disorder and the associated changes in neural functioning. *Experimental Neurology*, 284, 153-167.
- Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & Van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, 42, 1-15.

- Bolles, R. C., & Fanselow, M. S. (1980). PDR-a multi-level model of fear and pain. *Behavioral and Brain Sciences*, 3, 315-323.
- Borglin, S., Joyner, D., DeAngelis, K. M., Khudyakov, J., D'haeseleer, P., Joachimiak, M. P., & Hazen, T. (2012). Application of phenotypic microarrays to environmental microbiology. *Current opinion in biotechnology*, 23, 41-48.
- Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1999). Affect and the startle reflex. In M. E. Dawson, A. Schell, & A. Boehmelt (Eds.), *Startle modification: Implications for neuroscience, cognitive science and clinical science*. New York: Cambridge University Press.
- Brewer, M. B. (1988). A dual process model of impression formation. In T.K. Srull & R. S. Wyer (Eds.), *Advances in social cognition*, (pp. 1-36). Hillsdale, NJ: Erlbaum
- Brodeur, M.B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS ONE*, 5, e10773
- Brosch, T., & Sharma, D. (2005). The role of fear-relevant stimuli in visual search: a comparison of phylogenetic and ontogenetic stimuli. *Emotion*, 5, 360-364.
- Buchanon, R. L., & Phillips, J. G. (1990). Response surface model for predicting the effects of temperature pH, sodium chloride content, sodium nitrite concentration and atmosphere on the growth of *Listeria monocytogenes*. *Journal of Food Protection*, 53, 370-376.

- Buchanan, R. L., Whiting, R. C., & Damert, W. C. (1997). When is simple good enough: a comparison of the Gompertz, Baranyi, and three-phase linear models for fitting bacterial growth curves. *Food microbiology*, 14, 313-326.
- Calder, A. J., Lawrence, A.D., & Young, A.W. (2001). Neuropsychology of fear and loathing. *Nature Reviews Neuroscience*, 2, 352-363.
- Canli, T., & Lesch, K. P. (2007). Long story short: the serotonin transporter in emotion regulation and social cognition. *Nature Neuroscience*, 10, 1103-1109.
- Chaiken, S. (1980). Heuristic versus systematic processing and the use of source versus message cues in persuasion. *Journal of Personality & Social Psychology*, 39, 752–766.
- Cheon, B. K., Livingston, R. W., Hong, Y. Y., & Chiao, J. Y. (2014). Gene x environment interaction on intergroup bias: the role of 5-HTTLPR and perceived outgroup threat. *Social Cognitive and Affective Neuroscience*, 9, 1268-1275.
- Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: the quad model of implicit task performance. *Journal of Personality and Social Psychology*, 4, 469-487.
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 6, 1314-1329.
- Costa, T., Cauda, F., Crini, M., Tatu, M.-K., Celeghin, A., de Gelder, B., & Tamietto, M. (2014). Temporal and spatial neural dynamics in the perception of basic emotions

- from complex scenes. *Social Cognitive and Affective Neuroscience*, 9, 1690–1703.
- Cunningham, W. A., & Brosch, T. (2012). Motivational salience: Amygdala tuning from traits, needs, values, and goals. *Current Directions in Psychological Science*, 21, 54-59.
- Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science*, 15, 806-813.
- Cunningham, W. A., Packer, D. J., Kesek, A., & Van Bavel, J. J. (2009). Implicit measurement of attitudes: A physiological approach. In R. E. Petty, R. H. Fazio, & P. Briñol(Eds.), *Insights from the new implicit measures*. (pp. 485-512). New York, NY: Psychology Press.
- Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, 25, 736-760.
- Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annual review of neuroscience*, 15, 353-375.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5-18.
- Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science*, 1, 14-18.

- Dillon, D. G., & LaBar, K. S. (2005). Startle modulation during conscious emotion regulation is arousal-dependent. *Behavioral Neuroscience*, 119, 1118-1124.
- Dimberg, U., & Öhman, A. (1996). Behold the wrath: Psychophysiological responses to facial stimuli. *Motivation and Emotion*, 20, 149-182.
- Donders, N. C., Correll, J., & Wittenbrink, B. (2008). Danger stereotypes predict racially biased attentional allocation. *Journal of Experimental Social Psychology*, 44, 1328-1333.
- Eastwood, J. D., Smilek, D., & Merikle, P. M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion. *Perception & psychophysics*, 63, 1004-1013.
- Ebbinghaus H. (1885/1964). *Memory: A Contribution to Experimental Psychology*. New York, NY: Dover
- Ellwart, T., Rinck, M., & Becker, E. S. (2006). From fear to love: Individual differences in implicit spider associations. *Emotion*, 1, 18-27.
- Esteves, F., Dimberg, U., & Öhman, A. (1994). Automatically elicited fear: Conditioned skin conductance responses to masked facial expressions. *Cognition & Emotion*, 5, 393-413.
- Esteves, F., Parra, C., Dimberg, U., & Öhman, A. (1994). Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. *Psychophysiology*, 31, 375-385.
- Fanselow, M. S. (1994). Neural organization of the defensive behavior system responsible for fear. *Psychonomic Bulletin & Review*, 1, 429-438.

- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology*, 23, 75-109.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013-1027.
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 1, 297-327.
- Fazio, R. H., & Olson, M. A. (2014). The MODE model: Attitude-Behavior Processes as a Function of Motivation and Opportunity. In Sherman, J. W., Gawronski, B., & Trope, Y. (Eds.). *Dual-process theories of the social mind*. New York: Guilford Press.
- Filion, D. L., Dawson, M. E., & Schell, A. M. (1998). The psychological significance of human startle eyeblink modification: a review. *Biological Psychology*, 47, 1-43.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23, pp. 1–74). New York: Academic Press.
- Fox, E., & Damjanovic, L. (2006). The eyes are sufficient to produce a threat superiority effect. *Emotion*, 6, 534-539.
- Fox, E., Griggs, L., & Mouchlianitis, E. (2007). The detection of fear-relevant stimuli: Are guns noticed as quickly as snakes? *Emotion*, 7, 691-696.

- Fox, A. S., Oler, J. A., Tromp, D. P., Fudge, J. L., & Kalin, N. H. (2015). Extending the amygdala in theories of threat processing. *Trends in Neurosciences*, 38, 319-329.
- Freeman, J. B., & Ambady, N. (2011). Hand movements reveal the time-course of shape and pigmentation processing in face categorization. *Psychonomic Bulletin & Review*, 18, 705-712.
- Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, Discrimination, and Racism* (pp. 61–89). Orlando, FL: Academic Press.
- Garrido, M. I., Barnes, G. R., Sahani, M., & Dolan, R. J. (2012). Functional evidence for a dual route to amygdala. *Current Biology*, 22, 129-134.
- Garthright, W. E. (1991). Refinements in the prediction of microbial growth curves. *Food Microbiology*, 8, 239-248.
- Garvert, M. M., Friston, K. J., Dolan, R. J., & Garrido, M. I. (2014). Subcortical amygdala pathways enable rapid face processing. *Neuroimage*, 102, 309-316.
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological Bulletin*, 132, 692-731.
- Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE model. *Social Cognition*, 25, 687-717.

- Gawronski, B., & Bodenhausen, G. V. (2009). Operating principles versus operating conditions in the distinction between associative and propositional processes. *Behavioral and Brain Sciences*, 32, 207-208.
- Gawronski, B., & Bodenhausen, G. V. (2014). The associative–propositional evaluation model: Operating principles and operating conditions of evaluation. In J. W. Sherman, B. Gawronski & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 188–203). New York: Guilford Press
- Gibson, A. M., Bratchell, N., & Roberts, T. A. (1988). Predicting microbial growth: growth responses of salmonellae in a laboratory medium as affected by pH, sodium chloride and storage temperature. *International journal of food microbiology*, 6, 155-178.
- Gortner, E., Berns, S. B., Jacobson, N. S., & Gottman, J. M. (1997). When women leave violent relationships: Dispelling clinical myths. *Psychotherapy: Theory, Research, Practice, Training. Special Issue: Psychotherapy: Violence and the family*, 4, 343-352.
- Gould, S. J., & Vrba, E. S. (1982). Exaptation—a missing term in the science of form. *Paleobiology*, 8, 4-15.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4–27.
- Grillon, C., Pellowski, M., Merikangas, K. R., & Davis, M. (1997). Darkness facilitates the acoustic startle reflex in humans. *Biological Psychiatry*, 42, 453-460.

- Guglielmi, R. S. (1999). Psychophysiological assessment of prejudice: Past research, current status, and future direction. *Personality and Social Psychology Review*, 3, 123-157.
- Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *Neuroreport*, 11, 2351-2354.
- Haselton, M. G., & Buss, D. M. (2000). Error management theory: a new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, 78, 81-91.
- Hayes, S., & Jeffries, S. (2013). Why do they keep going back? Exploring women's discursive experiences of intimate partner abuse. *International Journal of Criminology and Sociology*, 2, 57-71.
- Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in human fear conditioning. *Biological Psychiatry*, 60, 361-368.
- Hermans, D., Vansteenwegen, D., & Eelen, P. (1999). Eye movement registration as a continuous index of attention deployment: Data from a group of spider anxious students. *Cognition & Emotion*, 13, 419-434.
- Hitchcock, J. H., & Davis, M. (1991). Efferent pathway of the amygdala involved in conditioned fear as measured with the fear-potentiated startle paradigm. *Behavioral neuroscience*, 105, 826-842.

- Huijding, J., & De Jong, P. J. (2009). Implicit and explicit attitudes toward spiders: Sensitivity to treatment and predictive value for generalization of treatment effects. *Cognitive Therapy and Research*, 33, 211-220.
- Ito, T. A., Cacioppo, J. T., & Lang, P. J. (1998). Eliciting affect using the International Affective Picture System: Trajectories through evaluative space. *Personality and Social Psychology Bulletin*, 24, 855-879.
- Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, 75, 887-900.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 11, 1254-1259.
- 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.
- Jost, J. T., Gaucher, D., & Stern, C. (2015). “The world isn’t fair”: A system justification perspective on social stratification and inequality. In J. Dovidio & J. Simpson (Eds.), *APA handbook of personality and social psychology* (pp. 317–340). Washington, DC: American Psychological Association.
- Kahn, K. B., & Davies, P. G. (2011). Differentially dangerous? Phenotypic racial stereotypicality increases implicit bias among ingroup and outgroup members. *Group Processes & Intergroup Relations*, 14, 569-580.

- Kveraga, K., Boshyan, J., Adams, R. B., Mote, J., Betz, N., Ward, N., ... & Barrett, L.F. (2015). If it bleeds, it leads: separating threat from mere negativity. *Social Cognitive and Affective Neuroscience*, 10, 28-35.
- Knight, D. C., Waters, N., & Bandettini, P. A. (2009). Neural substrates of explicit and implicit fear memory. *Neuroimage*, 45, 208-214.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review*, 97, 377-395.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8*. University of Florida, Gainesville, FL.
- LeDoux, J. E. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York, NY: Simon and Schuster.
- LeDoux, J. E. (2012). Rethinking the emotional brain. *Neuron*, 73, 653-676.
- LeDoux, J. E. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences*, 111, 2871-2878.
- LeDoux, J. E. (2015). *Anxious: Using the brain to understand and treat fear and anxiety*. Penguin.
- Lewis, P. A., Critchley, H. D., Rotshtein, P., & Dolan, R. J. (2007). Neural correlates of processing valence and arousal in affective words. *Cerebral Cortex*, 17, 742-748.
- Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflexion and reflection: A social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology*, 34, 199-249.

- Lorenz, K. (1965). *Evolution and Modification of Behaviour*. Chicago, IL: University of Chicago Press.
- Löw, A., Lang, P. J., Smith, J. C., & Bradley, M. M. (2008). Both predator and prey: Emotional arousal in threat and reward. *Psychological Science*, 19, 865–873.
- Löw, A., Weymar, M., & Hamm, A. O. (2015). When threat is near, get out of here: dynamics of defensive behavior during freezing and active avoidance. *Psychological Science*, 26, 1706-1716.
- Lundqvist, D., Esteves, F., & Öhman, A. (1999). The face of wrath: Critical features for conveying facial threat. *Cognition & Emotion*, 13, 691-711.
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior research methods*, 47, 1122-1135.
- March, D. S., & Graham, R. (2015). Exploring implicit ingroup and outgroup bias toward Hispanics. *Group Processes & Intergroup Relations*, 18, 89-103.
- March, D. S., Gaertner, L., & Olson, M. A. (2017). In harm's way: On preferential response to threatening stimuli. *Personality and Social Psychology Bulletin*, 43, 1519-1529.
- March, D. S., Gaertner, L., & Olson, M. A. (2018). On the prioritized processing of threat in a Dual Implicit Process Model of evaluation. *Psychological Inquiry*, 19, 1-13.
- McConnell, A. R., & Rydell, R. J. (2014). The systems of evaluation model: A dual-systems approach to attitudes. In J. W. Sherman, B. Gawronski, & Y. Trope

- (Eds.), *Dual Process Theories of the Social Mind* (pp. 204-217). New York: Guilford
- Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the International Affective Picture System. *Behavior Research Methods*, 37, 626-630.
- Mobbs, D., Petrovic, P., Marchant, J. L., Hassabis, D., Weiskopf, N., Seymour, B., ... & Frith, C. D. (2007). When fear is near: Threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science*, 317, 1079-1083.
- Morris, J. S., Öhman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating “unseen” fear. *Proceedings of the National Academy of Sciences*, 96, 1680-1685.
- Murphy, S., & Zajonc, R. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, 5, 723-739.
- Nesse, R. M. (2005). Natural selection and the regulation of defenses: A signal detection analysis of the smoke detector principle. *Evolution and Human Behavior*, 26, 88–105.
- Öhman, A. (2005). The role of the amygdala in human fear: automatic detection of threat. *Psychoneuroendocrinology*, 30, 953-958.
- Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, 3, 466-478.

- Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, 3, 381-396.
- Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review*, 108, 483-522.
- Öhman, A., & Mineka, S. (2003). The malicious serpent: Snakes as a prototypical stimulus for an evolved module of fear. *Current Directions in Psychological Science*, 12, 5-9.
- Olatunji, B. O., Williams, N. L., Tolin, D. F., Abramowitz, J. S., Sawchuk, C. N, Lohr, J. M., & Elwood, L. S. (2007). The Disgust Scale: item analysis, factor structure, and suggestions for refinement. *Psychological assessment*, 19, 281-297.
- Olson, M. A., Crawford, M. T., & Devlin, W. (2009). Evidence for the underestimation of implicit in-group favoritism among low-status groups. *Journal of Experimental Social Psychology*, 45, 1111-1116.
- Olson, M. A., & Fazio, R. H. (2009). Implicit and explicit measures of attitudes: The perspective of the MODE model. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 19 – 63). New York, NY: Psychology Press.
- Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience*, 10, 1095-1102.
- Oosterwijk, S., Lindquist, K. A., Adebayo, M., & Barrett, L. F. (2015). The neural representation of typical and atypical experiences of negative images: comparing

- fear, disgust and morbid fascination. *Social Cognitive and Affective Neuroscience*, 11, 11-22.
- Pascoe, J. P., & Kapp, B. S. (1985). Electrophysiological characteristics of amygdaloid central nucleus neurons during Pavlovian fear conditioning in the rabbit. *Behavioural Brain Research*, 16, 117-133.
- Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81, 181.
- Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: From a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews Neuroscience*, 11, 773-783.
- Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in Western Europe. *European Journal of Social Psychology*, 25, 57–75.
- Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (pp. 123–205). New York, NY: Academic Press.
- Petty, R. E., Tormala, Z. L., Briñol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST model. *Journal of Personality and Social Psychology*, 1, 21-41.
- Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race

- evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, 12, 729-738.
- Pitman, R. K., & Orr, S. P. (1986). Test of the conditioning model of neurosis: Differential aversive conditioning of angry and neutral facial expressions in anxiety disorder patients. *Journal of Abnormal Psychology*, 95, 208-213.
- Pratto, F., & John, O. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, 61, 380-391.
- Prinstein, M. J. (2008). Introduction to the special section on suicide and nonsuicidal self-injury: A review of unique challenges and important directions for self-injury science. *Journal of consulting and clinical psychology*, 76(1), 1.
- Purkis, H. M., & Lipp, O. V. (2007). Automatic attention does not equal automatic fear: Preferential attention without implicit valence. *Emotion*, 2, 314-323.
- Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, 85, 618-660.
- Reagh, Z., & Knight, D. (2013). Negative, but not positive emotional images modulate the startle response independent of conscious awareness. *Emotion*, 4, 782-791.
- Rimé, B., Delfosse, C., & Corsini, S. (2005). Emotional fascination: Responses to viewing pictures of September 11 attacks. *Cognition & Emotion*, 19, 923-932.
- Rinck, M., & Becker, E. S. (2006). Spider fearful individuals attend to threat, then quickly avoid it: Evidence from eye movements. *Journal of Abnormal Psychology*, 115, 231-238.

- Robinson, J. D., & Vrana, S. R. (2000). The time course of emotional and attentional modulation of the startle eyeblink reflex during imagery. *International Journal of Psychophysiology*, 37, 275-289.
- Rosen, J. B., Hitchcock, J. M. Sananes, C. B., Miserendino, M. J., & Davis, M. (1991). A direct projection from the central nucleus of the amygdala to the acoustic startle pathway: Anterograde and retrograde tracing studies. *Behavioral Neuroscience*, 105, 817-825.
- 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.
- Rubenking, B., & Lang, A. (2014). Captivated and grossed out: An examination of processing core and sociomoral disgusts in entertainment media. *Journal of Communication*, 64, 543-565.
- Rusbult, C. E., & Martz, J. M. (1995). Remaining in an abusive relationship: An investment model analysis of nonvoluntary dependence. *Personality & Social Psychology Bulletin*, 6, 558-571.
- 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.
- Schaller, M., & Neuberg, S. L. (2012). Danger, disease, and the nature of prejudice(s). In J. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (pp. 1-55). Burlington, VT: Academic Press.

- Schaller, M., Park, J. H., & Mueller, A. (2003). Fear of the dark: Interactive effects of beliefs about danger and ambient darkness on ethnic stereotypes. *Personality and Social Psychology Bulletin*, 29, 637-649.
- Schutte, N. S., Malouff, J. M., & Doyle, J. S. (1988). The relationship between characteristics of the victim, persuasive techniques of the batterer, and returning to a battering relationship. *Journal of Social Psychology*, 5, 605-610.
- Schofield, T. P., Deckman, T., Garriss, C. P., DeWall, C. N., & Denson, T. F. (2015). Evidence of ingroup bias on the shooter task in a Saudi sample. *SAGE Open*, 5, 1-6.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22.
- Sloman, S. A. (2014). Two systems of reasoning, an update. In Sherman, J., Gawronski, B., & Trope, Y. (Eds.). *Dual-process theories of the social mind*. Guilford Press.
- Smith, E. R. (1993). Social identity and social emotions: Toward new conceptualizations of prejudice. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping* (pp. 297–316). New York: Academic Press
- Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible attitudes influence categorization of multiply categorizable objects. *Journal of Personality and Social Psychology*, 5, 888-898.
- Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613-629.

- Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23–46). Hillsdale, NJ: Erlbaum.
- Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. *Personality and Social Psychology Bulletin*, 10, 1332-1345.
- Straube, T., Pohlack, S., Mentzel, H. J., & Miltner, W. H. (2008). Differential amygdala activation to negative and positive emotional pictures during an indirect task. *Behavioural Brain Research*, 191, 285-288.
- Strube, M. J. (1988). The decision to leave an abusive relationship: empirical evidence and theoretical issues. *Psychological Bulletin*, 104, 236.
- Swinnen, I. A. M., Bernaerts, K., Dens, E. J., Geeraerd, A. H., & Van Impe, J. F. (2004). Predictive modelling of the microbial lag phase: a review. *International Journal of Food Microbiology*, 94, 137-159.
- Teachman, B. A., & Addison, L. M. (2008). Training non-threatening interpretations in spider fear. *Cognitive Therapy and Research*, 3, 448-459.
- Teachman, B. A., & Woody, S. R. (2003). Automatic processing in spider phobia: Implicit fear associations over the course of treatment. *Journal of Abnormal Psychology*, 1, 100-109.
- Turner, S.A. Jr., & Silvia, P.J. (2006). Must interesting things be pleasant? A test of competing appraisal structures. *Emotion*, 6, 670-674.

- Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, 93, 239-257.
- Vanman, E. J., Ryan, J. P., Pedersen, W. C., & Ito, T. A. (2013). Probing prejudice with startle eyeblink modification: A marker of attention, emotion, or both. *International journal of psychological research*, 6, 30-41.
- Van Orden, K. A., Witte, T. K., Cukrowicz, K. C., Braithwaite, S. R., Selby, E. A., & Joiner Jr, T. E. (2010). The interpersonal theory of suicide. *Psychological Review*, 117, 575.
- Vasey, M. W., Harbaugh, C. N., Buffington, A. G., Jones, C. R., & Fazio, R. H. (2012). Predicting return of fear following exposure therapy with an implicit measure of attitudes. *Behaviour Research and Therapy*, 50, 767-774.
- Vrana, S. R., Spence, E. L., & Lang, P. J. (1988). The startle probe response: a new measure of emotion? *Journal of Abnormal Psychology*, 97, 487-491.
- Vuilleumier, P., Armony, J. L., Driver, J., Dolan, R. J. (2001). Effects of attention and emotion on faces processing in the human brain: an event-related fMRI study. *Neuron* 30, 829–841.
- Wallace, P. (2007). How can she still love him? Domestic violence and the Stockholm Syndrome. *Community Practitioner*, 80, 32-34.
- Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. *Handbook of implicit social cognition: Measurement, theory, and applications*, 95-116.

- West, G. L., Anderson, A. A., & Pratt, J. (2009). Motivationally significant stimuli show visual prior entry: Evidence for attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1032-1042.
- Whalen, P. J., Kagan, J., Cook, R. G., Davis, F. C., Kim, H., Polis, S., ... & Johnstone, T. (2004). Human amygdala responsivity to masked fearful eye whites. *Science*, 306, 2061-2061.
- Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *The Journal of Neuroscience*, 18, 411-418.
- White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Anxiety enhances threat processing without competition among multiple inputs: a diffusion model analysis. *Emotion*, 10, 662-677.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107, 101-126.
- Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, 81, 815.
- Yang, E., Zald, D. H., & Blake, R. (2007). Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion*, 7, 882-886.
- Zwietering, M. H., Jongenburger, I., Rombouts, F. M., & Van't Riet, K. (1990). Modeling of the bacterial growth curve. *Appl. Environ. Microbiol.*, 56, 1875-1881.

、

APPENDIX

Stimuli from Studies 1-4

Images indicated by “*” were not used in Study 3 or 4.

Negative Images – 1 = dead/injured animals, 2 = repulsive objects

2



2



2*



1



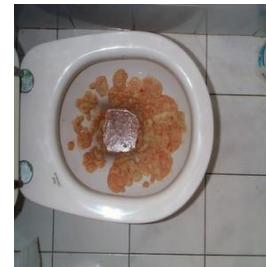
1



1



2



2



1



1



2



2



2*



2



2*



1



1



1



2



2



2



2



2



2*



2



2



2



1



1



1



1



1



1



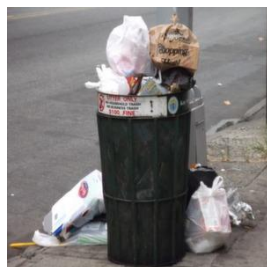
1



1



2



1



2



2*



2



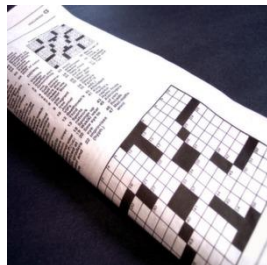
1

Neutral Images.

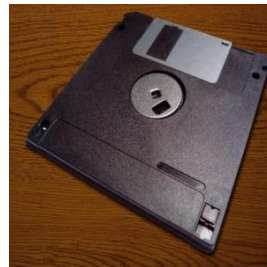
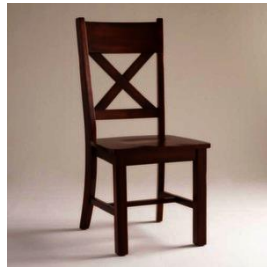
*



*



*



*



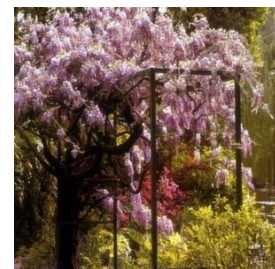


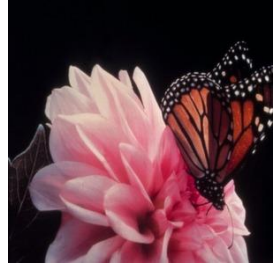
*



Positive Images.

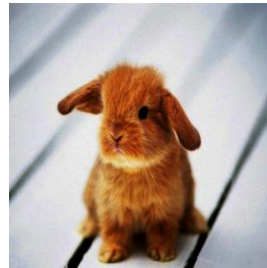
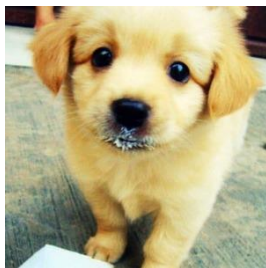
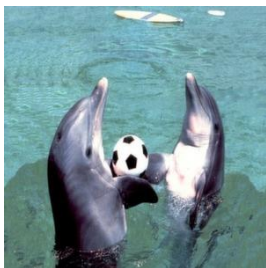
*





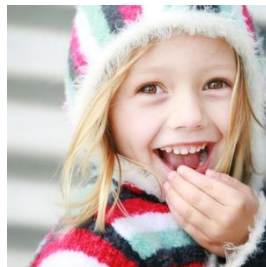
*

*



*

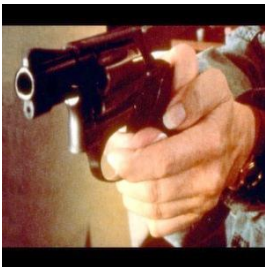
*





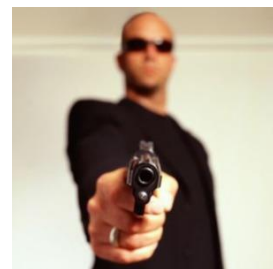
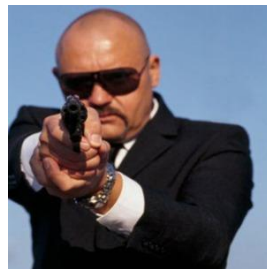
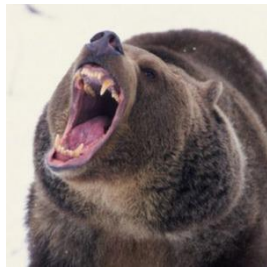
Threatening Images.

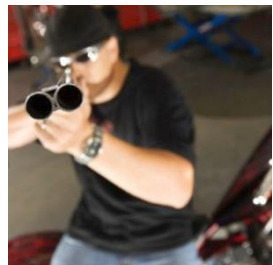
*



*







*

*



Mouse-Tracking Metrics

Mouse-tracker provides metrics that provide information about overall speed and shape of responses including reaction time (RT), area under the curve (AUC), maximum deviation (MD), and maximum deviation time (MD time). All predicted means are presented in Table A.1. All raw mouse-tracking response paths are presented in figure A.1. Patterns mostly mimic those of lag time with some exceptions (those that differ are indicated with an *).

I submitted each metric to separate 3 (race) x 3 (paired-expression) multi-level regressions using PROC MIXED of SAS and Kenwood-Rodgers degrees of freedom. In most instances the model would not converge with random race slopes, and never converged with random race X paired-expression slopes (models that converged with a random race effect are denoted with a #). All models contain a random intercept and random slopes for distractor.

Angry Target Face

Reaction time[#]. When viewing angry faces there was a main effect of race, $F(2,234) = 11.08, p < .0001$, a main effect of distractor, $F(2,334) = 35.01, p < .0001$, and no interaction, $F(4,466) = 1.23, p = .2971$. The race main effect indicates that angry Black faces ($M = 970\text{ms}$) led to quicker identification than did White ($M = 987\text{ms}$), $t(234) = -2.85, p = .0048$, or Asian faces ($M = 997\text{ms}$), $t(234) = 4.67, p < .0001$, and the latter two did not differ, $t(234) = 1.82, p = .0698$. The lack of an interaction indicates that reaction times for angry Black than White or Asian faces did not vary as a function of the distractor. The distractor main effect indicates that happy distractors ($M = 954\text{ms}$) led to

quicker identification of angry faces than did neutral ($M = 975\text{ms}$), $t(233) = -2.40$, $p = .0172$, or sad distractors ($M = 1024\text{ms}$), $t(233) = -8.14$, $p < .0001$, and more quickly to neutral than sad distractors, $t(233) = -5.72$, $p < .0001$.

Area under the curve. When viewing angry faces there was a main effect of race, $F(2,690) = 34.83$, $p < .0001$, a main effect of distractor, $F(2,322) = 5.18$, $p = .0063$, and no interaction, $F(4,690) = .34$, $p = .8533$. The race main effect indicates that angry Black faces ($M = .944$) led to smaller total area of divergence than did White ($M = 1.63$), $t(690) = -6.74$, $p < .0001$, or Asian faces ($M = 1.19$), $t(690) = 7.63$, $p < .0001$, and the latter two did not differ, $t(690) = .90$, $p = .3687$. The lack of an interaction indicates that area under the curve for angry Black than White or Asian faces did not vary as a function of the distractor. The distractor main effect indicates that happy distractors ($M = 1.023$) led to smaller total area of divergence than did sad distractors ($M = 1.184$), $t(222) = -3.21$, $p = .0015$. Area under the curve did not differ to happy vs. neutral distractors ($M = 1.093$) did not differ, $t(222) = -1.39$, $p = .166^*$, or neutral vs. sad distractors, $t(222) = -1.81$, $p = .0711^*$.

Maximum deviation. When viewing angry faces there was a main effect of race, $F(2,688) = 33.01$, $p < .0001$, a main effect of distractor, $F(2,221) = 3.91$, $p = .0214$, and no interaction, $F(4,688) = .25$, $p = .9126$. The race main effect indicates that angry Black faces ($M = .516$) led to smaller deviation from the optimal path than did White ($M = .593$), $t(689) = -6.67$, $p < .0001$, or Asian faces ($M = .601$), $t(688) = 7.35$, $p < .0001$, and the latter two did not differ, $t(689) = .68$, $p = .4949$. The lack of an interaction indicates that maximum deviation for angry Black than White or Asian faces did not vary as a

function of the distractor. The distractor main effect indicates that happy distractors ($M = .544$) led to smaller deviation from the optimal response path than did sad distractors ($M = .596$), $t(221) = -2.80$, $p = .0056$. Maximum deviation did not differ to happy vs. neutral distractors ($M = .571$), $t(222) = -1.46$, $p = .1463$ *, or neutral vs. sad distractors, $t(221) = -1.33$, $p = .1842$ *.

Maximum deviation time[#]. When viewing angry faces there was a main effect of race, $F(2,233) = 12.51$, $p < .0001$, a main effect of distractor, $F(2,233) = 35.09$, $p < .0001$, and no interaction, $F(4,466) = .62$, $p = .6511$. The race main effect indicates that angry Black faces ($M = 488\text{ms}$) led to an earlier time of maximum deviation than did White ($M = 499\text{ms}$), $t(233) = -3.21$, $p = .0015$, or Asian faces ($M = 506\text{ms}$), $t(234) = 4.93$, $p < .0001$, and the latter two did not differ, $t(233) = 1.72$, $p = .0873$. The lack of an interaction indicates that maximum deviation times for angry Black than White or Asian faces did not vary as a function of the distractor. The distractor main effect indicates that happy distractors ($M = 480\text{ms}$) led to an earlier time of maximum than did neutral ($M = 490\text{ms}$), $t(234) = -1.91$, $p = .0580$ *, or sad distractors ($M = 523\text{ms}$), $t(233) = -8.02$, $p < .0001$, and earlier to neutral than sad distractors, $t(233) = -6.10$, $p < .0001$.

Sad Target Face

Reaction time. When viewing sad faces, there was no main effect of race, $F(2,696) = .26$, $p = .7696$, but was a main effect distractor, $F(2,232) = 55.41$, $p < .0001$, and an interaction, $F(4,696) = 6.61$, $p < .0001$. The distractor main effect indicates that happy distractors ($M = 936\text{ms}$) were more quickly identified than were neutral ($M =$

1026ms), $t(232) = -9.63$, $p < .0001$ or angry distractors ($M = 1014$ ms), $t(230) = 8.29$, $p < .0001$, and the latter two did not differ, $t(233) = -1.34$, $p = .182^*$.

The interaction indicates that angry distractors led to *slower* responses when the sad face was Black ($M = 1027$ ms) than when it was Asian ($M = 1003$ ms), $t(695) = -2.65$, $p = .0083$. Angry distractors did not differentially affect the reaction time to sad Black and White ($M = 1012$ ms), $t(696) = 1.68$, $p = .0939^*$, or Asian and White faces, $t(696) = -.96$, $p = .3366$. Happy distractors led to *quicker* responses when the sad face was Black ($M = 915$ ms) than when it was Asian ($M = 942$ ms), $t(695) = 2.95$, $p = .0032$, or White ($M = 915$ ms), $t(695) = -3.90$, $p < .0001$, and the latter two did not differ, $t(695) = -.94$, $p = .3460$. Neutral distractors did not differentially affect the reaction time to identifying sad faces as a function of race, t 's < 1.78 , p 's $> .0747$.

Area under the curve. When viewing sad faces, there was no main effect of race, $F(2,695) = 1.61$, $p = .1999$, but was a main effect of distractor, $F(2,229) = 21.98$, $p < .0001$, and an interaction, $F(4,695) = 11.50$, $p < .0001$. The distractor main effect indicates happy distractors ($M = .9273$) led to a smaller total area of divergence than did angry ($M = 1.118$), $t(229) = 3.41$, $p = .0008$, or neutral distractors ($M = 1.298$), $t(228) = -6.63$, $p < .0014$, and smaller to angry than neutral distractors, $t(229) = -3.22$, $p = .0014$.

The interaction indicates angry distractors led to a *larger* total area of deviance when the sad face was Black ($M = 1.27$) than when it was Asian ($M = 1.02$), $t(695) = -3.99$, $p < .0001$, or White ($M = 1.07$), $t(696) = 3.11$, $p = .0020$, and the latter two did not differ, $t(696) = -.88$, $p = .3814$. Happy distractors led to a *smaller* total area of deviance when the sad face was Black ($M = .741$) than when it was Asian ($M = .996$), $t(695) =$

4.08, $p < .0001$, or White ($M = 1.045$), $t(695) = -4.87$, $p < .0001$, and the latter two did not differ, $t(695) = -.79$, $p = .4290$. Neutral distractors led to a smaller area of total deviance when the sad face was Black ($M = 1.229$) than when it was Asian ($M = 1.358$), $t(696) = 2.05$, $p = .041^*$. Neutral distractors did not lead to a differential area under the curve to sad Black vs. White ($M = 1.308$), $t(696) = -1.26$, $p = .2095$, or Asian vs. White faces $t(696) = .79$, $p = .4277$.

Maximum deviation. When viewing sad faces, there was no main effect of race, $F(2,694) = 2.22$, $p = .1089$, but was a main effect of distractor, $F(2,229) = 23.14$, $p < .0001$, and an interaction, $F(4,694) = 14.23$, $p < .0001$. The distractor main effect indicates that happy distractors ($M = .505$) led to smaller maximum deviation from the optimal path than did angry ($M = .584$), $t(229) = 4.02$, $p < .0001$, or neutral distractors ($M = .638$), $t(228) = -6.76$, $p < .0001$. Angry distractors led to smaller maximum deviation than did neutral distractors, $t(229) = -2.75$, $p = .0064$.

The interaction indicates that angry distractors led to a larger maximum deviation from the optimal path when the sad face was Black ($M = .639$) than when it was Asian ($M = .550$), $t(694) = -4.26$, $p < .0001$, or White ($M = .561$), $t(695) = 3.71$, $p = .0002$, and the latter two did not differ, $t(695) = -.54$, $p = .5878$. Happy distractors led to smaller maximum deviation from the optimal when the sad face was Black ($M = .434$) than when it was Asian ($M = .539$), $t(694) = 5.00$, $p < .0001$, or White ($M = .541$), $t(694) = -5.07$, $p < .0001$, and the latter two did not differ, $t(695) = -.08$, $p = .9402$. Neutral distractors led to smaller maximum deviation from the optimal path when the sad face was Black ($M =$

.608) than when it was Asian ($M = .656$), $t(695) = 2.27$, $p = .0234^*$ or White ($M = .648$), $t(695) = -1.90$, $p = .0573^*$, and the latter two did not differ, $t(694) = .37$, $p = .7122$.

Maximum deviation time. When viewing sad faces, there was no main effect of race, $F(2,696) = 1.06$, $p = .3483$, but was a main effect of paired-expression, $F(2,233) = 51.12$, $p < .0001$, and an interaction, $F(4,695) = 11.50$, $p < .0001$. The distractor main effect indicates that happy distractors ($M = 470\text{ms}$) led to earlier deviation from the optimal path than did angry ($M = 515\text{ms}$), $t(233) = 7.78$, $p < .0001$, or neutral distractors ($M = 525\text{ms}$), $t(232) = -9.49$, $p < .0001$, and the latter two did not differ, $t(229) = -1.72$, $p = .0875^*$.

The interaction indicates that angry distractors led to later deviation from the optimal path when the sad face was Black ($M = 527\text{ms}$) than when it was Asian ($M = 504\text{ms}$), $t(696) = -3.59$, $p = .0004$, or White ($M = 515\text{ms}$), $t(699) = 1.93$, $p = .0538$, and the latter two did not differ, $t(697) = -1.65$, $p = .0994$. Happy distractors led to earlier deviation from the optimal path when the sad face was Black ($M = 459\text{ms}$) than when it was Asian ($M = 471\text{ms}$), $t(696) = 1.93$, $p = .0543$, or White ($M = 480\text{ms}$), $t(696) = -3.43$, $p = .0006$, and the latter two did not differ, $t(696) = -1.51$, $p = .1326$. Neutral distractors led to later deviation from the optimal path when sad the face was Black ($M = 532\text{ms}$) than when it was White ($M = 517\text{ms}$), $t(697) = 2.29$, $p = .0226^*$. Neutral distractors did not differentially affect the maximum deviation time to sad Black vs. Asian ($M = 526\text{ms}$), $t(697) = -.81$, $p = .4172$, or Asian vs. White faces, $t(696) = 1.48$, $p = .1399$.

Happy Target Face

Reaction time. When viewing happy faces, there was a main effect of distractor, $F(2,233) = 6.56, p = .0017^*$, a main effect of race, $F(2,698) = 12.49, p < .0001$, but no interaction, $F(2,698) = 1.09, p = .3583^*$. The race main effect indicates that happy Black faces ($M = 983\text{ms}$) led to longer reaction times than did White ($M = 955\text{ms}$), $t(698) = 4.97, p < .0001$, or Asian faces ($M = 967\text{ms}$), $t(698) = -2.94, p = .0034$, and also longer to happy Asian than White faces, $t(698) = 2.02, p = .0436^*$. The lack of an interaction indicates that reaction times for happy Black than White or Asian faces did not vary as a function of the distractor. The distractor main effect indicates that neutral distractors ($M = 986\text{ms}$) led to longer reaction times to identify happy faces than did angry ($M = 956\text{ms}$), $t(233) = 3.48, p = .0006$, or sad distractors ($M = 963\text{ms}$), $t(234) = -2.94, p = .0034$, and the latter two did not differ, $t(234) = -.88, p = .3807$.

Area under the curve[#]. When viewing happy faces, there was no main effect of distractor, $F(2,225) = .05, p = .9475$, but was a main effect of race, $F(2,225) = 12.93, p < .0001$, and an interaction, $F(4,461) = 2.83, p = .0242$. The race main effect indicates that happy Black faces ($M = 1.191$) led to a larger area of total deviance than did White ($M = .983$), $t(225) = 4.85, p < .0001$, or Asian faces ($M = 1.031$), $t(226) = -3.74, p = .0002$, and the latter two did not differ, $t(226) = 1.11, p = .2672$.

The interaction indicates angry distractors led to larger total area of divergence when the happy face was Black ($M = 1.241$) than when it was White ($M = .927$), $t(665) = 4.64, p < .0001$, or Asian ($M = 1.011$), $t(666) = -3.43, p = .0007$, and the latter two did not differ, $t(666) = 1.20, p = .2290$. Sad distractors led to a larger total area of divergence when the happy face was Black ($M = 1.227$) than when it was White ($M = .993$), $t(666) =$

3.47, $p = .0005$, or Asian ($M = .982$), $t(666) = -3.64$, $p = .0003$, and the latter two did not differ, $t(668) = -.16$, $p = .8711$. Neutral distractors did not differentially affect the area under the curve to happy faces as a function of race, $ts < 1.20$, $ps > .2304$.

Maximum deviation[#]. When viewing happy faces, there was no main effect of distractor, $F(2,230) = .75$, $p = .9297$, but was a main effect of race, $F(2,224) = 22.19$, $p < .0001$, and an interaction, $F(4,464) = 3.51$, $p = .0078$. The race main effect indicates that happy Black faces ($M = .606$) led to larger deviation from the optimal path than did White ($M = .512$), $t(224) = 6.48$, $p < .0001$, or Asian faces ($M = .539$), $t(225) = -4.57$, $p < .0001$, and larger deviations to happy Asian than White faces, $t(225) = 1.90$, $p = .0581^*$.

The interaction indicates that angry distractors led to larger deviation from the optimal path to when the happy face was Black ($M = .631$) than when it was White ($M = .498$), $t(688) = 5.52$, $p < .0001$, or Asian ($M = .535$), $t(689) = -3.98$, $p < .0001$, and the latter two did not differ, $t(689) = 1.53$, $p = .1274$. Sad distractors also led to larger deviation from the optimal path when the happy face was Black ($M = .619$) than when it was White ($M = .508$), $t(689) = 4.62$, $p < .0001$, or Asian ($M = .519$), $t(689) = -4.17$, $p < .0001$, and the latter two did not differ, $t(691) = .46$, $p = .6467$. Neutral distractors did not differentially affect response deviation to happy faces as a function of race, $ts < 1.52$, $ps > .1290$.

Maximum deviation time[#]. When viewing happy faces, there was a main effect of distractor, $F(2,235) = 5.67$, $p = .0039^*$, a main effect of race, $F(2,233) = 14.19$, $p < .0001$, and an interaction, $F(4,465) = 5.48$, $p = .0003$. The distractor main effect indicates that neutral distractors ($M = 495\text{ms}$) led to later deviation from the optimal path

than did angry ($M = 480\text{ms}$), $t(234) = -3.32$, $p = .0011$, or sad distractors ($M = 485\text{ms}$), $t(235) = 2.16$, $p = .0225$, and the latter two did not differ, $t(235) = -1.16$, $p = .2490$. The race main effect indicates that happy Black faces ($M = 495\text{ms}$) led to later deviation from the optimal path than did White ($M = 477\text{ms}$), $t(233) = 5.31$, $p < .0001$, or Asian faces ($M = 488\text{ms}$), $t(234) = -2.30$, $p = .0225$, and later to happy Asian than White faces, $t(234) = 3.02$, $p = .0029^*$.

The interaction indicates that angry distractors led to a later maximum deviation from the optimal path when the happy face was Black ($M = 500\text{ms}$) than when it was White ($M = 464\text{ms}$), $t(698) = 6.23$, $p < .0001$, or Asian ($M = 477\text{ms}$), $t(699) = -4.01$, $p < .0001$, and later when it was Asian than White, $t(699) = 2.21$, $p = .0275$. Neutral distractors led to a later maximum deviation from the optimal path when the happy face was Asian ($M = 502\text{ms}$) than when it was White ($M = 489\text{ms}$), $t(698) = 2.24$, $p = .0251^*$. Neutral distractors did not differentially affect the time of maximum deviation to happy Black ($M = 493\text{ms}$) vs. Asian, $t(699) = 1.51$, $p = .1325$, or Black vs. White faces, $t(699) = .73$, $p = .4638$. Sad distractors led to later deviation from the optimal path when the happy face was Black ($M = 493\text{ms}$) than when it was White ($M = 479\text{ms}$), $t(699) = 2.41$, $p = .0163$. Sad distractors did not differentially affect the maximum deviation time to happy Black vs. Asian ($M = 484\text{ms}$), $t(699) = -1.55$, $p = .1220^*$, or Asian vs. White faces, $t(700) = .86$, $p = .3916$.

Neutral Target Face

Reaction time[#]. When viewing neutral faces, there was no main effect of race, $F(2,229) = .36$, $p = .6979^*$, but was a main effect of distractor, $F(2,233) = 8.36$, $p < .0001$,

and an interaction, $F(2,461) = 5.80, p < .0001$. The distractor main effect indicates that sad distractors ($M = 1025\text{ms}$) led to longer reaction times than did angry ($M = 992\text{ms}$), $t(233) = -3.71, p = .0003$, or happy distractors ($M = 995\text{ms}$), $t(233) = -3.34, p = .0010$, and the latter two did not differ, $t(233) = -.38, p = .7014$.

The interaction indicates that angry distractors led to slower responses when the neutral face was Black ($M = 1003\text{ms}$) than when it was White ($M = 981\text{ms}$), $t(697) = 2.54, p = .0112$. Anger distractors did not differentially affect the reaction time to neutral Black vs. Asian ($M = 990\text{ms}$), $t(697) = -1.53, p = .1266^*$, or Asian vs. White faces, $t(696) = 1.01, p = .3114$. Happy distractors led to quicker responses when the neutral face was Asian ($M = 980\text{ms}$) than when it was Black ($M = 1000\text{ms}$), $t(698) = -2.36, p = .0183^*$, or White ($M = 1005\text{ms}$), $t(698) = -2.99, p = .0029^*$, and the latter two did not differ, $t(697) = -.63, p = .5306$. Sad distractors led to slower responses when the neutral face was Asian ($M = 1037\text{ms}$) then when it was Black ($M = 1015\text{ms}$), $t(699) = 2.55, p = .0110^*$, or White ($M = 1021\text{ms}$), $t(697) = 1.87, p = .0615^*$, and the latter two did not differ, $t(699) = -.67, p = .5006^*$.

Area under the curve. When viewing neutral faces, there was a main effect of race, $F(2,698) = 8.15, p = .0003$, a main effect of distractor, $F(2,233) = 3.38, p = .0356$, and an interaction, $F(4,698) = 6.09, p < .0001$. The race main effect indicates that the total area of divergence was larger when the neutral face was Black ($M = 1.196$) than when it was Asian ($M = 1.083$), $t(699) = -3.55, p = .0004$, or White ($M = 1.087$), $t(698) = 3.44, p = .0006$, and the latter two did not differ, $t(698) = -.12, p = .9074$. The distractor main effect indicates that sad distractors ($M = 1.191$) led to larger total area of divergence than

did happy ($M = 1.073$), $t(233) = -2.49$, $p = .0133$, or angry distractors ($M = 1.102$), $t(233) = -1.89$, $p = .0605$, and the latter two did not differ, $t(233) = .60$, $p = .5482$.

The interaction indicates that anger distractors led to a larger total area of deviance when the neutral face was Black ($M = 1.28$) than when it was Asian ($M = 1.04$), $t(698) = -4.48$, $p < .0001$, or White ($M = .984$), $t(698) = 5.45$, $p < .0001$, and the latter two did not differ, $t(698) = .98$, $p = .3296$. Neither sad nor happy distractors lead to differential areas under the curve as a function of race, $ts < 1.85$, $ps > .0644$.

Maximum deviation. When viewing neutral faces, there was a main effect of distractor, $F(2,233) = 3.73$, $p = .0254$, a main effect of race, $F(2,698) = 10.75$, $p < .0001$, and an interaction, $F(4,698) = 7.57$, $p < .0001$. The race main effect indicates that neutral Black faces ($M = .618$) led to larger deviation from the optimal path than did White ($M = .565$), $t(698) = 4.15$, $p < .0001$, or Asian faces ($M = .568$), $t(699) = -3.87$, $p = .0001$, and the latter two did not differ, $t(698) = .28$, $p = .7832$. The distractor main effect indicates that sad distractors ($M = .607$) led to larger total area of divergence than did happy ($M = .561$), $t(233) = -2.69$, $p = .0076$, or angry distractors ($M = .577$), $t(233) = -1.74$, $p = .0832$, and the latter two did not differ, $t(233) = .95$, $p = .3445$.

The interaction indicates that angry distractors led to larger deviation from the optimal path to when the neutral face was Black ($M = .649$) than when it was White ($M = .532$), $t(698) = 5.94$, $p < .0001$, or Asian ($M = .550$), $t(689) = -4.98$, $p < .0001$, and the latter two did not differ, $t(689) = .96$, $p = .3381$. Sad distractors led to larger deviation from the optimal path when the neutral face was Black ($M = .636$) than when it was White ($M = .582$), $t(700) = 2.76$, $p = .0059$, or Asian ($M = .602$), $t(700) = -1.72$, $p = .0861$,

and the latter two did not differ, $t(698) = 1.04$, $p = .2975$. Happy distractors did not differentially affect response deviation to happy faces as a function of race, $t_s < 1.54$, $p_s > .1246$.

Maximum deviation time[#]. When viewing happy faces, there was a main effect of distractor, $F(2,231) = 10.30$, $p < .0001$, a main effect of race, $F(2,231) = 4.53$, $p = .0118$, and an interaction, $F(4,464) = 5.96$, $p = .0001$. The distractor main effect indicates that sad distractors ($M = 524\text{ms}$) led to later maximum deviation from the optimal path than did angry ($M = 508\text{ms}$), $t(231) = -3.41$, $p = .0008$, or happy distractors ($M = 504\text{ms}$), $t(231) = -4.30$, $p < .0001$, and the latter two did not differ, $t(231) = .88$, $p = .3804$. The race main effect indicates that neutral Black faces ($M = 517\text{ms}$) led to later maximum deviation from the optimal path than did White ($M = 510\text{ms}$), $t(230) = 2.03$, $p = .0438$, or Asian faces ($M = 508\text{ms}$), $t(231) = -2.94$, $p = .0036$, and the latter two did not differ, $t(231) = -.91$, $p = .3627$.

The interaction indicates that angry distractors led to a later maximum deviation from the optimal path when the neutral face was Black ($M = 520\text{ms}$) than when it was White ($M = 500\text{ms}$), $t(694) = 3.67$, $p = .0003$, or Asian ($M = 502\text{ms}$), $t(694) = -3.29$, $p = .0011$, the latter two did not differ, $t(694) = .38$, $p = .7020$. Happy distractors led to an earlier maximum deviation from the optimal path when the neutral face was Asian ($M = 494\text{ms}$) than when it was Black ($M = 504\text{ms}$), $t(695) = -1.98^*$, $p = .0479$, or White ($M = 513\text{ms}$), $t(695) = -3.60$, $p = .0003^*$, and the latter two did not differ, $t(694) = -1.63$, $p = .1042$. Sad distractors did not differentially affect the time of maximum response deviation as a function of race, $t_s < 1.59$, $p_s > .1131^*$.

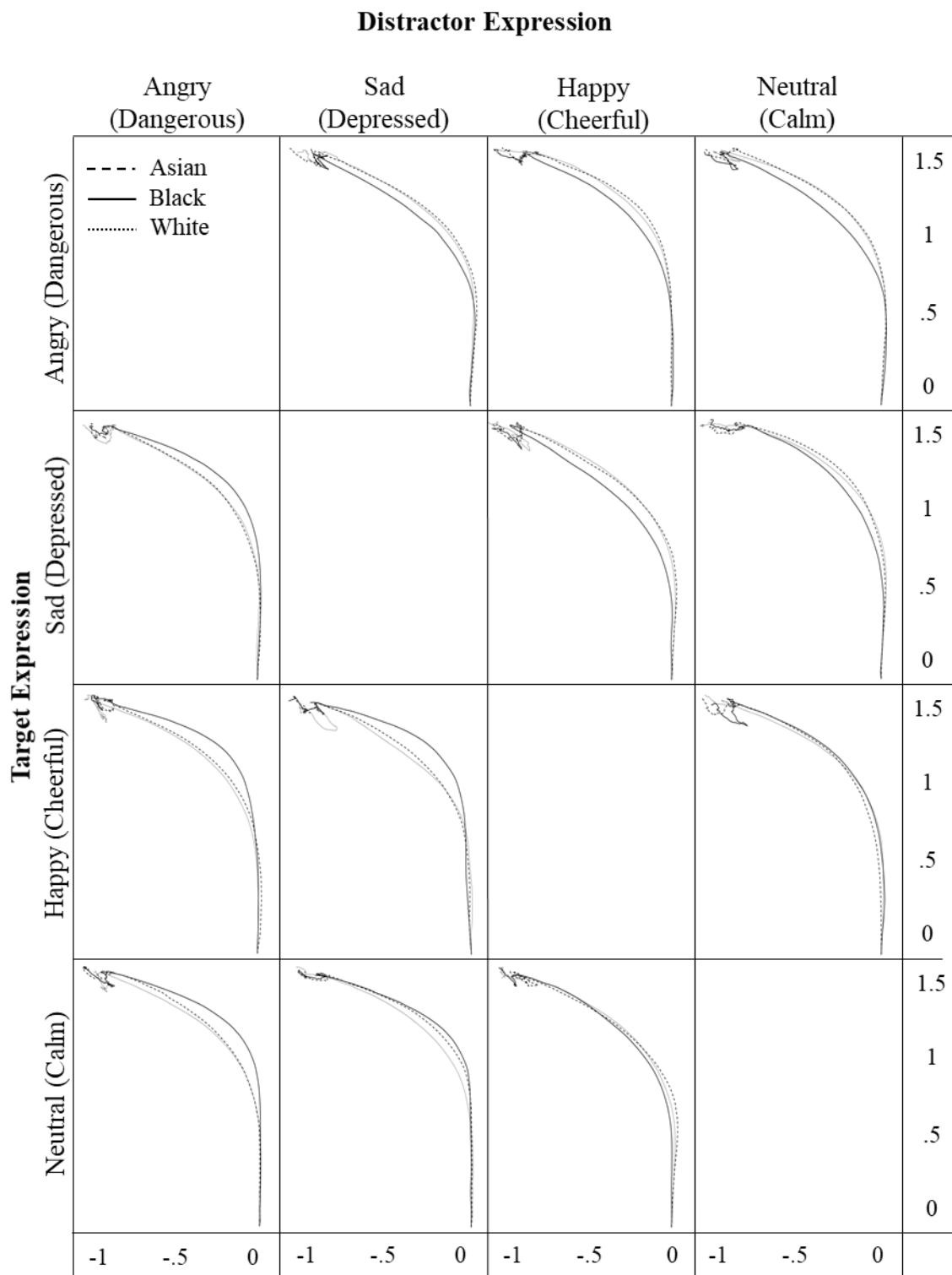


Figure A.1 .Raw mouse-tracking paths. Note: All paths have been horizontally re-mapped to the left. X-axis: 0 = center of screen, 1 = left edge of screen. Y-axis: 0 = bottom of screen, 1.5 = top edge of screen. Asian and White lines often overlap.

Table A.1. Mouse-tracking metrics of each Target x Distractor pairing by race of the face

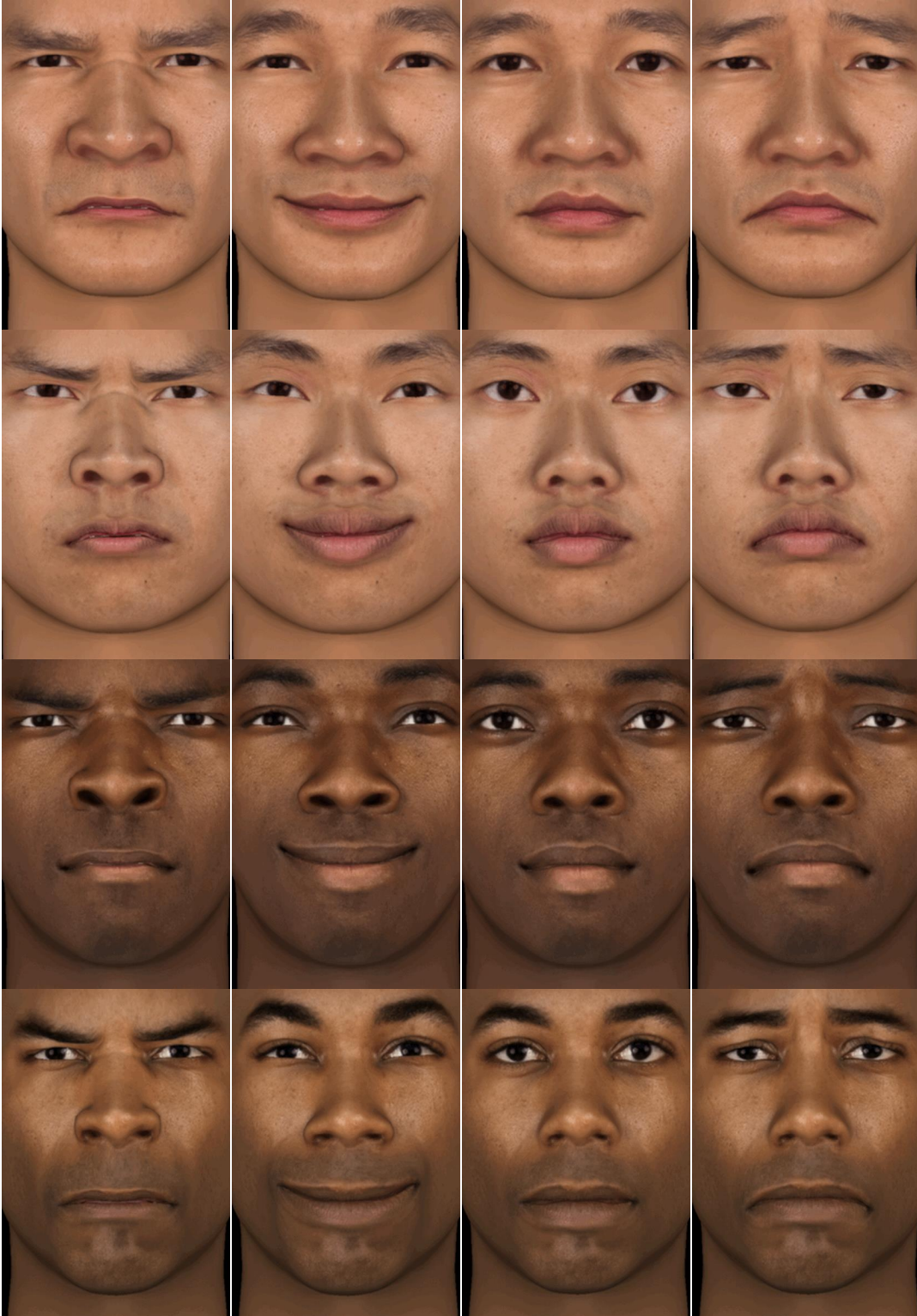
Target Expression		Distractor Expression											
		Angry (Dangerous)			Sad (Depressed)			Happy (Cheerful)			Neutral (Calm)		
		Asian	Black	White	Asian	Black	White	Asian	Black	White	Asian	Black	White
Angry (Dangerous)	Lag Time				574	516	552	511	483	502	530	496	527
	RT				1045	1006	1022	960	944	959	985	961	979
	AUC				1.27	1.02	1.26	1.10	.88	1.09	1.21	.93	1.36
	MD				.63	.54	.62	.57	.49	.57	.61	.51	.59
	MD Time				535	512	523	485	472	484	500	479	493
Sad (Depressed)	Lag Time	519	573	523				498	467	512	577	558	554
	RT	1003	1024	1012				942	915	951	1025	1035	1019
	AUC	1.02	1.27	1.07				1.00	.74	1.05	1.36	1.23	1.31
	MD	.55	.64	.56				.54	.44	.54	.66	.61	.65
	MD Time	504	527	515				471	459	480	527	532	517
Happy (Cheerful)	Lag Time	490	544	483	500	548	501				532	522	510
	RT	950	976	941	958	982	951				992	992	973
	AUC	1.01	1.24	.93	.98	1.23	.99				1.10	1.11	1.03
	MD	.53	.63	.50	.52	.62	.51				.57	.57	.53
	MD Time	477	500	464	484	493	479				502	493	489
Neutral (Calm)	Lag Time	524	568	516	549	573	530	520	521	535			
	RT	990	1003	981	1037	1015	1021	980	1000	1005			
	AUC	1.04	1.28	.98	1.17	1.25	1.15	1.04	1.05	1.13			
	MD	.55	.65	.53	.60	.64	.58	.55	.55	.58			
	MD Time	502	520	500	527	526	518	494	504	513			

Note. All values are estimates of least square means. RT = reaction time in milliseconds; AUC = area under the curve; MD = maximum deviation; MD Time = time of maximum deviation.

Facial Stimuli from Study 5

















VITA

David March was born in Wilmington, DE, to Shelley Spero March and Paul March. He is the second child of five children – brother to Jennifer March Augustine, Shawn March, Douglas March, and Andrew March. He attended Zephyrhills High School and then was a bad undergraduate at Florida State University. He floundered for 5 years in the hippie dive of Austin, TX, where he “worked” as a file clerk. He eventually got it together to get his Master of Science in Psychology at Texas State University in 2010. He then worked as a bad scientist at a media research firm, where he learned how not to do research. David was accepted by philanthropic academics into the Social Psychology program at the University of Tennessee, where he earned a PhD in 2019. David is continuing his academic career as an assistant professor back at Florida State University.