Lions, and Tigers, and Implicit Measures, Oh My!

Implicit Assessment and the Valence vs. Threat Distinction

David S. March
Florida State University

Michael A. Olson & Lowell Gaertner
University of Tennessee

Word Count: 3,449
Corresponding Author Email: march@psy.fsu.edu

Abstract

Physically threatening objects are negative, but negative objects are not necessarily threatening. Moreover, responses elicited by threats to physical harm are distinct from those elicited by other negatively (and positively) valenced stimuli. We discuss the importance of the threat versus valence distinction for implicit measurement both in terms of the activated evaluation and the design of the measure employed to assess that evaluation. We suggest that accounting for the distinct evaluations of threat and valence better enable implicit measures to provide understanding and prediction of subsequent judgement, emotion, and behavior.

We recently argued that the mind uniquely evaluates threatening stimuli relative to other negatively (and positively) valenced stimuli (March, Gaertner, & Olson, 2018a, b). Based on this evaluative difference, we suggest that the distinction between physical threat (the potential to cause injury or death) and valence (the evaluative continuum from negative to positive) is important in the interpretation of indirect measures as they relate to human evaluative responses, and that existing measures differ in their propensity to uncover threat responses versus evaluative responses.

Threat Processing

The threat processing literature suggests that organisms that were faster to detect and react to threats to *immediate bodily harm* were more likely to survive and, consequently, a neural threat-system evolved that advantages the processing of survival threats, both phylogenetically evolved (e.g., snakes, spiders) and ontogenetically learned (e.g., guns, knives), relative to nonthreatening stimuli (Blanchette, 2006; Fox & Damjanovic, 2006; Öhman, Flykt, & Esteves, 2001; cf. Lipp, Derakshan, Waters, & Logies, 2004). That advantage manifests in a faster and stronger perceptual, physiological, and ocular response to threatening stimuli (for a review see March et al., 2018a). The threat we consider is not to self-esteem or happiness but is confined to subjective threats of bodily harm (i.e., stimuli the perceiver construes as immediate dangers to physical safety). An object may be construed as a threat through “prepared” responses (e.g., phylogenesis), which may be relatively “fixed," as well as through learning, which may be context-dependent (e.g., when one learns a close other is prone to violent outbursts, but only when intoxicated). It is only physically threatening stimuli that require quick processing for adaptive responding. We further distinguish the quick evaluation of threat from the “fear” emotion (e.g., LeDoux, 2014), which is a downstream product of both implicit and explicit

processing (Russell & Barrett, 1999). We confine our discussion to the relevance of the automatic evaluation of threat (i.e., assessing whether a stimulus is immediately harmful/deadly) vs. valence (i.e., assessing whether a stimulus is positive or negative) to implicit measurement.

While threatening stimuli are clearly negative, not all negative stimuli are immediately threatening. Our research indicates that the mind preferentially processes threatening stimuli, not simply negatively valent stimuli. For example, relative to nonthreatening negative, positive, and neutral stimuli, threatening stimuli are (a) more quickly detected in an embedded image task, (b) more frequent targets of initial eye-gaze, and (c) stronger elicitors of startle-eyeblink responses (March, Gaertner, & Olson, 2017). The unique processing advantages afforded threatening stimuli would be overlooked without cleaving evaluation into threat and valence dimensions. Extant measures often do not cleave in this manner, which, depending on the aim of the researcher, is often appropriate. However, as we will discuss below, it is often not appropriate to ignore this distinction.

On Matters of Constructs and Construction

The proliferation of implicit measures affords researchers increased flexibility in measurement design, stimuli, response options, timing parameters, and context. Such choices affect measurement outcomes (Fazio & Olson, 2003), including which attitudinal components the measure reveals (Cunningham, Preacher, & Banaji, 2001). For example, in the domain of prejudice and stereotyping, good-bad target judgments in priming tasks reflect overall prejudice, whereas lexical decision (i.e., identification) tasks reflect stereotype content (Wittenbrink et al., 2001; Barden, Maddux, Petty, & Brewer, 2004). Sometimes the broad brushstroke of valence (e.g., good vs. bad) may provide sufficient specificity for the question. However, because of how they are constructed, measures can fail to capture the construct of interest or, perhaps worse,

confound constructs such as threat and valence which have different implications for downstream perception, judgment, and behavior.

Reaction Time Measures

Consider evaluative priming tasks, which rely on semantic/affective congruence between primes and targets to measure their strength of association (Fazio et al., 1986; Fazio, Jackson, Dunton, & Williams, 1995; analogous points can be made about the Implicit Association Test (IAT); Greenwald, Nosek, & Banaji, 2003). Congruent prime-target pairs lead to faster categorization of the target than do incongruent pairings. Yet, such measures often do not include well-differentiated target categories, and as such are agnostic to the nature of the priming effect beyond valence. Imagine a task which involves categorizing "good" (e.g., puppies, rainbows) or "bad" (e.g., guns, dead animals) targets following Black and White face primes. One might expect faster responses on Black-negative and White-positive trials than on incongruent trials (Dovidio, Kawakami, & Gaertner, 2002; Dovidio, Kawakami, Johnson, & Howard, 1997; Fazio, et al., 1995; Ito, Willadsen-Jense, Kaye, & Park, 2011). Such a finding is often interpreted as reflecting valence-based prejudice. Though not incorrect, this conclusion is imprecise because it fails to account for the heterogeneity of the target stimuli; some are merely negative while others are threatening.

To illustrate the importance of this distinction, imagine a person who: (1) was once terrorized by a knife-wielding funhouse clown and now has an automatic threat response whenever they see one; and (2) also has an automatic negative evaluation of mimes absent any deliberate thought. If this person undertook a priming measure with good/bad targets, we would expect both clowns and mimes to facilitate "bad." But, the source of facilitation is distinct: threat for the clown and negative valence for the mime. Such a consideration is usually not considered

in implicit prejudice work. Indeed, by functionally distinguishing threat vs. valence, we suggest that associations with Black individuals could represent dislike and/or a survival threat (March et al., 2018a). Consequently, Black faces might not facilitate responses to all negative targets, and instead might facilitate only negative targets that are relevant to threat or valence (e.g., Donders et al., 2008; Wittenbrink et al., 2001).

Our own work supports this position (March, Gaertner, & Olson, 2020a). Across two priming studies, we found that Black (vs. White) faces facilitated reaction to threatening objects (e.g., predators, masked gunman) but not negative objects (e.g., hurt animals, feces). Furthermore, across two studies using mouse-tracking, we found that (a) Black (vs. White) is more strongly associated with the concept "Dangerous," (b) White (vs. Black) is more associated with the concept "Positive," and (c) neither race is differentially associated with the concept "Negative". All four studies unambiguously indicated that White Americans automatically associate Black individuals with survival threat (an effect that did not extend to another outgroup, i.e., Asians). This suggests that threat, not negative valence, underlies much anti-Black prejudice. Thus, when the aim is to make claims about the processing of threat specifically, it is not sufficient to divide stimuli and responses simply by valence, but to also differentiate negativity from threat.

In addition to matching the stimuli to the process of interest (i.e., threat or valence), consideration of the interpretational latitude of the stimuli themselves is important. A picture, as the saying goes, is worth a thousand words, and might offer more specificity than do words. Black (vs. White) faces, for example, facilitate the identification of negative words (e.g., horrible, terrible) as "bad" (Dovidio et al, 1997; Fazio et al., 1995). Horrible and terrible are both applicable to the description of a cockroach and a murderer, but only a murderer poses an

immediate threat to survival. Here the use of words might obscure the underlying associations that cause Black faces to facilitate the identification of "negative" stimuli.

The need to carefully consider the threat value vs. valence of stimuli is especially apparent when considering sequential-priming measures such as the weapons identification and shooter tasks. Rather than using good/bad responses, these measures use response labels that target the specifically activated construct. In the weapons task, Black and White faces precede images of either weapons or innocuous objects, which participants distinguish using a label to describe the threatening (e.g., "threatening", "gun", "dangerous") or innocuous object (e.g., "safe", "tool", "toy", "non-dangerous"). Participants are quicker to identify threatening versus innocuous objects after Black vs. White primes (Payne, 2001; Thiem, Neel, Simpson, & Todd, 2019; Todd, Thiem, & Neel, 2016; Valla, Bossi, Fox, Ali, & Rivolta, 2018). Similarly, in a shooter task, participants are faster to "shoot" armed Black than White men and slower to "not shoot" unarmed Black than White men (Correll, Park, Judd, & Wittenbrink, 2002; Sadler, Correll, Park, & Judd, 2012). Weapons identification and shooter responses are typically interpreted as reflecting the activation of a stereotype linking Black more than White with guns. Yet, in both measures, one response label clearly connotes threat (e.g., "shoot," "weapon") and one connotes safety (e.g., "don't shoot," "tool"). Though stereotypes indeed link Black with guns, our research would suggest that much of these effects are driven by an underlying association between Black and threat. Often absent from these paradigms are non-threatening negative and non-weapon threatening stimuli that would provide understanding of what Black primes – i.e., the specific concept "gun," a broader evaluation of threat, or an evaluation of negativity.

Indeed, as often constructed, a Black-negative (i.e., prejudice) rather than Black-threat evaluation would produce the same patterns of anti-Black response bias in these tasks. This limitation was the motivation for a study that explored valenced vs. stereotype-based responses to Black vs. White primes (Judd, Blair, & Chapleau, 2004). Finding that Black (vs. White) primes facilitated identification of (a) guns but not insects, and (b) sports but not fruit, Judd et al. suggest there is a specific association between Black and Gun (i.e., stereotype) but not necessarily an evaluative association (i.e., prejudice). What is not clear, however, is what would have occurred had their task required participants to *evaluate* the targets as "Good" or "Bad" rather than semantically *identify* them as "Gun" or "Insect" and "Sports" or "Fruit," which, as Wittenbrink et al (2001) demonstrates, is critical. Indeed, as we previously discussed, our priming work (March et al, 2020a) indicates that Black (vs. White) primes facilitated the evaluation not only of stereotype-congruent threats (e.g., guns) but also of stereotype-incongruent threats (e.g., spiders, bears). Without including multiple instances of threatening objects, we would not have been able to determine whether Black primes facilitated only guns or threats more broadly.

Physiological and Neurological Measures

Physiological and neurological measures index autonomic, central nervous, or musculoskeletal responses without the need for an overt behavioral response (Jackson et al., 1996; Lavine, Thomsen, Zanna, & Borgida, 1998; Stangor, Sullivan, & Ford, 1991). This tact is often seen in research on prejudice, which uses differences in neurophysiological response patterns to reflect distinctions in underlying attitude representations (Amodio, Harmon-Jones, & Devine, 2003). Some of these measures may better distinguish threat from valence.

Consider the startle reflex, which is a behavioral component of the self-protective startle cascade involving both muscular and autonomic responses (Vanman, Paul, Ito, & Miller, 1997). In a typical startle paradigm, people view valenced or neutral primes during a subset of which a loud startling probe occurs. The amplitude of the startle eyeblink is considered a physiological marker of affective state (Vanman et al., 1997) whereby relatively larger startle responses to probes occur during negative primes and relatively smaller responses occur during positive primes (compared to neutral primes; see Bradley, Cuthbert, & Lang, 1999, for a review). The startle reflex has been used to discriminate negative vs. positive responses to different races with research showing that White Americans exhibit larger startle (i.e., negative) responses when primed by Black or Hispanic than White faces (Amodio et al., 2003; March & Graham, 2015; Phelps et al., 2000). Yet, in our own work (March et al., 2017), we presented stimuli that were either neutral, positive, negative, *or* threatening. In an initial study, we found larger startle responses to threatening stimuli than to neutral stimuli, and *reduced* startle to negative stimuli than to neutral stimuli. We have since replicated this finding using subliminally presented stimuli (March, Gaertner, & Olson, 2020b) such that threatening stimuli yielded larger startle than did neutral stimuli and there were no differences in startle to the positive, negative, and neutral stimuli. Considering these patterns then, larger startle eyeblink responses to groups stereotyped as violent or aggressive seem more likely to reflect underlying threat responses and not merely the activation of negative valence. This is reasonable; relatively larger responses would be expected when an organism is in a defensive versus an appetitive state, and defense is a natural response to threat. This is conceptually similar to work focusing on motor components of automatic response to racial groups (Amodio & Devine, 2006; Dovidio, Kawakami, & Gaertner, 2002). Startle may therefore be a useful index of threat responses but not mere negativity.

Furthermore, we replicated the startle effect to subliminally presented threats with another physiological measure of autonomic arousal – skin conductance responses (SCR). Historically, SCRs were believed to occur in response to any highly arousing (positive or negative) stimulus (Rankin & Campbell, 1955). However, our work (March et al., 2020b) observed elevated SCRs only to subliminally presented threatening stimuli, and no differences among SCR to subliminally presented negative, positive, and neutral stimuli. Because we differentiated stimuli in regard to threat and valence, we were able to realize that the SCR response was triggered by threat, not valence (at least with subliminally presented stimuli).

Neuroimaging techniques (e.g., fMRI) are used to explore neural structures involved in prejudice (e.g., the amygdala; Amodio, 2014). Though the amygdala is attuned to the initial processing of threatening information (Cunningham, Packer, Kesek, & van Bavel, 2009), it is also involved in processing affective and motivationally relevant information, and novel (Cunningham & Brosch, 2012) and positive stimuli (Garavan et al., 2001). This begs the question – when the amygdala lights up, is it in response to threat, negativity, or something else? It could be any or all depending on the design and context of the particular study. For example, fMRI research generally finds more amygdala activation in response to Black versus White faces (Cunningham et al., 2004; Phelps et al. 2000). Given the amygdala's role in processing threat, one might expect that such a response reflects a relatively higher threat response to Black versus White faces.

Yet, it is notoriously difficult to interpret brain activations as indicating a direct link between stimuli and threat (e.g., Amodio, 2008). Without considering of all aspects of study design, such ambiguity often leads to careless use of reverse inferenced fallacious associations between activation and cognitive process (Poldrack, 2006). Indeed, the same brain region can be

activated as a consequence of many cognitive processes. Though, similar to our position, Hutzler (2014) suggests that increased functional specificity can be achieved by accounting for the task-setting which may result in more accurate attributions of biological responses to evaluation.

Disentangling Threat and Valence Improves Explanation and Prediction

Threat and valence responses are functionally distinct as they respond to different types of stimuli, have unique purposes, and result in distinct outcomes. This distinction is relevant for interpreting the outcomes of implicit measures used in numerous domains, including those frequently explored by attitude researchers (e.g., prejudice), as well as domains into which attitudes researchers have been more reticent to venture. Many phenomena, like phobias, suicide, and intimate partner violence, arguably involve functionally distinct threat and valence characteristics. But, as we have suggested, methods and measures that obscure the threat vs. valence distinction can produce spurious conclusions.

An illustration of this comes from one of the most documented IAT effects, which is White individuals' relative preference for White over Black names and faces (e.g., Greenwald et al., 1998). More nuanced approaches have provided insight into the nature of the prejudice. For example, in a study employing separate IATs assessing the associations between Black/White with bad/good and threat/nonthreat, respectively, although Black was more strongly associated with both unpleasant and threat, after controlling for the Black/threat association, the Black/bad effect disappeared (Schaller, Park, Mueller, 2003; not reported is whether the threat effect disappears after controlling for valence). This implies that many race IATs utilizing good/bad outcomes may be tapping threat and not simply a negative evaluation. The utility of the typical race IAT may therefore be limited as it measures the negatively valanced outcome of some summary evaluation (i.e., prejudice) while ignoring the possible threat origin of that evaluation.

Again, it may be that all a researcher wants is to measure valanced associations between groups, and in some situations, this is likely sufficient. But we contend that accurately distinguishing the measurement of a threat vs. valence evaluation is important to not only understanding the functional origins of prejudice, but to combatting such prejudice. Indeed, empirical work shows that anti-Black prejudice interventions have little to no lasting change on implicit measurement (Lai et al., 2016). We suggest that such practices are based on a limited understanding of prejudice that assumes that changing the valenced association will lead to less biased outcomes.

The threat vs. valence distinction is also apparent when considering phobia. In a study that employed separate IATs assessing associations between Snakes/Spiders with bad/good and danger/safety, respectively, the association between snakes and spiders persisted even after controlling for the good-bad judgment (Teachman, Gregg, & Woody, 2001). These authors concluded that the "fear-emotive association [i.e., threat] captures individual differences above and beyond the simple effects of negative [i.e., valence] evaluation." (p.231). However, the threat vs. valence distinction is often overlooked in research on phobia reduction techniques. That is, if as we propose, threat and valence are independent components of evaluation (March et al., 2018a), people may become less automatically negative toward a phobic stimulus while maintaining the automatic threat response (i.e., evaluatively, emotionally, and/or physiologically). Indeed, research has shown that phobia treatments may affect the implicitly measured valenced evaluation of the phobic object and perhaps not one's implicitly measured threat response (Teachman & Woody, 2003; cf. Dirikx, Hermans, Vanteenwegen, Baeyens, & Eelen, 2004). By carving up valence into threat- and non-threat-relevant components, this distinction has potential to not only inform the nature of automatic evaluations, but also improve behavioral prediction: implicit measures indexing change in valence may suggest progress in

phobia reduction when minimal underlying change is occurring to the threat evaluation. Hence, spontaneous recovery and reinstatement that seems unpredictable (Hermans, Craske, Mineka, & Lovibond, 2005) may seem more sensible when considering that the threat response remains.

In regard to suicide, both the threat and valence are likely independent antecedents. That is, suicide likely involves self-directed negative valence (i.e., self-hatred) and simultaneously overcoming an association between threat and avoidance to enable self-injurious behavior. Research has shown that repeated exposure to threatening experiences (e.g., combat exposure) can increase the probability of suicidal behaviors, but only among individuals with suicidal ideation (Van Orden et al., 2010). If a diminished threat response occurs independent of a valenced self-evaluation, then indirect measures that differentially assess threat and self-valence could provide insight into suicidal intentions. If implicit measurement of the threat vs. valence evaluation is possible, more efficacious interventions may be developed that target the evolved and socially reinforced associations between threat and avoidance in addition to those that address self-evaluations (e.g., Franklin et al., 2016).

In regard to relationship violence, the threat vs. valence distinction implies the possibility that a victim can simultaneously fear and love the abuser. That is, threat and valence need not conform. Indeed, victims often explicitly endorse both positive views and fear of their abuser (Wallace, 2007), suggesting that both threat and valence are unique and interactive components of the stay/leave decision. This disconnect means that depending on how one implicitly measures a victim's attitudes toward an abusive partner, one may expect to see a positive evaluation, a threat response, or perhaps a mixture (i.e., a summary evaluation) that obscures independent contributions of threat and valence. Previous work has shown that emotional abuse is more predictive of stay/leave decisions than is physical abuse (Gortner, Berns, Jacobson, & Gottman,

1997), suggesting that implicit measures of partner-valance may be better predictors of stay/leave decisions than implicit measures of partner-threat.

As these examples make clear, the threat vs. valence distinction has implications for disparate fields. Disentangling the roles of these arguably distinct processes will provide a more accurate window into not only their underlying nature but also their treatments.

Conclusion

Threatening stimuli elicit unique evaluations relative to other negatively (and positively) valenced stimuli. Yet, the unique influence of threat vs. valence on evaluation is often not considered by work employing implicit measures. We have proposed that researchers may need to consider which measures are best suited to assess threat or valence or something else entirely. Indeed, our threat vs. valence distinction may be but one example of a much-needed discussion about the usefulness of considering more fine-grained distinctions than the typical positive vs. negative distinction. There are many dimensions that go beyond the valence dimension that could be useful for research using implicit measures (e.g., disgust vs. general valence). The threat distinction, however, is particularly relevant to implicit measurement because of its role in very fast processing and reactivity. Through these considerations, indirect measures will provide more accurate insights into and prediction of downstream judgments and behaviors.

References

Allport, G.W. (1954). The historical background of modern social psychology. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 1, pp. 3-56). Cambridge, MA: Addison-Wesley.

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology, 91*, 652-661.

Amodio, D. M. (2008). The social neuroscience of intergroup relations. *European Review of Social Psychology, 19*, 1-54.

Amodio, D. M. (2010). Can neuroscience advance social psychological theory? Social neuroscience for the behavioral social psychologist. *Social Cognition, 28*, 695-716.

Amodio, D. M., Harmon-Jones, E., & Devine, P. G. (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.

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*, 15, 670-682.

Barden, J., Maddux, W., Petty, R., & Brewer, M. (2004) Contextual Moderation of Racial Bias: The Impact of Social Roles on Controlled and Automatically Activated Attitudes. *Journal of Personality and Social Psychology, 87*, 5-22

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.

Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1999). Affect and the startle reflex. *Startle modification: Implications for neuroscience, cognitive science, and clinical science*, 157-183.

Brownstein, M., Madva, A., & Gawronski, B. (2019). What do implicit measures measure? *Cognitive Science*, e1501.

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, 83*, 1314-1329.

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., Packer, D. J., Kesek, A., & Van Bavel, J. (2009). Implicit measurement of attitudes: A physiological approach. *In Attitudes: Insights from the new implicit measures* (pp. 485-512). Psychology Press.

Cunningham, W. A. et al. (2004) Separable neural components in the processing of black and white faces. *Psychological Science, 15*, 806–813.

De Houwer, J. (2003). A structural analysis of indirect measures of attitudes. *The psychology of evaluation: Affective processes in cognition and emotion*, 219-244.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. Learning & Memory, 11, 549-554.

Donders, N.C., Correll, J., & Wittenbrink, B. (2008). Danger stereotypes predict racially biased attentional allocation. *Journal of Experimental Social Psychology, 44*, 1328-1333.

Dovidio, J. F., Gaertner, S. E., Kawakami, K., & Hodson, G. (2002). Why can't we just get along? Interpersonal biases and interracial distrust. *Cultural Diversity and Ethnic Minority Psychology, 8,* 88-102.

Dovidio, J., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). The nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33,* 510–540.

Fazio, R. H. (2007). Attitudes as object–evaluation associations of varying strength. *Social cognition, 25,* 603-637.

Fazio, R. H., & Dunton, B. C. (1997). Categorization by race: The impact of automatic and controlled components of racial prejudice. *Journal of Experimental Social Psychology, 33,* 451-470.

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., Ledbetter, J. E., & Towles-Schwen, T. (2000). On the costs of accessible attitudes: Detecting that the attitude object has changed. *Journal of Personality and Social Psychology, 78,* 197-210.

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.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual review of psychology, 54,* 297-327.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of personality and social psychology, 50,* 229-238.

Fox, E. & Damjanovic, L. (2006). The eyes are sufficient to produce a threat superiority effect. *Emotion, 6,* 534-539.

Franklin, J. C., Fox, K. R., Franklin, C. R., Kleiman, E. M., Ribeiro, J. D., Jaroszewski, A. C., Hooley, J. M., & . Nock, M. K. (2016). A brief mobile app reduces nonsuicidal and suicidal self-injury: Evidence from three randomized controlled trials. *Journal of Consulting and Clinical Psychology, 84,* 544-557.

Gaertner, S.L., & McLaughlin, J.P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly, 46,* 23-30.

Garavan, H., Pendergrass, J. C., Ross, T. J., Stein, E. A., & Risinger, R. C. (2001). Amygdala response to both positively and negatively valenced stimuli. *Neuroreport, 12,* 2779-2783.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 5,* 692-731

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.

Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology, 40,* 357-365.

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of personality and social psychology, 79,* 1022-1038.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in human fear conditioning. *Biological Psychiatry, 60,* 361-368.

Hutzler, F. (2014). Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data. *Neuroimage, 84,* 1061-1069.

Jackson, L., Hodge, C., Gerard, D., Ingram, J., Ervin, K., & Sheppard, L. (1996). Cognition, affect, and behavior in the prediction of group attitudes. *Personality and Social Psychology Bulletin*, *22*, 306–316.

Judd, C. M., Blair, I. V., & Chapleau K. M., (2004). Automatic stereotypes vs. automatic prejudice: Sorting out the possibilities in the Payne (2001) weapon paradigm. *Journal of Experimental Social Psychology, 40*, 75-81.

LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E., & Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: A mixed-trial fMRI study. *Neuron, 20,* 937-945.

Lavine, H., Thomsen, C. J., Zanna, M. P., & Borgida, E. (1998). On the primacy of affect in the determination of attitudes and behavior: The moderating role of affective-cognitive ambivalence. *Journal of Experimental Social Psychology, 34,* 398-421.

Lipp, O. V., Derakshan, N., Waters, A. M., & Logies, S. (2004). Snakes and cats in the flower bed: Fast detection is not specific to pictures of fear-relevant animals. *Emotion, 4*, 233.

Maison, D., Greenwald, A. G., & Bruin, R. H. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behavior. *Journal of consumer psychology*, *14*, 405-415.

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. (2018a). On the prioritized processing of threat in a Dual Implicit Process model of evaluation. *Psychological Inquiry, 29,* 1-13

March, D. S., Gaertner, L., & Olson, M. A. (2018b). Clarifying the explanatory scope of the Dual Implicit Process model. *Psychological Inquiry, 29,* 38-43.

March, D. S., Gaertner, L., & Olson, M. A. (2020a). Distinguishing threat from negative valence as sources of prejudice against Black Americans. Manuscript under review.

March, D. S., Gaertner, L., & Olson, M. A. (2020b). Unique physiological and behavioral responses to nonconsciously processed threatening and non-threatening stimuli. Manuscript in preparation.

March, D. S. & Graham, R. (2015). Exploring implicit ingroup and outgroup bias toward Hispanics. *Group Processes and Intergroup Relations, 18,* 89-103.

Nock, M. K., Park, J. M., Finn, C. T., Deliberto, T. L., Dour, H. J., & Banaji, M. R. (2010). Measuring the suicidal mind: Implicit cognition predicts suicidal behavior. *Psychological science, 21*, 511-517.

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in cognitive sciences, 15*, 152-159.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General, 3*, 466-478.

Olson, M. A., & Fazio, R. H. (2004). Trait inferences as a function of automatically-activated racial attitudes and motivation to control prejudiced reactions. *Basic and Applied Social Psychology, 26,* 1-12.

Olson, M. A., Kendrick, R. V. & Fazio, R. H. (2009). Implicit covariation learning in evaluative vs. non-evaluative dimensions. *Journal of Experimental Social Psychology, 45,* 398-403.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning* (No. 47). University of Illinois press.

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,* 181-192.

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.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10, 59*-63.

Rankin, R. E., & Campbell, D. T. (1955). Galvanic skin response to Negro and White experimenters. *Journal of Abnormal and Social Psychology, 51,* 30-33.

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.

Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and social psychology bulletin*, *26*, 1315-1328.

Russell, J.A., & Barrett, L.F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76,* 805-819.

Sadler, M. S., Correll, J., Park, B., & Judd, C. M. (2012). The world is not black and white: Racial bias in the decision to shoot in a multiethnic context. *Journal of Social Issues, 68,* 286-313.

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

Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible attitudes influence categorization of multiply categorizable objects. *Journal of Personality and Social Psychology, 71*, 888-898.

Stangor, C., Sullivan, L. A., & Ford, T. E. (1991). Affective and cognitive determinants of prejudice. *Social Cognition, 9*, 359-380.

Teachman, B. A., Gregg, A. P., & Woody, S. R. (2001). Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology, 110,* 226-235.

Teachman, B. A., & Woody, S. R. (2003). Automatic processing in spider phobia: Implicit fear associations over the course of treatment. *Journal of Abnormal Psychology*, *112*, 100-109.

Thiem, K. C., Neel, R., Simpson, A. J., & Todd, A. R. (2019). Are Black women and girls associated with danger? Implicit racial bias at the intersection of target age and gender. *Personality and Social Psychology Bulletin, 10, 1427*-1439.

Todd, A. R., Thiem, K. C., & Neel, R. (2016). Does seeing faces of young Black boys facilitate the identification of threatening stimuli? *Psychological Science*, *27*, 384-393.

Valla, L. G., Bossi, F., Calì, R., Fox, V., Ali, S. I., & Rivolta, D. (2018). Not only Whites: Racial priming effect for Black faces in Black people. *Basic and Applied Social Psychology, 40*, 195-200.

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*, 575-600.

Vanman, E. J., Paul, B. Y., Ito, T. A., & Miller, N. (1997). The modern face of prejudice and structural features that moderate the effect of cooperation on affect. *Journal of Personality and Social Psychology, 73*, 941-959.

Wallace, P. (2007). How can she still love him? Domestic violence and the Stockholm Syndrome. *Community Practitioner, 80*, 32-34.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81*, 815-827.

Young, A. I., & Fazio, R. H. (2013). Attitude accessibility as a determinant of object construal and evaluation. *Journal of Experimental Social Psychology*, *49*, 404-418.