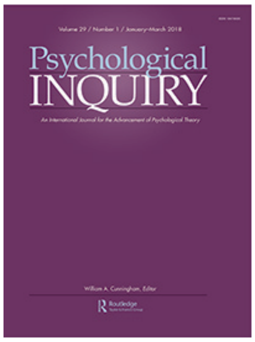

Psychological Inquiry

An International Journal for the Advancement of Psychological Theory

ISSN: 1047-840X (Print) 1532-7965 (Online) Journal homepage: http://www.tandfonline.com/loi/hpli20

Dual and Single-Process Perspectives on the Role of Threat Detection in Evaluation

Benjamin Ruisch, Jeremy Cone, Xi Shen & Melissa J. Ferguson

To cite this article: Benjamin Ruisch, Jeremy Cone, Xi Shen & Melissa J. Ferguson (2018) Dual and Single-Process Perspectives on the Role of Threat Detection in Evaluation, Psychological Inquiry, 29:1, 27-27, DOI: 10.1080/1047840X.2018.1435619

To link to this article: https://doi.org/10.1080/1047840X.2018.1435619

Published online: 21 Mar 2018.

PSYCHOLOGICAL INQUIRY
2018, VOL. 29, NO. 1, 22–27
https://doi.org/10.1080/1047840X.2018.1435619

Routledge
Taylor & Francis Group

COMMENTARIES

Check for updates

Dual and Single-Process Perspectives on the Role of Threat Detection in Evaluation

Benjamin Ruisch[a], Jeremy Cone[b], Xi Shen[a], and Melissa J. Ferguson[a]

[a]Department of Psychology, Cornell University, Ithaca, New York; [b]Department of Psychology, Williams College, Williamstown, Massachusetts

We applaud March, Gaertner, and Olson for their efforts to integrate the literature on threat detection and evaluation—two literatures that have been surprisingly and perhaps unjustifiably disconnected and that we agree could benefit from additional crosstalk. We also thank the authors more generally for their attempts to bring other literatures to bear on existing models of evaluation. We wholeheartedly agree that we as a field would benefit from integrating findings from other areas of research in order to test and refine our existing models of evaluation.

Our goal in this commentary is to outline two broad theoretical issues that we consider to be especially relevant for evaluating and interpreting the dual implicit processes model (DIPM). First, we consider the conceptual distinction between threatening and other negative stimuli. We then discuss the extent to which the current evidence supports the authors' conclusion that threatening and nonthreatening stimuli exhibit *qualitatively* different processing characteristics, or whether these differences might be better characterized as quantitative in nature. We focus particularly on the question of whether it is currently empirically justified to posit the existence of a distinct process responsible for threat detection that is independent of other implicit evaluative processes.

Second, we wish to broaden the theoretical "playing field" beyond the dual-process models reviewed by March and colleagues. Specifically, we consider alternative models that, like the DIPM, propose the existence of multiple, dissociable implicit processes that may operate in the context of implicit evaluation (e.g., Amodio, 2014; Amodio & Ratner, 2011). We also consider single-process models that propose that the dissociations between implicit and explicit evaluation are not explained by a distinction in underlying mental representation, process, or structure (see Cunningham, Zelazo, Packer, & Van Bavel, 2007; De Houwer, 2014; Ferguson, Mann, & Wojnowicz, 2014; Ferguson & Wojnowicz, 2011; Hughes, Barnes-Holmes, & De Houwer, 2011; Wojnowicz, Ferguson, Dale, & Spivey, 2009). Finally, we examine whether single-process models may be able to successfully account for the apparent privileged processing of threat in implicit evaluation, and what these models might suggest about how threat processing may operate, if not via a distinct threat detection process or system.

Is Threat Detection a Distinct Process?

March and colleagues argue that threatening stimuli can be meaningfully distinguished from other kinds of negative stimuli. Whereas all threats are presumably negative, not all negative stimuli are threatening (e.g., their examples of "nonthreatening negative stimuli" include sick or dead animals, rotten teeth, and other cues of sickness or disease). The authors posit a kind of qualitative distinction between threatening and nonthreatening stimuli and argue that stimuli that could potentially cause "immediate bodily harm" are processed via a specialized threat detection system, whereas those that do not pose this kind of threat are not. Yet there is reason to believe that the concept of threat may be less discrete in nature than the authors suggest.

Where Does "Immediate" begin (and How does this Matter for Threat Response)?

In the authors' model, stimuli that pose the threat of *immediate* bodily harm (presumably within a few seconds) are said to be processed via the threat detection system, whereas those that hold the promise of a more delayed impact (e.g., those that may cause sickness or disease) are not. In many cases, this distinction seems to intuitively make sense. For example, a snarling Rottweiler seems to pose a clear threat of immediate bodily harm, whereas a sick person does not. However, although these two particular examples exhibit a clear difference in the "imminence" of the threat that they pose (immediate vs. delayed), it is not clear that the underlying threat that they represent (incapacitation and/or death) is different, or that they would require a qualitatively different threat response.

Specifically, it is not clear to us that there would be an evolutionary advantage to avoiding a dog versus avoiding a sick person. Although the threat posed by the dog is more immediate, in both cases it is necessary to *act* in order to avoid the possibility of getting bitten or sick. To avoid bodily harm, we must sometimes act immediately, even if the negative impact of the threat may itself be delayed. It isn't necessarily clear to us that an evaluative system would be advantaged by drawing a firm boundary between the two. Thus, in our view, these two kinds of negative stimuli—at least in terms of the outcomes they bring about—may be less distinct than the authors suggest.

Further, although the two preceding examples—an aggressive dog and a sick person—differ on a number of important dimensions, the difference between what is threatening and what is merely negative is often simply a matter of degree. For example, although a snarling Rottweiler clearly has the potential to cause immediate bodily harm, a snarling Chihuahua

CONTACT Melissa J. Ferguson mjf44@cornell.edu 211 Uris Hall, Department of Psychology, Cornell University, Ithaca, NY 14853, USA.

likely does not. But at what size, between Chihuahua and Rottweiler, does a dog begin to constitute a threat of immediate bodily harm, activating the threat detection system and thus exerting an influence on implicit evaluations? Shih Tzu? Poodle? Similarly, at what height does a potential fall cross the threshold from nonthreatening to threatening? Two feet? Five feet? Ten? Twenty?

In What Other Ways do Threatening Stimuli Differ from Negative Stimuli?

The authors outline a diverse array of stimuli that they argue meet the threshold for inducing a threat response: animals such as bears, snakes, and spiders; sudden loud noises; members of stereotypically threatening racial groups; weapons; and physically abusive romantic partners, just to name a few. However, it's not immediately clear how outgroup members would pose a qualitatively distinct threat to bodily harm relative to other kinds of evaluatively significant stimuli, nor is it clear, for example, whether there ought to be a qualitative distinction between an abusive romantic partner who might induce bodily harm versus those that are, say, verbally abusive.

It may indeed be the case that the stimuli that past researchers and theorists have categorized as threatening could elicit privileged processing compared with those that have been classified as simply negative. However, we would argue that the current state of the evidence does not convincingly demonstrate that this privileged processing is necessarily due to any qualitative difference between these two classes of stimuli. There is relatively strong evidence—both in the authors' own review and elsewhere in the literature (e.g., Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001)—that negative information *in general* elicits stronger and faster responses than positive or neutral information. However, the authors' more important claim—that threatening information elicits privileged responding *relative to negative information*—has been much less convincingly demonstrated. There are other dimensions on which threatening and negative stimuli are likely to (quantitatively) differ, which have largely not been accounted for in previous research on threat processing. Important to note, some of these factors have also been shown to elicit privileged processing (i.e., stronger and faster responses), raising the possibility that some of these confounding factors may be driving the observed threat advantage.

Extremity. One salient dimension on which threatening and negative stimuli are likely to quantitatively differ is extremity—that is, their degree of positivity or negativity. As an example, in addition to categorizing one as threatening and one as simply negative, we would also expect that the snarling Rottweiler would be seen as *more negative* than the (relatively harmless) snarling Chihuahua. Given that the literature on negativity dominance (e.g., Baumeister et al., 2001; Rozin & Royzman, 2001) has suggested that there is a processing advantage (e.g., stronger and faster responding) for negative stimuli, then it may well be this quantitative difference in the extremity of the stimuli—rather than a qualitative difference between threatening versus not threatening—that drives the privileged processing elicited by the more threatening dog. If extremity is in fact the driving factor behind these observed effects, then it would perhaps suggest that there is little need to posit the existence of a separate threat detection process.

Arousal. Another potentially important factor to take into account relates to the degree of physiological arousal elicited by the stimulus. Other work has shown that more arousing stimuli also elicit privileged processing (e.g., both threatening images *and* erotic images have been shown to elicit stronger responding; Buodo, Sarlo, & Palomba, 2002; Lang, Bradley, & Cuthbert, 1997; Schimmack & Derryberry, 2005). It seems likely that threatening stimuli may, on average, differ from negative stimuli in the degree of arousal that they induce. If the processing advantages afforded to threatening stimuli are driven by systematic differences in arousal, then this, too, may obviate the need to posit the existence of a distinct evaluative process that is specific to threat.

Rarity. In addition, threatening stimuli may generally be less common than nonthreatening stimuli. Other work has shown a processing advantage for novel (vs. familiar) stimuli (e.g., Bradley, Lang, & Cuthbert, 1993). Therefore, to the extent that the frequency of threats versus nonthreats systematically differs in the environment, we should expect there to be differential prioritization of threatening stimuli, with threats commanding more attention or interest *because of their novelty*. Again, this would not suggest a categorical processing advantage for threat, but rather a continuous difference in degree of privileged processing as a function of a stimulus's position on the continuum from common to rare.

To be sure, some of these dimensions may be difficult to disentangle, and it may be the case that some subset of these covariates is inherent to the distinction between negativity and threat; for example, it may be that threatening stimuli *necessarily* induce greater arousal than nonthreatening stimuli. Yet the DIPM must successfully account for these potentially confounding factors if it is to establish that threat is the true driver behind this privileged responding, as well as to determine whether this privileged processing might be better characterized as a continuous—rather than qualitatively discrete—difference.

Is Threat Detection "Privileged" and Does it Necessitate the Existence of a Distinct Process?

Even if we assume that threats do indeed generate stronger and faster responses relative to other classes of stimuli—including other negative evaluative stimuli—another question that still remains for the DIPM is the extent to which this requires a separate threat detection *process* or whether these effects can be successfully accommodated by current models of evaluation. Of key concern is whether the privileged responding that we observe for threat-related stimuli is the result of *qualitatively* (i.e., computationally) different processes. Although dual process models often differ in their specifics and in their proposals for what distinguishes each process (as the authors outline), a common feature is that processes differ in how information is *computed* or *integrated*. A number of dual mode models across cognitive and social psychology propose that implicit processes are based on associative mechanisms that operate as a function of spatiotemporal proximity between stimuli, whereas explicit processes operate on the basis of rule-based propositional mechanisms (for definitions of those terms, see, e.g., Chomsky, 1980, 1986; Hahn & Chater, 1998; Marcus et al., 1995; Pylyshyn, 1980; Searle, 1980; Sloman, 1996; E. E. Smith, Langston, & Nisbett, 1992). Although there is considerable discussion and disagreement about how exactly these processes operate and are implemented, they all are fundamentally concerned with how information from different sources is integrated.

Even the task of establishing a distinction between associative and propositional processes—which would seem to be more computationally disparate phenomena than threat and negative evaluation—has been a matter of debate in the evaluation literature (e.g., De Houwer, 2014; Gawronski & Bodenhausen, 2011, 2014b). Some work from our own lab (Ferguson et al., 2014) has identified potential limitations of attempting to use behavioral data (using measures that differ in their structural properties and other characteristics; Cunningham & Zelazo, 2007; Payne, 2005; Payne, Burkley, & Stokes, 2008; Sloman, 1996)—including, for example, speed and strength of response—to draw such conclusions about underlying computation and operating characteristics. Rather, our own view (as we have articulated elsewhere; Cone, Mann, & Ferguson, 2017; Ferguson et al., 2014) is that computational modeling—which makes highly specific predictions that can then be tested against behavioral data to assess their fit—provides a stronger test of the underlying processing characteristics that drive evaluation, rather than attempting to draw inferences on behavioral data alone. Important to note, there is relatively little evidence in the literature of this kind of computational distinction between threat and negative evaluative processes, and indeed, if the distinction between the two is primarily predicated on the notion that threats are processed more rapidly and elicit stronger responses than other negative stimuli, then we would contend that such evidence would be difficult to recruit. After all, a similar processing advantage exists between positive and negative stimuli (i.e., negativity dominance), and yet this would not necessarily lead us to conclude that negative and positive evaluation are handled by distinct processes. It could, of course, be the case that a computational model that classifies threat and negative evaluation as qualitatively distinct processes could make predictions that differ from standard dual process models; these cases can serve as important tests of the DIPM. However, we would contend that no such evidence currently exists and that models that characterize the distinction between threat and negativity as a simple *quantitative* difference are equally plausible at this point.

Ultimately, it may be helpful to more clearly specify and make more explicit what constitutes a distinct process in the DIPM and in what ways these processes differ from one another with respect to their nature and operation.

Do Threats Activate Distinct Neural Networks and Physiological Outputs?

Another common strategy for differentiating processes in dual mode models is to point to distinct neural substrates (e.g., Cushman, Young, & Greene, 2010; Lieberman, Gaunt, Gilbert, & Trope, 2002; Satpute & Lieberman, 2006; Smith & DeCoster, 2000; Spunt & Lieberman, 2014). March and colleagues (this issue) contend that the neural substrates of threat detection are different from those underlying the detection of nonthreatening negative stimuli and propose that these dissociations implicate the existence of an independent threat detection process—based on LeDoux's (1996, 2012) "low road"—which they contend is thought to respond specifically to a diverse array of immediate threats to bodily harm, including stereotypically threatening outgroup members, loud noises, snakes, and spiders.

Of interest, however, recent work suggests that there are distinct neural circuits involved in processing different forms of threat and that different threats can elicit different patterns of physiological and behavioral response (e.g., Fanselow, 1994; Mobbs et al., 2007). For example, among the simplest threat recognition and response systems are those involved in the human startle reflex. The neural circuitry underlying this reflex is composed of a (relatively) simple and direct neural network and involves a relatively fixed and predictable pattern of defensive response (e.g., blinking, hunching of the shoulders, forward motion of the head; Davis, 1984; Koch, 1999). However, the startle response is triggered *only by a very narrow subset of stimuli* (Davis, 1984; Koch, 1999), such as a rapid burst of noise or a close-range, rapidly moving visual stimulus. Furthermore, the startle system is relatively insensitive to more nuanced *content-related* features of stimuli—and thus is *not* specific to threats of immediate bodily harm (e.g., a quickly moving teddy bear would elicit much the same startle response as a quickly moving bear).

Beyond the startle reflex, there seem to be many forms of more complex threat responses. For example, simple threatening stimuli such as depictions of animals, guns, and other clearly recognizable threats of immediate bodily harm seem to generally exhibit far greater complexity in their processing and behavioral responses than do more reflexive threat responses like the startle response. Work from Fanselow (e.g., Fanselow, 1994; Fanselow & Lester, 1988) and Mobbs (e.g., Mobbs et al., 2007) and their respective colleagues, for instance, has demonstrated that there is no single, monolithic response to threat; rather, there are distinct stages of threat response, each of which are characterized by distinct patterns of neural activation, and each of which elicits a distinct suite of physiological and behavioral responses.

Similarly, other work has demonstrated that other factors related to the specific class, nature, and context of a threat also elicit distinct patterns of threat response. For example, different forms of threat-relevant facial expressions—such as perceiving fearful faces versus angry faces—have been shown to involve distinct forms of neural activation (Pinchon, de Gelder, & Grèzes, 2009), different attentional patterns (Williams, Moss, Bradshaw, & Mattingley, 2005), and different physiological responses (Springer et al., 2007). Other distinct forms of threat that have been shown to elicit different patterns of neural, physiological, and/or behavioral responses include threat conveyed with verbal (vs. visual) information (McMenamin & Marsolek, 2013), exposure to threatening faces versus bodies (Kret, Pichon, Grèzes, & de Gelder, 2011), and threats that are visible versus temporarily out of sight (Rigoli et al., 2016).

It should also be noted that all of the preceding effects have been found in research using relatively simple threat stimuli (e.g., simple line drawings or photographs of snakes, angry faces, etc., such as those used in the work of Öhman, LeDoux, and others). It is unclear whether more complex threat stimuli involve different neural networks or elicit different physiological responses (LeDoux, 2014).

Research suggests that there may be other important distinctions as well. For example, the evidence cited in the authors' review conflates response to *actual* threats like pain and injury (usually from work on mice) with response to *hypothetical* threats like images of weapons and dangerous animals (usually from work on humans). Yet it seems reasonable to suspect that the hypotheticality of threat may matter, with real threat and depictions of threat being processed via different networks and eliciting different patterns of response. More generally, other researchers have shown that threat response in mice and other nonprimates differs in important ways from human threat response, raising questions about whether and to what degree we can generalize from the former to the latter (Pessoa & Adolphs, 2010).

Even if a distinct common neural circuitry for threat processing exists, distinct neural networks do not necessarily indicate differences in evaluative process. After all, computational processes are not a one-to-one mapping onto unique regions of neural circuitry; distinct neural substrates do not necessarily point to differences in the ways in which information is handled computationally. The level of analysis (Marr, 1982) here seems relevant. A consideration of the neural regions involved in threat detection might be seen at the level of implementation, whereas the specifics of how we integrate information (e.g., associative vs. propositional) is understood at a different level (algorithmic). If, like other models of evaluation (cf. Amodio & Ratner, 2011; Cunningham et al., 2007), March and colleagues (this issue) are interested in an algorithmic level of analysis (cf. Amodio & Ratner, 2011; Cunningham et al., 2007), then, on our reading, there is no clear mapping of the (possibly multiple) neural circuitries involved in threat detection to associative versus propositional processes posited in theories of evaluation. More generally, making use of neural dissociations as evidence for distinct processes or systems has some noteworthy limitations (e.g., Dunn & Kirsner, 2003; Plaut, 1995). Increasingly, research highlights the phenomenon of *neural reuse* (see Anderson, 2007, 2010), which refers to the fact that many (cognitive) functions are not tied to any particular neural region but instead depend on dispersed cortical activity. Our reading of this literature is that this research undermines the role of neural specificity in the argument for dual mode approaches.

Dual Mode Versus Single Mode Approaches to Evaluation

In thinking about whether it is necessary to propose a new process for threat detection versus evaluation, it might be useful to take a step back and consider dual mode models of evaluation more generally. Before adding another process to a dual mode approach, we may ask if there is even a convincing need for dual-mode models of evaluation in the first place. If not, and there are viable alternatives, then it might call into question the strategy of explaining threat detection by adding another process.

Just as in other areas of social psychology, dual mode claims in the attitudes literature are typically based on assumptions about information integration in terms of associative versus propositional processing (e.g., see Gawronski & Bodenhausen, 2011, 2014b; McConnell & Rydell, 2014). However, these assumptions are frequently based on dissociations in behavioral data that are collected with structurally different measures. For instance, evaluations that are captured implicitly have been argued to change more slowly in response to new information compared with evaluations that are measured explicitly (for a review, see Cone et al., 2017). This dissociation in learning rates has been used as confirmation that implicit evaluations operate on the basis of slow learning systems or processes whereas explicit evaluations operate according to fast learning systems or processes. The problem is that the difference in measures could explain this and other kinds of dissociations (Ferguson et al., 2014; Payne et al., 2008), without claiming anything about the process in terms of in terms of whether it is associative versus propositional. The nature of the measure can greatly influence and determine the attitude object, the context, demand effects, motivation, and so on, all of which may explain any observed differences in responses.

In addition, no individual given measure is going to be able to capture a single kind of process—associative, propositional, or otherwise. Measures typically are not "process pure", but rather capture multiple ways of combining, weighting, and transposing information (Bishara & Payne, 2009; Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Cunningham & Zelazo, 2007; Payne, 2005; Payne et al., 2008; Sherman, 2006, 2009; Sloman, 1996). This issue adds to the limitations of using only behavioral methods to test claims about process, at least in terms of distinguishing between associative versus propositional.

If the evidence for dual modes is weakened by these varied concerns, what is the alternative? Dual mode approaches (to evaluation and many other topics) have been incredibly generative and productive, both explaining the considerable evidence of dissociation between measures, as well as inspiring new claims and hypotheses (see, e.g., Chaiken & Trope, 1999; Sherman, Gawronski, & Trope, 2014). But given the concerns we have outlined here, we believe that it is useful to consider single (and other) mode approaches in the ongoing research and theory concerning evaluation. A number of theorists have recently articulated exactly these kinds of approaches. For example, in their memory systems model, Amodio (2014) and Amodio and Ratner (2011) advocated for abandoning the idea that implicit cognition consists only of a slow-learning type of associative processing, and instead argued that it reflects multiple types of processes. They defined processes in terms of numerous attributes including learning rate, extinction rate, memory systems, behavioral expression, and neural substrates. De Houwer (2014) argued for a model that assumes only propositional processing for both implicit and explicit evaluation. This approach is in line with newer work suggesting that implicit evaluations can be updated rapidly and durably (e.g., Brannon, Sacchi, & Gawronski 2016; Cone & Ferguson, 2015; Cone et al., 2017; Mann & Ferguson, 2015, 2017; Wyer, 2010, 2016), going against multiple assumptions of implicit cognition as a slow-learning process. Finally, some of us have argued elsewhere for a dynamic model of social cognition (Ferguson & Wojnowicz, 2011; Ferguson et al., 2014; Wojnowicz et al., 2009; see also Cunningham et al., 2007) that assumes that mental processing meets the same set of neurobiologically plausible principles that have been used to describe various brain regions (O'Reilly, 1998). This approach can explain some of the evidence of dissociation (e.g., in evaluation) but does not need to postulate different representational formats (although they can certainly differ) or a distinction between rule-based versus associative processing to do so. In our view, this set of approaches—though they of course have their own explanatory challenges—are an important addition to contemporary theories of evaluation. They also, it is important to note, raise the possibility that just as we might be able to explain explicit and implicit evaluation without invoking two computationally distinct processes (or systems), we similarly might be able to explain threat detection without posting a new type of process.

Conclusion

We reiterate our thanks to March and colleagues for their efforts in bridging the literatures of threat detection and evaluation. We have considered two main issues with their proposal, one concerning the strength of the evidence that threat detection is empirically separable (in a categorical way) from evaluation, and the other concerning whether evaluation more generally needs to be explained with two processes (or systems). Although we identified some alternative explanations for the results reviewed by March and colleagues, and outlined some other theoretical approaches to thinking about process, we appreciate the discussion about these issues and look forward to the empirical work that will undoubtedly provide more constraints and correctives for theory building.

References

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*, *15*(10), 670–682. doi:10.1038/nrn3800

Amodio, D. M., & Ratner, K. G. (2011). A memory systems model of implicit social cognition. *Current Directions in Psychological Science*, *20* (3), 143–148. doi:10.1177/0963721411408562

Anderson, M. L. (2007). Evolution of cognitive function via redeployment of brain areas. *The Neuroscientist*, *13*(1), 13–21. doi:10.1177/1073858406294706

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, *33*(4), 245–266. doi:10.1017/S0140525X10000853

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. doi:10.1037/1089-2680.5.4.323

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, *1*(3), 276–298. doi:10.1037/1528-3542.1.3.276

Bradley, M. M., Lang, P. J., & Cuthbert, B. N. (1993). Emotion, novelty, and the startle reflex: Habituation in humans. *Behavioral Neuroscience*, *107*(6), 970. doi:10.1037/0735-7044.107.6.970

Bradley, M. M., Sabatinelli, D., Lang, P. J., Fitzsimmons, J. R., King, W., & Desai, P. (2003). Activation of the visual cortex in motivated attention. Behavioral neuroscience, *117*(2), 369–380.

Brannon, S. M., Sacchi, D. L., & Gawronski, B. (2017). (In) consistency in the eye of the beholder: The roles of warmth, competence, and valence in lay perceptions of inconsistency. *Journal of Experimental Social Psychology*, 70, 80–94.

Buodo, G., Sarlo, M., & Palomba, D. (2002). Attentional resources measured by reaction times highlight differences within pleasant and unpleasant, high arousing stimuli. *Motivation and Emotion*, *26*(2), 123–138. doi:10.1023/A:1019886501965

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. Guilford Press.

Chomsky, N. (1980). Rules and representations. *Behavioral and brain sciences*, *3*(1), 1–15. doi:10.1017/S0140525X00001515

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Greenwood Publishing Group.

Cone, J., & Ferguson, M. J. (2015). He did what? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology*, *108*(1), 37. doi:10.1037/pspa0000014

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. In *Advances in Experimental Social Psychology* (Vol. 56, pp. 131–199). Academic Press.

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, 89*(4), 469–487.

Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends in Cognitive Sciences, 11*(3), 97–104.

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. doi:10.1521/soco.2007.25.5.736

Cushman, F., Young, L., & Greene, J. D. (2010). Our multi-system moral psychology: Towards a consensus view. *The Oxford handbook of moral psychology*, 47–71.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report. *Biological Psychology*, *52* (2), 95–111. doi:10.1016/S0301-0511(99)00044-7

Davis, M. (1984). The mammalian startle response. In *Neural mechanisms of startle behavior* (pp. 287–351). New York, NY: Springer

De Houwer, J. (2014). Why a propositional single-process model of associative learning deserves to be defended. In J. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual processes in social psychology* (pp. 530–541). New York, NY: Guilford.

Dunn, J. C., & Kirsner, K. (2003). What Can We Infer from Double Dissociations? *Cortex*, *39*(1), 1–7.

Fanselow, M. S. (1994). Neural organization of the defensive behavior system responsible for fear. *Psychonomic Bulletin & Review*, *1*(4), 429–438. doi:10.3758/BF03210947

Fanselow, M. S., & Lester, L. S. (1988). A functional behavioristic approach to aversively motivated behavior: Predatory imminence as a determinant of the topography of defensive behavior. In R. C. Bolles & M. D. Beecher (Eds.), *Evolution and learning* (pp. 185–212). Hillsdale, NJ: Erlbaum.

Ferguson, M. J., Mann, T. C., & Wojnowicz, M. (2014). Rethinking duality: Criticisms and ways forward. In J. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 578–594). New York, NY: Guilford Press.

Ferguson, M. J., & Wojnowicz, M. (2011). The when and how of evaluative readiness: A social cognitive neuroscience perspective. *Personality and Social Psychology Compass*, *5*, 1018–1038. doi:10.1111/j.1751-9004.2011.00393.x

Gawronski, B., & Bodenhausen, G. V. (2009). Operating principles versus operating conditions in the distinction between associative and propositional processes. *Behavioral and Brain Sciences*, *32*(2), 207–208. doi:10.1017/S0140525X09000958

Gawronski, B., & Bodenhausen, G. V. (2011). The Associative–Propositional Evaluation Model: Theory, Evidence, and Open Questions. In J. M. Olson & M. P. Zanna (Eds.), Advances in Experimental Social Psychology (Vol. 44, pp. 59–127). Academic Press. "https://doi.org/10.1016/B978-0-12-385522-0.00002-0"

Gawronski, B., & Bodenhausen, G. V. (2014). Implicit and explicit evaluation: A brief review of the associative-propositional evaluation model.

Social and Personality Psychology Compass, *8*(8), 448–462. doi:10.1111/spc3.12124

Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? exhaustive? empirically distinguishable?. *Cognition*, *65*(2), 197–230. doi:10.1016/S0010-0277(97)00044-9

Harris, C. R., & Pashler, H. (2004). Attention and the processing of emotional words and names: Not so special after all. *Psychological Science*, *15*(3), 171–178. doi:10.1111/j.0956-7976.2004.01503005.x

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorizing in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record*, *61*(3), 465–496. doi:10.1007/BF03395772

Koch, M. (1999). The neurobiology of startle. *Progress in Neurobiology*, *59* (2), 107–128. doi:10.1016/S0301-0082(98)00098-7

Kret, M. E., Pichon, S., Grèzes, J., & de Gelder, B. (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *Neuroimage*, *54*(2), 1755–1762. doi:10.1016/j.neuroimage.2010.08.012

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. In P. J. Lang, R. F. Simons, & M. T. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 97–135). Mahwah, NJ: Erlbaum.

LeDoux, J. (2012). Rethinking the Emotional Brain. *Neuron, 73*(4), 653–676.

LeDoux J.E. (1996). The Emotional Brain. New York: Simon & Schuster.

LeDoux, J. E. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences, 111*(8), 2871–2878.

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. doi:10.1016/S0065-2601(02)80006-5

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology*, *108*(6), 823. doi:10.1037/pspa0000021

Mann, T., & Ferguson, M. J. (2017). Reversing implicit first impressions through reinterpretation after a two-day delay. *Journal of Experimental Social Psychology*, *68*, 122–127. doi:10.1016/j.jesp.2016.06.004

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: Freeman.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. Cognitive psychology, *Cognitive psychology, 29(3)*, 189–256

McConnell, A. R., & Rydell, R. J. (2014). The systems of evaluation model. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 204–217). New York, NY: Guilford.

McMenamin, B. W., & Marsolek, C. J. (2013). Can theories of visual representation help to explain asymmetries in amygdala function?. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(2), 211–224. doi:10.3758/s13415-012-0139-1

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* (5841), 1079–1083. doi:10.1126/science.1144298

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*(1), 5–9.

O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends in cognitive sciences, 2*(11), 455–462.

Payne, B. K. (2005). Conceptualizing control in social cognition: how executive functioning modulates the expression of automatic stereotyping. *Journal of personality and social psychology, 89*(4), 488–503.

Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology, 20*(1), 272–314.

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, *94*(1), 16. doi:10.1037/0022-3514.94.1.16

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*(11), 773–783.

Pichon, S., de Gelder, B., & Grèzes, J. (2009). Two different faces of threat. Comparing the neural systems for recognizing fear and anger in dynamic body expressions. *Neuroimage*, *47*(4), 1873–1883. doi:10.1016/j.neuroimage.2009.03.084

Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 17*(2), 291–321.

Pylyshyn, Z. W. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, *3*(1), 111–132. doi:10.1017/S0140525X00002053

Rigoli, F., Ewbank, M., Dalgleish, T., & Calder, A. (2016). Threat visibility modulates the defensive brain circuit underlying fear and anxiety. *Neuroscience Letters, 612*, 7–13.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. doi:10.1207/S15327957PSPR0504_2

Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, *1079*(1), 86–97. doi:10.1016/j.brainres.2006.01.005

Schimmack, U., & Derryberry, D. E. (2005). Attentional interference effects of emotional pictures: Threat, negativity, or arousal? *Emotion*, *5*(1), 55. doi:10.1037/1528-3542.5.1.55

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. doi:10.1017/S0140525X00005756

Sherman, J. W. (2006). On building a better process model: It's not only how many, but which ones and by which means?. *Psychological Inquiry*, *17*(3), 173–184.

Sherman, J. W. (2009). Controlled influences on implicit measures: Confronting the myth of process-purity and taming the cognitive monster. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.),Attitudes: Insights from the new wave of implicit measures (pp. 391–426). Hillsdale, NJ: Erlbaum.

Sherman, J. W., Gawronski, B., & Trope, Y. (Eds.). (2014). *Dual-process theories of the social mind*. Guilford Publications.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3. doi:10.1037/0033-2909.119.1.3

Smith, E. E., Langston, C., & Nisbett, R. E. (1992). The case for rules in reasoning. *Cognitive Science*, *16*, 1–40. doi:10.1207/s15516709cog1601_1

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*, 108–131. doi:10.1207/S15327957PSPR0402_01

Springer, U. S., Rosas, A., McGetrick, J., & Bowers, D. (2007). Differences in startle reactivity during the perception of angry and fearful faces. *Emotion, 7*(3), 516.

Spunt, R. P., & Lieberman, M. D. (2014). Automaticity, control, and the social brain. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 279–298). New York, NY: Guilford Press.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247. doi:10.1207/s15327957pspr0803_1

Williams, M., Moss, S., Bradshaw, J., & Mattingley, J. (2005). Look at me, I'm smiling: Visual search for threatening and nonthreatening facial expressions. *Visual Cognition*, *12*, 29–50. doi:10.1080/13506280444000193

Wojnowicz, M., Ferguson, M. J., Dale, R., & Spivey, M. (2009). The self-organization of deliberate evaluations. *Psychological Science*, *20*, 1428–1435. doi:10.1111/j.1467-9280.2009.02448.x

Wyer, N. A. (2010). You never get a second chance to make a first (implicit) impression: The role of elaboration in the formation and revision of implicit impressions. *Social Cognition*, *28*(1), 1–19.

Wyer, N. A. (2016). Easier done than undone… by some of the people, some of the time: The role of elaboration in explicit and implicit group preferences. *Journal of Experimental Social Psychology*, *63*, 77–85.