

Swarthmore College

Works

Psychology Faculty Works

Psychology

2021

The Negativity Bias, Revisited: Evidence From Neuroscience Measures And An Individual Differences Approach

Catherine Norris

Swarthmore College, cnorris2@swarthmore.edu

Follow this and additional works at: <https://works.swarthmore.edu/fac-psychology>



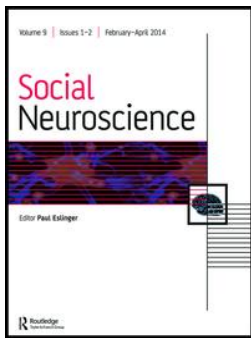
Part of the [Psychology Commons](#)

Let us know how access to these works benefits you

Recommended Citation

Catherine Norris. (2021). "The Negativity Bias, Revisited: Evidence From Neuroscience Measures And An Individual Differences Approach". *Social Neuroscience*. Volume 16, DOI: 10.1080/17470919.2019.1696225
<https://works.swarthmore.edu/fac-psychology/1125>

This work is brought to you for free by Swarthmore College Libraries' Works. It has been accepted for inclusion in Psychology Faculty Works by an authorized administrator of Works. For more information, please contact myworks@swarthmore.edu.



The Negativity Bias, Revisited: Evidence from Neuroscience Measures and an Individual Differences Approach

Catherine J. Norris

To cite this article: Catherine J. Norris (2019): The Negativity Bias, Revisited: Evidence from Neuroscience Measures and an Individual Differences Approach, Social Neuroscience, DOI: [10.1080/17470919.2019.1696225](https://doi.org/10.1080/17470919.2019.1696225)

To link to this article: <https://doi.org/10.1080/17470919.2019.1696225>



Accepted author version posted online: 21 Nov 2019.



Submit your article to this journal [↗](#)



Article views: 23



View related articles [↗](#)



View Crossmark data [↗](#)

Publisher: Taylor & Francis & Informa UK Limited, trading as Taylor & Francis Group

Journal: *Social Neuroscience*

DOI: 10.1080/17470919.2019.1696225

Running Header: NEGATIVITY BIAS, REVISITED

The Negativity Bias, Revisited:

Evidence from Neuroscience Measures and an Individual Differences Approach

Catherine J. Norris, Ph.D.

Swarthmore College

DATE: 11/20/2019 10:56:00 AM

WORD COUNT: 10770

Address correspondence to: Catherine J. Norris, Ph.D., Associate Professor, Department of Psychology, Swarthmore College, 500 College Avenue, Swarthmore PA 19081, cnorris2@swarthmore.edu.

Abstract

Past research has provided support for the existence of a negativity bias, the tendency for negativity to have a stronger impact than positivity. Theoretically, the negativity bias provides an evolutionary advantage, as it is more critical for survival to avoid a harmful stimulus than to pursue a potentially helpful one. The current paper reviews the theoretical grounding of the negativity bias in the Evaluative Space Model, and presents recent findings using a multilevel approach that further elucidate the mechanisms underlying the negativity bias and underscore the importance of the negativity bias for human functioning.

Keywords: ERPs, fMRI, neuroticism, personality, gender, age

The Negativity Bias, Revisited:

Evidence from Neuroscience Measures and an Individual Differences Approach

When all other factors are held constant, bad is stronger than good. Negative information has stronger effects on attention, perception, memory, physiology, affect, behavior, motivation, and decision-making than does equally extreme and arousing positive information. This effect has been termed the *negativity bias*, and is intuitive in folk psychology; most individuals acknowledge that negativity has the power to wield a stronger influence on our behavior and our survival than does positivity. Furthermore, empirical evidence supporting the negativity bias has been accumulated from a wide variety of paradigms and research fields, and summarized in multiple review articles published 20-30 years ago (Taylor, 1991; Rozin & Royzman, 2001; Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). Although each of these reviews outlined theoretical perspectives on the negativity bias, there existed no comprehensive model of affect or emotion that accounted for most (if not all) manifestations of the negativity bias until Cacioppo and Berntson (1994; Cacioppo, Gardner & Berntson, 1997, 1999) proposed the Model of Evaluative Space (ESM). The current paper first summarizes the principles of the ESM that give rise to the negativity bias, and then turns to recent research supporting the functioning of a negativity bias using multiple measures, drawing from different research domains, and focusing on (a) evidence for a negativity bias from the field of social neuroscience, and (b) individual differences in the negativity bias, as well as the effects of individual differences (e.g., age, gender) on the negativity bias.

Two of John Cacioppo's legacies – regardless of topic area – are the implications of his work for broad fields of research, including social psychology (and its many subfields), psychoneuroimmunology, psychophysiology, neuroscience, clinical psychology, and many others; and his multilevel approach to studying and understanding these wide-ranging questions about human nature. Research on the negativity bias is illustrative of both of these legacies. First, it has far-reaching consequences spanning from marketing and advertising to treatment of mental health disorders. Second, the integration of multiple methodologies and levels of inquiry allows for the investigation of mechanisms and the possibility for the development of interventions or other applications.

Some of the initial evidence for a negativity bias came from the burgeoning field of behavioral economics, in which Kahneman and Tversky (1979; Tversky & Kahneman, 1991) showed that whether a choice was framed in terms of losses or gains had an impact on decisions, and that the value of a loss and a gain of equal amounts is different. This work is integral for understanding, for example, the efficacy of advertisements, movie reviews, and information about health behaviors, and was strongly influential for some of John's earlier work on the ESM and the negativity bias in particular. For example, Cacioppo and Gardner (1993) examined (separable) positive and negative attitudes toward blood, bone-marrow, and organ donation, and found that "Negative beliefs and fears may constitute a particularly difficult obstacle to inducing donor behaviors (p.269)." Much of John's earlier thinking about the negativity bias was influenced by research in the realm of attitudes and persuasion.

John's own work on the negativity bias has also influenced perspectives on affective processing in clinical disorders. In collaboration with Jackie Gollan and her colleagues, we

have published a series of papers examining differences in the negativity bias as a function of clinical diagnosis, and the effects of treatment (i.e., Behavioral Activation therapy, most strongly) on biases in affective processing (Gollan et al., 2014, 2016a, 2016b; Rosebrock et al., 2016). Taken together, these studies demonstrate the utility of applying a theory about affective processing to understanding differences in those processes in healthy vs. clinically depressed and anxious samples to, ultimately, effects of treatment on regulating those processes.

Furthermore, the incorporation of physiological and neural measures to research on the negativity bias has allowed for investigations into its underlying mechanisms. Both ERP and fMRI data have suggested that the negativity bias emerges relatively early in processing, driving attention towards unpleasant vs. pleasant stimuli (Smith, Cacioppo, Larsen, & Chartrand; 2003; Norris & Cacioppo, 2019), but have also demonstrated that this bias is maintained in later stages (e.g., the late positive potential of the ERP, associated with affective categorization; Ito et al., 1998), and that it may even affect recovery from unpleasant stimuli in individuals higher in neuroticism (Norris, Larsen & Cacioppo, 2007). By taking a multimethod, multilevel approach to studying the negativity bias, John's work has opened doors to understanding the instantiation and consequences of a bias towards unpleasant and aversive stimuli.

In sum, John's theoretical and empirical work on the negativity bias is illustrative of his contributions to multiple fields and has far-reaching implications for understanding human behavior.

The Evaluative Space Model

Based on their review of research on attitudes and evaluative processes (i.e., the “operations by which organisms discriminate threatening from nurturant environments”), Cacioppo and Berntson (1994; Cacioppo, Gardner & Berntson, 1997, 1999) proposed that positivity and negativity are separable dimensions of the affect system that, when combined, produce an overall summary approach or avoidance motivation (see also Norris, Gollan, Berntson, & Cacioppo, 2010). This hypothesis was based on extensive coverage of findings from multiple methods and levels of analysis, from Brown (1948) and Miller’s (e.g., 1959) work on approach-withdrawal motivation (and their co-activation) in albino rats, up to work on attitudes and the limitations of bipolar rating scales (cf. Klopfer & Madden, 1980), and is the central guiding principle of the Model of Evaluative Space (ESM; Cacioppo & Berntson, 1994). The ESM includes a broad set of related postulates, many of which are beyond the scope of the current paper (see Norris et al., 2010, Table 1); here we focus on the separability of positivity and negativity, and the predictions that arise from this separability.

First, if positivity and negativity are separable, then they are not bound to be equivalent in their constitution, operations, or consequences (Norris et al., 2010). The separability of positivity and negativity allows for distinct *activation functions*, which characterize the relationship between input to a system and output of that system. In other words, the same amount of input to the positivity and negativity sub-systems can produce differential output. Specifically, the ESM predicts that the negativity sub-system produces a stronger output than does the positivity system, given equivalent input. This negativity bias is predicted to be a difference in *gain* of the systems, meaning that at very low levels of input (i.e., in response to neutral or very slightly emotional stimuli), little to no difference

in output of the positivity and negativity systems is observed¹; but that for every additional unit of input, negativity increases at a faster rate than does positivity, resulting in greater output of the negative system than the positive system at higher levels of input.

Furthermore, the activation functions for positivity and negativity are predicted to be negatively-accelerating, meaning that their slopes decrease (i.e., flatten) as input increases. The ESM also predicts a negativity bias in the degree of negative deceleration, such that the slope of the activation function for negativity is predicted to flatten at a slower rate than the activation function for positivity (see Figure 1 for an illustration of the activation functions or gradients for positivity and negativity). Combined, these predictions imply that each additional unit of input to the negativity sub-system carries *even* greater weight than does each additional unit of input to the positivity system (for evidence supporting this conclusion in the realm of political attitudes, see Holbrook, Krosnick, Visser, Gardner, & Cacioppo, 2001). Norris, Gollan, Berntson and Cacioppo (2010) have previously summarized the wide-ranging evidence both for (a) the separability of positivity and negativity, and (b) a negativity bias in affective processing.

An Evolutionary Argument for the Negativity Bias. The ESM, along with other theoretical approaches or reviews of the literature (e.g., Rozin & Royzman, 2001), argues that the negativity bias is adaptive, in that it confers an evolutionary advantage that prioritizes survival over all other motivations. Such an argument is necessarily post hoc and untestable (therefore also unfalsifiable), but also sets forth some important predictions

¹ The ESM predicts a second asymmetry in affective processing, the *positivity offset*, which is dominant at low levels of input to the systems and results in a slightly greater output for positivity than for negativity; this second affective asymmetry is beyond the scope of the current paper and will not be discussed further. See Norris et al. (2010) for a theoretical perspective on the positivity offset and Norris, Larsen, Crawford, and Cacioppo (2011) for empirical evidence.

and is thus worth addressing briefly here. In their review of research on “negative potency”, Rozin and Royzman (2001) describe this perspective well:

“Avoiding risks of death must be a matter of the highest priority in the evolutionary scheme: the peak of vigilance and investment would well be oriented to escape death. It is true that reproduction is the final measure of evolutionary success, but there are usually multiple opportunities to reproduce, and death terminates these options.”

The ESM similarly assumes that survival carries greater weight than an opportunity unpursued.

The possibility that the negativity bias has been sculpted by evolution also results in at least three (somewhat testable) hypotheses, each of which is addressed in the current review:

1. If natural selection has favored individuals who exhibit a negativity bias, then (a) on average, most individuals should exhibit stronger responses to aversive than to appetitive stimuli, and (b) some variation in the negativity bias should exist among individuals, given that variation is the engine of natural selection (i.e., variance is a necessary part of competition between and selection of adaptive traits). In other words, stable, general individual differences in the negativity bias should exist and might have consequences for fitness (e.g., mate selection, mental and/or physical health; this latter hypothesis is beyond the scope of the current paper and will not be addressed further).
2. Assuming that there are differential fitness pressures and/or goals as a function of gender (whether biologically- or socially-driven), there may be functional gender differences in the magnitude of the negativity bias. For example, given that males may be more responsible for ensuring reproductive success (either biologically, socially, or culturally), they may be more driven by strongly appetitive stimuli, including possible

mates (i.e., erotica) and/or other rewards that may contribute to this success (e.g., food, resources, money). Thus, males may exhibit a weaker negativity bias than do females.

3. Similarly, as fitness needs change over the course of life, we may expect developmental differences in the magnitude of the negativity bias. First, the negativity bias may emerge later in development (i.e., not be innate), as infants become more mobile and less dependent on caregivers. Second, the negativity bias may weaken toward the end of life, after the age of primary reproductive ability (especially for females) and as we approach mortality; or at the very least, responses to certain categories of affective stimuli may change (e.g., babies, attractive conspecifics). On the other hand, a contradictory hypothesis would state that if survival is prioritized above all else, the negativity bias may emerge early in development and be relatively maintained throughout life.

Although evolutionary theories are difficult to support with empirical evidence, the argument for an evolutionary explanation for the negativity bias is an important motivator for much of the research covered here, and these three hypotheses have driven a number of the studies summarized in this review.

The Negativity Bias: Empirical Considerations

Before we turn to new findings in research on the negativity bias, it is necessary to provide a working definition of the term *negativity bias*. Based on the theoretical predictions of the ESM, the negativity bias results from asymmetries in the activation functions for negativity and positivity, such that the activation function for negativity is both steeper and more stable (i.e., has a lower rate of deceleration). In other words, holding input to the two sub-systems constant, the negativity bias would manifest as more extreme

output from the negativity system than for the positivity system at higher levels of input.²

To test this prediction, the following conditions must be met in a single study:

1. Both positive/pleasant and negative/unpleasant stimuli must be included.
2. Positive and negative stimuli must fall on the higher end of the input scale (i.e., moderately to extremely pleasant/unpleasant).
3. Positive and negative stimuli must be demonstrably equivalent in all dimensions except for valence.
4. The measure of output must be equally valid and specific for positive and negative stimuli.

Critical for the demonstration of the negativity bias are careful attention to and matching of the (positive and negative) stimuli used, whether words, images, sounds, events, personality characteristics, gamble outcomes, or others. Without adequate matching, results may be attributable to any difference between positive and negative stimuli, including but not limited to: arousal, dominant color, complexity, or familiarity.

Indeed, Taylor (1991) addressed this issue well in her early review of the negativity bias:

“There is an issue of calibration involved in comparing negative and positive events: How does one know that the negative stimuli (events, trait words, and the like) are as negative as the positive stimuli are positive? The strongest case can be made in studies in which the positive and negative stimuli involved occur on the same interval scale (e.g., the costs or gains in dollars of a wager): Any inequivalency of the positive and negative stimuli is psychological, and therefore is part of the phenomenon, not a confound. A less strong but defensible case can be made when the positive and negative stimuli are rendered equivalent on some scale related to the inference to be drawn (p. 68).”

² Note that this is only one (arguably the strongest) prediction of the ESM regarding the negativity bias. Based on other aspects of the model, we would also predict that (a) the slopes of the activation functions relating input to output should differ, such that the negativity slope is steeper than the positivity slope, and (b) the degree of curvature (i.e., negative deceleration) of the activation functions should differ, such that the positivity slope decelerates at a higher rate than the negativity slope. All three predictions are made by the ESM; but comparing output of the two systems at higher levels of (equivalent) input is the most straightforward, empirically valid hypothesis.

Many studies that claim to test the negativity bias fail to adequately match positive and negative stimuli, which is required for essentially any category of affective stimulus with the exception of currency (e.g., \$10 is \$10 regardless of whether it is won or lost, and yet a loss of \$10 may have a stronger impact on emotional responses than a gain of \$10; Kahneman & Tversky, 1984; Norris et al., 2011). Often matching may be accomplished by relying on published (or collected) normative affective ratings, as with the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008); in which case negative and positive stimuli may be chosen to be *equidistant from the midpoint of the normative valence scale* to properly calibrate *extremity*.³

Finally, consideration must also be given to the measurement of output, or of *reactivity* to each stimulus. Self-report measures (e.g., rating scales) may be used, but ideally not bipolar scales (i.e., ranging from *very unpleasant* to *very pleasant*). Separate unipolar scales for positivity and negativity are arguably equivalent measures of the output of the sub-systems. The Evaluative Space Grid (ESG; Larsen, Norris, McGraw, Hawkey, & Cacioppo, 2009) was developed to more efficiently collect positive and negative responses using a single-item measure. In brief, the ESG is a 5x5 grid in which positivity is represented on the x-axis and negativity on the y-axis; participants select one of the 25 cells

³ Note that other approaches to studying differences in emotional responses to pleasant and unpleasant stimuli have provided evidence both in support of and at odds with the functioning of a negativity bias. For example, research using facial expressions (e.g., happy vs. sad/angry/fearful faces) often has shown that negative/unpleasant expressions elicit stronger behavioral, neural, and physiological responses to the latter than the former (e.g., Leppänen et al., 2007; Adolphs, 2008; Yang et al., 2012; but cf. Kauschke, Bahn, Vesker, & Schwartz, 2019 for a more nuanced view across development). To the contrary, studies using categories of IAPS images have sometimes shown equivalent or even stronger responses to pleasant (e.g., erotica) than to unpleasant (e.g., mutilation, snakes) images (e.g., Weinberg & Hajcak, 2010; for sex differences in this approach cf. Bradley, Codispoti, Sabatinelli, & Lang, 2001; Sabatinelli et al., 2004). Both types of evidence, however, fail to meet the criterion of equivalency and are therefore beyond the scope of the current paper.

that best corresponds to their current feelings. McGraw, Larsen, Kahneman, and Schkade (2010) have also introduced a common unipolar intensity scale, in which individuals indicate the intensity of their feelings on a scale ranging from *no effect* to *a very large effect* that effectively measures reactivity to both positive and negative stimuli. In addition to self-report measures, indirect assessments, such as psychophysiological measures (e.g., ERPs, fMRI) may be used to measure output of the positivity and negativity affect sub-systems. Critically, not all psychophysiological measures are appropriate. Facial electromyography (EMG), for example, is a validated measure of positive and negative affect (i.e., over the *zygomaticus major* and *corrugator supercilii* muscle groups, respectively; Larsen, Norris & Cacioppo, 2003); however, the fact that different muscle groups are differently sensitive to positive and negative affect (and that the relationship between muscle activity and valence differs for each, as well) is problematic for measuring asymmetrical responses. Similarly, neuroimaging approaches that focus on neural regions and/or networks strongly associated with either positive (e.g., nucleus accumbens) or negative (e.g., insula) affect are subject to the same criticism. In sum, researchers must be thoughtful in both the selection of stimuli (input) and of measurement of reactivity (output) in order to provide a true test of the negativity bias.

The term *negativity bias* has occasionally been used to describe (or has been confused with) other documented affective phenomena. We briefly cover two of these – threat bias and interpretations of ambiguity – to better illustrate how each differs from the negativity bias as predicted by the ESM.

Threat bias. Attention is biased toward threatening (versus neutral) stimuli (cf. Öhman, Lundqvist, & Esteves, 2001); this bias is often called “threat bias” or “attention to

threat bias”, but has occasionally been termed a “negativity bias.” The dot-probe task (cf. MacLeod, Mathews & Tata, 1986; Mogg & Bradley, 1999), in which a pair of stimuli (typically faces; one neutral and one threatening) appear briefly on opposite sides of a display and are followed by a dot in the location of one of the stimuli that requires a response, is often used as a measure of threat bias (or threat avoidance, depending on trial timing). Anxious individuals exhibit strong threat bias (non-anxious sometimes do not; Bar-Haim, Lamy, Pergamin, Bakersman-Kranenburg, & van IJzendoorn, 2007), and interventions that target threat bias have been effective in reducing anxiety (Browning, Holmes, Murphy, Goodwin & Harmer, 2010). A great deal of research has examined the threat bias across the life span (e.g., in pediatric and adolescent anxiety; Britton et al., 2013), with different stimuli (e.g., sad, fear faces; Sylvester, Hudziak, Gaffrey, Barch, & Luby, 2016), in different psychiatric disorders (e.g., post-traumatic stress disorder [PTSD]; Bardeen & Orcutt, 2011), and using electrophysiological approaches (e.g., ERPs; O’Toole & Dennis, 2012). Although the functioning of a threat bias (in the dot-probe task as well as many others) is consistent with the basic premise of the negativity bias (that bad is stronger than good), such studies rarely include (a) both positive and negative stimuli that are (b) matched with regard to their normative valence. Thus, studies of threat bias typically do not qualify as tests of the negativity bias, and therefore fall outside of the scope of the current paper.

Interpretations of ambiguity. Individuals differ in their tendency to interpret ambiguous stimuli (e.g., surprise faces) as negative or positive (Kim et al., 2003).⁴ The term

⁴ It is important to note here the difference between “ambivalent” and “ambiguous” stimuli. “Ambivalent” stimuli are those that are “ambi-” (both) “valent”; or both positive and negative at the same

“negativity bias” (or “positivity-negativity bias”; Neta, Norris & Whalen, 2009; and more recently “valence bias”; Neta et al., 2017) has been used to describe this tendency. Research on this tendency has shown that: more negative interpretations of surprise faces are associated with increased amygdala activation (Kim et al., 2003) and increased activation over *corrugator supercilii* (i.e., the brow muscle typically associated with negative affect; Neta, Norris & Whalen, 2009); low-spatial-frequency (LSF; associated with early processes and increased ventral amygdala activation) presentations of surprise faces bias interpretations towards the negative, and this bias is stronger for those with a general positive bias (as those with a general negative bias consistently rate surprise faces as more negative; Neta & Whalen, 2010); the default interpretation of surprise faces is negative (Neta, Davis & Whalen, 2011); threat of shock (TOS; i.e., a negative context) increases the tendency to interpret surprise faces as negative (Neta, Cantelon, et al., 2017); and patterns of eye fixation predict positive (i.e., earlier fixation and longer fixation on the mouth) versus negative interpretations (Neta, Tong, et al., 2017). Furthermore, Tottenham and colleagues (2013) have argued that the negativity bias in interpretation of surprise faces is “first and fast” and can be overcome by regulatory control mechanisms over the course of development. They have found that children show a strong negativity bias in interpretation of surprise faces (i.e., tendency to interpret ambiguous surprise faces as negative), which is accompanied by greater *corrugator supercilii* activation; but that more positive interpretations emerge later in development (i.e., during adolescence), as control increases. Although these findings are certainly in line and supportive of a general negativity bias in affective processing (e.g., the *initial negativity default*; Tottenham et al., 2013), the

time. “Ambiguous” stimuli are those that do not inherently carry one specific valence but may be interpreted as *either* positive or negative (e.g., surprise faces). The line of research discussed here concerns the latter.

interpretation of ambiguous stimuli falls outside of our definition of the negativity bias (which is focused on asymmetrical responses to equally negative and positive stimuli) and a thorough review of this literature is beyond the scope of the current paper.

Recent Evidence Supporting a Negativity Bias

Given that a number of thorough, comprehensive reviews of past research supporting a negativity bias have previously been published (e.g., Taylor, 1991; Rozin & Royzman, 2001; Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001), the review of recent evidence provided here is meant to supplement and extend those previous findings and is not meant to provide complete coverage of all work supporting the negativity bias in the past 20 years. This review covers empirical work, much conducted by my colleagues and myself, in two areas. First, evidence from multiple methodologies in the field of social neuroscience (e.g., event-related brain potentials [ERPs], functional magnetic resonance imaging [fMRI]) supports the functioning of a negativity bias in neural responses to affective stimuli. Second, examining individual differences, both in personality (or factors that function like traits) and in age and sex, can shed light on the functioning of the negativity bias.

Evidence for a Negativity Bias from the Field of Social Neuroscience

Numerous studies have examined the negativity bias using electrophysiological, neuroimaging, and other measures to (a) bypass issues related to self-report methods and test the hypothesis that negativity has a stronger impact on (neural and bodily) responses than does positivity, and (b) explore the mechanisms that might give rise to the negativity bias. Furthermore, the ESM predicts that positivity and negativity are not only functionally separable, but that they arise from at least partially separable neurophysiological

substrates (Norris et al., 2010), giving rise to the prediction that positive/pleasant and negative/unpleasant stimuli should have differential effects on neural processes.

ERPs. In an initial ERP study using a modified oddball paradigm in which rare pleasant and unpleasant images (i.e., “targets”) taken from the IAPS (Lang, Bradley, & Cuthbert, 2008) were embedded in series of frequent neutral images, Ito, Larsen, Smith and Cacioppo (1998) found that the late positive potential (LPP), a component associated with evaluative categorization (Cacioppo, Crites, Berntson, & Coles, 1993) and related to the oddball P300 (Ito et al., 1998), was larger for unpleasant targets than for equally extreme, arousing, and rare pleasant targets. This finding indicates that, even when unpleasant and pleasant stimuli are equally unexpected, extreme, and arousing, unpleasant (i.e., negative) images have a stronger impact on recruitment of neural resources than do pleasant (positive) images. Ito and colleagues (1998) arguably provided the first neural evidence of a negativity bias in relatively late-stage affective processing.

Focusing on an earlier stage of processing, Smith and his colleagues (2003) found that the P1 component of the ERP, an index of attentional allocation that occurs approximately 100 ms post-stimulus, is larger to rare unpleasant than to rare pleasant images. Thus, negative information garners more neural resources than does equally rare, extreme, and arousing information both at early attentional stages (P1) and later evaluative categorization stages (LPP) of processing.

Huang and Luo (2006) investigated the negativity bias at multiple stages of processing in the same study. Participants viewed neutral (50%) and either positive or negative (50%; valence of pictures was blocked) images from the IAPS (Lang, Cuthbert, & Bradley, 2008), and categorized each as either neutral or positive/negative with a button

press (the response hand was also blocked to allow for an investigation of the lateralized readiness potential [LRP], the latency of which is an index of primed motor behavior; Hsieh & Yu, 2003). Huang and Luo (2006) found that (a) the P2, an early attention-related component, was larger in negative than positive blocks; (b) the LPP was larger for negative than for positive images (even though both were as frequent in this design as neutral images), and (c) the latency of the LRP was shorter in negative than positive blocks. Together, these data suggest that the negativity bias emerges early in attentional processes, but extends through a later evaluative-categorization stage, and results in increased motor priming for responses to negative stimuli (Huang & Luo, 2006).

fMRI. In addition to studying the time course and stages of processing at which a negativity bias functions in response to affective stimuli, neuroimaging research using fMRI may shed light on the neural networks implicated in the negativity bias. In two studies, undergraduate females were scanned while viewing neutral, negative, and positive images taken from the IAPS; all three categories of images were matched on dominant colors, complexity, and content, and negative and positive images were matched on extremity and arousal. In both studies, negative images elicited greater activation of BA 19 ($x = 58, y = -54, z = 12$), a region of visual association cortex implicated in selective attention, than did comparably extreme and arousing positive images (Norris, 2004; see Figure 2). In Study 1, a region of the anterior cingulate cortex (ACC; $x = 1, y = 25, z = 8$) also showed a negativity bias in activation; however, this did not replicate in Study 2 (which contained fewer images in the service of time; this may have reduced our power to detect differences in neural activation).

Connecting these fMRI results to previous findings using ERPs, Clark and Hillyard (1996) have previously localized the P1 to BA 19, suggesting that the negativity bias in early attentional ERP components previously reported by Smith and colleagues (2003) may be directly related to the negativity bias observed in activation of visual association cortex (BA 19; Norris, 2004). Negative images, therefore, appear to elicit greater selective attention than do well-matched positive images. Notably, the ACC is involved in executive attention, suggesting that the neural negativity bias observed in Study 1 in the ACC may be related to a bias to direct attention toward negative stimuli.

In a third study, we investigated the negativity bias using a repetition suppression paradigm, in which participants viewed and rated negative, neutral, and positive images, some of which were repeated (Norris, Monteleone, & Cacioppo, 2019). Repetition suppression refers to the finding that repeated presentations of stimuli (e.g., visual scenes; Epstein, Parker, & Feiler, 2008; Jessen et al., 2002) often result in decreased neural activation in broad regions of cortex, consistent with the interpretation that multiple stimulus presentations results in perceptual adaptation and decreased novelty. Repetition enhancement – in which neural activation *increases* with stimulus repetition – has also been reported. For example, Salimpoor, Chang, and Menon (2009) reported repetition enhancement in the hippocampus, posterior cingulate, precuneus, and supplementary motor areas for repeated math problems; neural enhancement in these regions was associated with improved response times on the problems, consistent with performance facilitation. We found that positive and neutral images showed typical repetition suppression in visual areas, including occipital and inferotemporal cortices; whereas negative images elicited repetition *enhancement* in attentional networks (posterior parietal

cortex, anterior cingulate), the insula, and primary motor areas (Norris, Monteleone, & Cacioppo, 2019). Ratings of negative images also decreased in latency and increased in negativity with repetition, suggesting that both neural and behavioral responses to negative stimuli are heightened with repetition, consistent with a negativity bias.

Contradictory Evidence. In contrast to results reported above, a number of studies have provided evidence suggesting either no differences in neural processing of pleasant and unpleasant stimuli or occasionally even a bias towards heightened responses to pleasant (e.g., erotica images) stimuli. Although many of these studies fail to meet the criterion of equivalency (or at least do not provide the appropriate statistical analyses to verify proper matching), a sampling is provided here to acknowledge that contradictory evidence does exist. Schupp and his colleagues (2000) used a slight modification of the oddball paradigm described above and again found larger LPP amplitudes to pleasant and unpleasant IAPS images (stimuli were not matched in any way); but no differences between the two. Using fMRI, Bradley and her colleagues (2003) examined activation of the occipital cortex in response to categories of stimuli including erotica, mutilation, and more neutral images (e.g., families, household objects), and found increased activation for images “related to primary motive states,” including erotica and mutilation images, which did not differ from each other. In addition, Bradley et al. (2003) included neutral and angry faces and found no difference in activation of the occipital cortex to these categories. Although these data are informative, they may simply indicate the lack of a negativity bias at the level of early visual processing and do not contradict the functioning of a negativity bias at higher levels (i.e., as their analyses were limited to dorsal and ventral primary visual

regions).⁵ Finally, Weinberg and Hajcak (2010) examined ERPs to unpleasant, neutral, and pleasant IAPS images (which were subsequently divided into content categories) and found no differences in early ERPs (i.e., N1 and the Early Posterior Negativity [EPN]), but did replicate Ito et al. (1998) in that larger LPP amplitudes were elicited by unpleasant than pleasant images. When analyzing based on content, however, erotica and mutilation images elicited comparable LPPs. These results are complicated by the fact that, arguably, (a) pleasant and unpleasant images were not well matched, with unpleasant images being more extreme ($M_{\text{valence}} = 1.90$; $M_{\text{extremity}} = 3.1$) than were pleasant images ($M_{\text{valence}} = 7.43$; $M_{\text{extremity}} = 2.43$), and normative ratings are not provided for the erotica and mutilation categories. For a thorough review of ERP findings in response to unpleasant/negative and pleasant/positive images, see Olofsson, Nordin, Sequeira, and Polich (2008); but note that this review does not pay specific attention to the criterion of equivalency.

In sum, evidence contracting the functioning of a negativity bias in behavioral and neural responses to unpleasant vs. pleasant stimuli does exist, with some studies showing no differences and others even showing heightened responses to pleasant stimuli (especially for males' responses to erotica); but many of these studies fail to demonstrate that they have met the criterion of equivalency.

Individual Differences

Process-based measures of the negativity bias. If the negativity bias is, as we have suggested (Norris et al., 2010), evolutionarily based, we would predict that most individuals should exhibit a negativity bias, on average; but that differences in the

⁵ It is also worth noting that participants in Bradley et al. (2003) were all male; we return to gender differences in the negativity bias and to neural and behavioral responses to erotica and mutilation pictures later.

magnitude of the negativity bias would emerge across individuals, given that variation is the engine of natural selection. To examine individual differences in the negativity bias, we developed a process-based measure of this asymmetry in affective functioning that relies on participants' emotional responses to stimuli rather than on introspection of personal traits (Norris, Larsen, Crawford & Cacioppo, 2011). Specifically, participants indicated how they felt using the ESG (Larsen et al., 2009) about pictures (IAPS; Lang, Cuthbert & Bradley, 2008), sounds (IADS; Bradley & Lang, 2007), and words (ANEW; Bradley & Lang, 2017) that ranged from extremely unpleasant to extremely pleasant, as based on normative ratings. Negative ratings of the most unpleasant images and positive ratings of the most pleasant (and equally extreme) images were used as estimates of individuals' reactivity to negative and positive stimuli; the difference between these represents an individual's process-based negativity bias (Norris et al., 2011).⁶

Participants did exhibit a negativity bias on average, but the magnitude of this measure differed across individuals. Importantly, the negativity bias emerged in responses to a wide variety of stimuli, including pictures, sounds, and words, and the magnitude of the bias was correlated across these independent measures, suggesting that it is a general asymmetry in affective processes. The negativity bias also exhibited significant temporal stability across a 2-week and even a 1-year interval. Finally, both our process-based measure of the negativity bias and more traditional inventory-based measures of affective processing (e.g., the BIS/BAS, PANAS) predicted behavior; but did so independently on

⁶ Note that we argue that both process-based measures, which do not rely on self-reports of how one "typically" feels or responds, and more traditional reflective measures (i.e., surveys; such as the Behavioral Inhibition and Activation Scales [BIS/BAS; Carver & White, 1994] and the Positive and Negative Affect Schedules [PANAS; Watson, Clark & Tellegen, 1988]) both are important for understanding individual differences in affective processing. Interested readers are referred to Norris and colleagues (2011) for more discussion of these issues.

separate measures of emotional processing. Furthermore, our measure of the negativity bias was not correlated with inventory-based measures of personality (e.g., neuroticism, BIS, NA), again suggesting a double dissociation. Thus, individual differences in the negativity bias are general processing characteristics that emerge in multiple contexts and in responses to broad categories of stimuli, are stable over time, exhibit predictive validity, and are not redundant with other inventory measures of personality traits (Norris et al., 2011).

We also assessed individual differences in this process-based measure of the negativity bias as part of two fMRI studies reported above (Norris, 2004). In Study 1, participants with a higher negativity bias exhibited (a) greater activation to negative than to positive images in BA 19, $r = .72$, $p < .01$ (see Figure 2), and (b) greater neural activation to negative than to positive images in the ACC, $r = .61$, $p < .05$. In Study 2, participants with a higher negativity bias exhibited greater activation to negative than to positive stimuli in BA 19, $r = .32$, one-tailed $p = .053$. Together, these findings suggest that individual differences in the negativity bias (as measured by self-reported affective responses) are related to neural biases toward negative stimuli).

Finally, the use of IAPS images (or IADS sounds or ANEW words) to examine the negativity bias does inherently present some complications. First, there are notable gender differences in how individuals rate certain categories of images. For example, females tend to rate images of mutilations, dead bodies, and other highly arousing unpleasant images as higher in arousal and negativity than do males; and males tend to rate erotica images as higher in positivity than do females (see normative ratings for IAPS in Lang, Bradley & Cuthbert, 2008; 2015). In addition, there is often more variance in how both males and

females rate erotica (as can be seen by higher standard deviations for these images in the normative ratings), which also complicates their use in such studies. We have often addressed this problem in the past by recruiting females only (although Norris, Monteleone, & Cacioppo, 2019 recruited only males), which allows us to both choose and match unpleasant and pleasant stimuli more carefully. Clearly, this presents an issue for generalization. A second approach that we have taken in the past (Norris et al., 2011; Study 2) is to examine the negativity bias in a game of chance, in which participants won or lost money (\$1, \$5, \$10, \$25, \$50, \$75, \$100, \$150) based on the toss of a coin. Arguably, given that \$1 is a dollar whether lost or gained, this study better matched positive (wins) and negative (losses) stimuli. Results still indicated a negativity bias, and replicated the slopes predicted by the ESM.

Other personality dimensions associated with a negativity bias.

Neuroticism. One of the most broadly studied of the Big Five Personality Dimensions, neuroticism is a trait characterized by high levels of negative affect, anxiety, and low emotional stability. Although we previously have found no relationship between neuroticism and our process-based measures of individual differences in the negativity bias (Norris et al., 2011), it is possible that individuals higher in neuroticism may exhibit a bias toward negative stimuli in other measures. Norris, Leaf and Fenn (2018) examined both an overall negativity bias in both correct and false memory using the Deese-Roediger-McDermott illusory memory paradigm (DRM task), as well as moderation of these findings by neuroticism. Across two studies, individuals showed a negativity bias in their correct memory, such that negative list words were remembered better than were positive list words. When tested immediately after the learning phase, individuals also exhibited

greater false memory for negative versus positive critical lures, consistent with a negativity bias. When tested after a 24-hr delay, only those individuals higher in neuroticism showed a maintained negativity bias in false memory, suggesting that neuroticism may increase rumination of (and subsequent “memory” for) critical lures associated with negative list words.

Gender differences. Many of the studies summarized above focused on female participants (Norris, 2004; Norris et al., 2011), but given the evolutionary function of the negativity bias, we might predict a gender difference. Specifically, given that males are expected to explore the environment and protect the social group, as well as being arguably more responsible for ensuring procreation (which may be biologically-, socially-, or culturally-motivated), we might predict that they have a somewhat smaller negativity bias than females. Indeed, we have found that in a mixed-gender sample, although a negativity bias was observed on average, males exhibited a smaller negativity bias than did females (Norris, 2019). This difference emerged for nonsocial pictures (i.e., those that did not depict people), but not for social pictures (i.e., those depicting people), indicating that social stimuli may be inherently more evocative and relevant for both genders.

Furthermore, the gender difference in the magnitude of the negativity bias was predicted by a measure of prenatal testosterone exposure. Specifically, fingers have receptors for sex hormones and prenatal hormone exposure can affect finger length. The 2nd to 4th digit ratio (2D:4D; length of the 2nd digit divided by length of the 4th digit of the dominant hand) has been found to be negatively correlated with prenatal testosterone levels and is lower in males than females (Manning, Scutt, Wilson, & Lewis-Jones, 1998; McIntyre, 2006). We found that 2D:4D predicted the magnitude of the negativity bias, such that lower 2D:4D

was associated with a smaller negativity bias (Norris, 2019). In sum, males exhibit a smaller negativity bias, possibly encouraging more risk taking behaviors (Sapienza, Zingales, & Maestripieri, 2009), and this sex difference is driven at least in part by prenatal exposure to testosterone.

Ashare and colleagues (2013) examined the independent and interactive effects of gender and polymorphisms of the serotonin receptor gene 102T>C on individual differences in the negativity bias. Females with a C/C genotype had a higher negativity bias than did those with a T/T genotype (and the pattern was dose-dependent, as those with a T/C genotype had middling negativity biases); males showed no differences in the negativity bias across genotypes. The C/C genotype is associated with impaired emotion regulation, including the development of depression (Jokela et al., 2007), schizophrenia (Abdolmaleky, Faraone, Glatt, & Tsuang, 2004), suicide attempts (Vaquero-Lorenzo et al., 2008), and anxiety-related traits (Golimbet, Alfimova, & Mityushina, 2004). Thus, there is some evidence to suggest that the serotonin system may be implicated in the negativity bias; and that higher negativity biases may lead to affect dysregulation (at least in females).

Age. As the ESM (and other theoretical accounts) have suggested that there is an evolutionary advantage for the negativity bias, one might predict that (a) even infants and young children may exhibit a bias towards negative stimuli, and (b) that this bias should persist throughout the lifespan, or at least through life stages in which survival is prioritized over other motivations. An alternative prediction would state that, as both motivation and needs shift over the course of life, such biases towards appetitive vs. aversive stimuli might also shift. For example, young infants who lack mobility and are dependent on caregivers might show stronger responses towards appetitive stimuli (e.g.,

food, helpful others) than towards aversive or harmful stimuli (e.g., hinderers); and again, in older age when mortality is near, perhaps another shift again towards the positive.

Evidence for a negativity bias in infants and children. Hamlin, Wynn and Bloom (2007; Study 1) found that 6- and 10-month old infants showed a behavioral preference for a “helper” over a “hinderer” (i.e., a figure that either helped or hindered a target stimulus from climbing up a hill), and indicated surprise (via longer looking times) when the climber approached the hinderer versus the helper. In an additional study, Hamlin and colleagues (2007; Study 3) showed both approach behavior toward the helper and avoidance of the hinderer as compared to a neutral figure. Chae & Song (2018) replicated these results and showed that 6- and 10-month infants looked longer at an agent who hindered a target as compared to a neutral agent (who was present but did not interact with the target); but exhibited equal looking times for a helping agent and a neutral agent. Together, these results indicate that infants make decisions based on the valence of acts observed by social agents, but that a negativity bias exists in attention toward hindering agents. In a follow-up to these results, Hamlin, Wynn and Bloom (2010; Study 2) examined the same behaviors in 3-month olds, and reported longer looking times for the neutral agent than the hindering agent, but equal looking times for the neutral and the helping agent. They interpret these results as being consistent with a negativity bias; infants as young as 3 months old learn more quickly about antisocial agents than prosocial agents. Although the results from Chae & Song (2018) and Hamlin and colleagues (2010) are somewhat contradictory (i.e., longer looking times for the hindering vs. neutral agent in the former; longer looking times for the neutral vs. the hindering agent in the latter), given the comparison with the other condition (in which looking times were equivalent for helping and neutral agents in both studies),

they do suggest that hindering agents (i.e., negative social stimuli) are preferentially distinguished at as young as 3 months of age.

Vaish, Grossmann and Woodward (2008) provide a comprehensive review of the argument for a negativity bias emerging early in development, as well as the evidence to date that supports a negativity bias in infants and young children, even as they state that “Despite the ubiquity of the negativity bias in adulthood, no explicit theoretical or empirical work has examined this bias in development (p. 385).” Vaish and colleagues (2008) summarize findings both in the realm of social referencing (as in Hamlin and colleagues’ work; 2007, 2010) as well as other developmental domains, and conclude that even in the absence of strong empirical work on the development of the negativity bias in infancy and childhood, research provides overwhelming evidence in support of an early-emerging negativity bias. In addition, they provide a thorough consideration of both the evolutionary and ontogenetic mechanisms that may give rise to a negativity bias in infants and young children, as well as a series of testable hypotheses regarding these mechanisms.

It is worth noting that developmental research on the negativity bias in infants and young children may not meet the criteria put forth at the beginning of this review; namely, it is difficult or even impossible to know whether positive and negative stimuli are “matched” in this population. Is the “helper” equivalent to the “hinderer” in every dimension except the act observed? Similarly, other relevant work in infants and children relies on facial expressions: but it is equally difficult to know whether a smiling/happy face is as arousing and extreme as is a frowning/angry face.

Having said that, research using facial expressions as stimuli and measuring ERPs in infants may provide additional support for the early emergence of a negativity bias.

Yrttiaho and colleagues (2014) acquired ERPs and attention disengagement data in response to fearful, happy, neutral and scrambled faces in the same infants at 5- and 7- months of age. ERPs indicated a bias toward fearful faces, with fear-sensitive activity over posterior sites starting with the N290 wave. In addition, the fear attention bias at 5 months was correlated with that at 7 months, indicating a stable individual difference. Xie and colleagues (2018) also found evidence for a fear bias in 5-, 7-, and 12-month old infants, as fear faces elicited larger N290 amplitudes than did happy faces; however, angry faces did not elicit a bias in this component, but did elicit larger P400 amplitudes than did fear and happy faces. Combined, these data suggest a bias in the processing of negative facial expressions at two different stages of processing.

Crepeo-Llado, Vanderwert, and Geangu (2018) examined ERPs in 8-month olds while they listened to peers' nonverbal vocalizations of emotion (i.e., crying, laughing, coughing). Ratings of arousal provided by university students suggested that crying and laughing were matched (at least for young adults). Crying produced larger N100 and late positive component (LPC) amplitudes than did laughing and coughing, providing evidence for a negativity bias in neural responses to nonverbal vocalizations in infants. Missana, Altvater-Mackensen, and Grossman (2017) also found an enhanced N200 when infants listened to a peer crying vs. laughing; however, laughing resulted in an enhanced P300, suggesting that an early attentional bias toward an unpleasant stimulus may be followed by selective attention toward a pleasant stimulus in infants as young as 8-months old. Taken together, these studies provide additional evidence for the early emergence of a negativity bias in development.

Switch to a 'positivity effect' in older adults? Seemingly in contrast to the general functioning of a negativity bias are findings suggesting a *positivity effect* in older adults. Carstensen, Mather and their colleagues (Carstensen & Mikels, 2005; Mather & Carstensen, 2005; Carstensen & DeLiema, 2018) have argued that there is a tendency for older individuals to experience fewer negative emotions, and that this tendency may be driven by strategic processes in their attention and memory that enhance positive information. Indeed, they have found that older adults exhibit memory distortions for positive information, whether autobiographical or presented as images in a laboratory (Mather & Carstensen, 2005). They have argued that this positivity effect reflects motivational changes in later life, and is not due to neural or cognitive decline (Carstensen & DeLiema, 2018).

Kisley and colleagues (Wood & Kisley, 2006; Kisley, Wood & Burrows, 2007) have used ERPs to examine the negativity bias in older (versus younger) adults. Using the same modified oddball paradigm as Ito and colleagues (1998; Study 2), Wood and Kisley (2006) replicated the finding that in a sample of young adults, the LPP was enhanced to rare pleasant and unpleasant IAPS images as compared to frequent neutral images, and that the LPP was larger to unpleasant than to pleasant images. A sample of older adults, however, showed generally attenuated LPPs to both unpleasant and pleasant images, and no difference between the two; providing evidence for an elimination of the negativity bias in later life. In a separate study, Kisley, Wood and Burrows (2007) used the same modified oddball paradigm but included participants that ranged in age from 18 to 81 to allow for an investigation of relationships between age and LPP amplitudes to neutral, unpleasant, and pleasant images. Across all participants, they found that the LPP was smallest to neutral,

middling to pleasant, and largest to unpleasant images, effectively replicating Ito and colleagues' (1998) results within a broader age range of participants, and supporting a negativity bias in evaluative categorization. Correlations, however, indicated that age was inversely correlated with LPP amplitudes to unpleasant images ($r = -.32$) but was not correlated with LPP amplitudes to pleasant images ($r = .00$); the difference between these correlations was significant. In sum, Kisley and colleagues' (2006, 2007) findings are consistent with a reduction (or elimination) of a neural measure of the negativity bias in older age, but they argue this reduction is due to *decreased* neural reactivity to negative stimuli, not to *increased* reactivity (or motivation) to positive stimuli. Thus, their results are somewhat in contradiction to Carstensen's theory of the positivity effect.

Arguably, some of the work done by Carstensen and her colleagues may not meet the very strict definition that we have employed for testing the negativity bias. We have, however, examined our process-based measures of the negativity bias in younger and older adults' responses to negative, neutral, and positive images (replication & extension of Norris et al., 2011), and have found *no differences* in the magnitude of the negativity bias across age groups (Norris & Flanary, 2019); both younger (18-22 year old) and older (65-80 year old) female adults exhibited a negativity bias in their responses to affective images.

Furthermore, we have found that individual differences in a global negativity factor (composed of neuroticism, negative affect, anxiety, depression) are associated with both structural and functional neural differences in older females. Specifically, in older individuals higher negativity is associated with (a) decreased grey matter density in the bilateral hippocampus; (b) greater activation of the hippocampus and amygdala in response to unpleasant stimuli; and (c) stronger arousal in response to unpleasant stimuli

(Ingbretsen & Norris, 2019). These findings suggest that, over the life-course, highly negative but otherwise healthy individuals may experience structural and functional changes in the hippocampus and amygdala that could have implications for emotional memory and reactivity. Of course, given that these data are cross-sectional and not longitudinal, we cannot conclude causality. The relationships between trait negativity, increased reactivity to unpleasant stimuli, and the grey matter density of subcortical and limbic structures in older (but not younger) women deserve further attention.

Summary. Reviewing the literature on the negativity bias over the lifespan has revealed mixed results. Some studies seem to show a relatively early emergence of a negativity bias (as early as 6 months of age) and a negativity bias that is maintained through older age (as late as 65-80 years of age). Other studies contradict these results and have either shown no bias or a bias towards appetitive stimuli (e.g., helpers, positive images) early and late in life. In addition, there are methodological issues with both of these populations in terms of the criterion of matching; it is difficult to know whether appetitive and aversive stimuli are matched for nonverbal infants, and normative values used to match stimuli for older populations are often based on emotional ratings provided by young adults (e.g., Lang, Bradley & Cuthbert, 2008, 2015). Given that motivations and needs do change over the course of life, future studies would do well to consider these issues and examine the developmental trajectory of the negativity bias.

Conclusion

The current paper has summarized the theoretical arguments for the existence of a negativity bias in affective processes, as well as the considerations required to provide strong empirical support for the functioning of a negativity bias. Furthermore, evidence

from neuroscience methods (ERPs and fMRI), and from an individual differences approach supports the conclusions that: the brain preferentially responds to negative versus positive stimuli, starting at early attentional processes but biasing motor responses and evaluative categorization at later stages; there are stable, general individual differences in the negativity bias that predict behavior; females tend to have a stronger negativity bias than males, and this may be associated with prenatal testosterone exposure and/or the C/C genotype of 102T>C (rs6313) in the HTR2A gene, which is related to reduced post-synaptic serotonin receptor expression; and a negativity bias seems to function in infants as young as 3 months old, and continues to function in older age (although some research supports the emergence of a positivity effect in late adulthood), consistent with the hypothesized survival function of the negativity bias. A multilevel approach to investigating the negativity bias is critical for better understanding its underlying mechanisms and its consequences for affect dysregulation (cf. Gollan et al., 2016a, 2016b).

Author Note

I would like to acknowledge my collaborators on the research reviewed in this paper, including (alphabetically): Kristin (Wood) Flanary, Zachary Ingbretsen, JS Irick, and George Monteleone. In addition, I thank Sky Deswert for her help with conducting multiple literature searches on the negativity bias, threat bias, and other related topics for this review.

Accepted Manuscript

References

- Abdolmaleky, H. M., Faraone, S. V., Glatt, S. J., & Tsuang, M. T. (2004). Meta-analysis of association between the T102C polymorphism of the 5HT2a receptor gene and schizophrenia. *Schizophrenia Research*, 67(1), 53–62.
- Ashare, R. L., Norris, C. J., Wileyto, E. P., Cacioppo, J. T., & Strasser, A. A. (2013). Individual differences in positivity offset and negativity bias: Gender-specific associations with two serotonin receptor genes. *Personality and Individual Differences*, 55(5), 469-473.
- Bar-Haim, Y., Lamy, D., Pergamin, L., Bakersman-Kranenburg, M.J., & van IJzendoorn, M.H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, 133, 1–24.
- Bardeen, J. R., & Orcutt, H. K. (2011). Attentional control as a moderator of the relationship between posttraumatic stress symptoms and attentional threat bias. *Journal of Anxiety Disorders*, 25(8), 1008-1018.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323-370.
- Bradley, M. M. & Lang, P. J. (2007). The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. Technical report B-3. University of Florida, Gainesville, FL.
- Bradley, M.M. & Lang, P.J. (2017). Affective Norms for English Words (ANEW): Instruction manual and affective ratings. Technical Report C-3. Gainesville, FL:UF Center for the Study of Emotion and Attention.
- Britton, J. C., Bar-Haim, Y., Clementi, M. A., Sankin, L. S., Chen, G., Shechner, T., ... & Pine, D. S. (2013). Training-associated changes and stability of attention bias in youth:

- implications for attention bias modification treatment for pediatric anxiety. *Developmental Cognitive Neuroscience*, 4, 52-64.
- Brown, J. S. (1948). Gradients of approach and avoidance responses and their relation to level of motivation. *Journal of Comparative and Physiological Psychology*, 41, 450-465.
- Browning, M., Holmes, E. A., Murphy, S. E., Goodwin, G. M., & Harmer, C. J. (2010). Lateral prefrontal cortex mediates the cognitive modification of attentional bias. *Biological Psychiatry*, 67(10), 919-925.
- Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, 115(3), 401-423.
- Cacioppo, J. T., Crites Jr, S. L., Berntson, G. G., & Coles, M. (1993). If attitudes affect how stimuli are processed, should they not affect the event-related brain potential?. *Psychological Science*, 4(2), 108-112.
- Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review*, 1(1), 3-25.
- Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology*, 76(5), 839-855.
- Carstensen, L. L., & DeLiema, M. (2018). The positivity effect: a negativity bias in youth fades with age. *Current Opinion in Behavioral Sciences*, 19, 7-12.

- Carstensen, L. L., & Mikels, J. A. (2005). At the intersection of emotion and cognition: Aging and the positivity effect. *Current Directions in Psychological Science*, 14(3), 117-121.
- Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of Personality and Social Psychology*, 67(2), 319.
- Chae, J. J. K., & Song, H. J. (2018). Negativity bias in infants' expectations about agents' dispositions. *British Journal of Developmental Psychology*, 36(4), 620-633.
- Clark, V. P., & Hillyard, S. A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *Journal of Cognitive Neuroscience*, 8(5), 387-402.
- Epstein, R. A., Parker, W. E., & Feiler, A. M. (2008). Two kinds of fMRI repetition suppression? Evidence for dissociable neural mechanisms. *Journal of Neurophysiology*, 99, 2877-2886.
- Golimbet, V., Alifimova, M., & Mityushina, N. (2004). Polymorphism of the serotonin 2A receptor gene (5HTR2A) and personality traits. *Molecular Biology*, 38(3), 337-344.
- Gollan, J. K., Hoxha, D., Hunnicutt-Ferguson, K., Norris, C. J., Rosebrock, L., Sankin, L., & Cacioppo, J. (2016a). Twice the negativity bias and half the positivity offset: Evaluative responses to emotional information in depression. *Journal of Behavior Therapy and Experimental Psychiatry*, 52, 166-170.
- Gollan, J. K., Hoxha, D., Hunnicutt-Ferguson, K., Norris, C. J., Rosebrock, L., Sankin, L., & Cacioppo, J. (2016b). The negativity bias predicts response rate to Behavioral Activation for depression. *Journal of Behavior Therapy and Experimental Psychiatry*, 52, 171-178.

- Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, *13*(6), 923-929.
- Holbrook, A. L., Krosnick, J. A., Visser, P. S., Gardner, W. L., & Cacioppo, J. T. (2001). Attitudes toward presidential candidates and political parties: Initial optimism, inertial first impressions, and a focus on flaws. *American Journal of Political Science*, 930-950.
- Hsieh, S., & Yu, Y. T. (2003). Switching between simple response-sets: inferences from the lateralized readiness potential. *Cognitive Brain Research*, *17*(2), 228-237.
- Huang, Y. X., & Luo, Y. J. (2006). Temporal course of emotional negativity bias: an ERP study. *Neuroscience Letters*, *398*(1-2), 91-96.
- Ingbretsen, Z. A. & Norris, C. J. (2019). *Structure and Function of the Hippocampus and Amygdala Vary With Age and Individual Differences in Negativity*. Unpublished manuscript, Swarthmore College, Swarthmore, PA.
- 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*(4), 887.
- Jessen, F., Manka, C., Scheef, L., Granath, D. O., Schild, H. H., & Heun, R. (2002). Novelty detection and repetition suppression in a passive picture viewing task: A possible approach for the evaluation of neuropsychiatric disorders. *Human Brain Mapping*, *17*(4), 230-236.
- Jokela, M., Keltikangas-Jarvinen, L., Kivimaki, M., Puttonen, S., Elovainio, M., Rontu, R., et al. (2007). Serotonin receptor 2A gene and the influence of childhood maternal

- nurturance on adulthood depressive symptoms. *Archives of General Psychiatry*, 64(3), 356-360.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39(4), 341-350.
- Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *Neuroreport*, 14(18), 2317-2322.
- Kisley, M. A., Wood, S., & Burrows, C. L. (2007). Looking at the sunny side of life: Age-related change in an event-related potential measure of the negativity bias. *Psychological Science*, 18(9), 838-843.
- Klopper, F. J., & Madden, T. M. (1980). The middlemost choice on attitude items: Ambivalence, neutrality, or uncertainty?. *Personality and Social Psychology Bulletin*, 6(1), 97-101.
- 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.
- Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, 40(5), 776-785.
- Larsen, J. T., Norris, C. J., McGraw, A. P., Hawley, L. C., & Cacioppo, J. T. (2009). The evaluative space grid: A single-item measure of positivity and negativity. *Cognition and Emotion*, 23(3), 453-480.

- MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*(1), 15-20.
- Manning, J. T., Scutt, D., Wilson, J., & Lewis-Jones, D. I. (1998). The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Human Reproduction, 13*(11), 3000-3004.
- Mather, M., & Carstensen, L. L. (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in Cognitive Sciences, 9*(10), 496-502.
- McGraw, A. P., Larsen, J. T., Kahneman, D., & Schkade, D. (2010). Comparing gains and losses. *Psychological Science, 21*(10), 1438-1445.
- McIntyre, M. H. (2006). The use of digit ratios as markers for perinatal androgen action. *Reproductive Biology and Endocrinology, 4*(1), 10.
- Miller, N. E. (1959). Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation and social learning. In S. Koch (Ed.), *Psychology: A study of a science, Study 1* (pp. 198-292). New York: McGraw-Hill.
- Mogg, K., & Bradley, B. P. (1999). Orienting of attention to threatening facial expressions presented under conditions of restricted awareness. *Cognition & Emotion, 13*(6), 713-740.
- Neta, M., Cantelon, J., Haga, Z., Mahoney, C. R., Taylor, H. A., & Davis, F. C. (2017). The impact of uncertain threat on affective bias: Individual differences in response to ambiguity. *Emotion, 17*(8), 1137.
- Neta, M., Davis, F. C., & Whalen, P. J. (2011). Valence resolution of ambiguous facial expressions using an emotional oddball task. *Emotion, 11*(6), 1425.

- Neta, M., Norris, C. J., & Whalen, P. J. (2009). Corrugator muscle responses are associated with individual differences in positivity-negativity bias. *Emotion, 9*(5), 640.
- Neta, M., Tong, T. T., Rosen, M. L., Enersen, A., Kim, M. J., & Dodd, M. D. (2017). All in the first glance: first fixation predicts individual differences in valence bias. *Cognition and Emotion, 31*(4), 772-780.
- Neta, M., & Whalen, P. J. (2010). The primacy of negative interpretations when resolving the valence of ambiguous facial expressions. *Psychological Science, 21*(7), 901-907.
- Norris, C. J. (2004). *Exploring the Negativity Bias: A Social Neuroscience Approach*. ProQuest Dissertation.
- Norris, C. J. (2019). *Prenatal Testosterone Levels Influence Emotional Biases*. Unpublished manuscript, Swarthmore College, Swarthmore, PA.
- Norris, C. J. & Flanary, K. W. (2019). *The Simple Things in Life: Age Differences in Emotional Biases*. Unpublished manuscript, Swarthmore College, Swarthmore, PA.
- Norris, C. J., Gollan, J., Berntson, G. G., & Cacioppo, J. T. (2010). The current status of research on the structure of evaluative space. *Biological Psychology, 84*(3), 422-436.
- Norris, C. J., Larsen, J. T., Crawford, L. E., & Cacioppo, J. T. (2011). Better (or worse) for some than others: Individual differences in the positivity offset and negativity bias. *Journal of Research in Personality, 45*(1), 100-111.
- Norris, C. J., Leaf, P. T., & Fenn, K. M. (2018). Negativity bias in false memory: moderation by neuroticism after a delay. *Cognition and Emotion, 1-17*.
- Norris, C. J., Monteleone, G., & Cacioppo, J. T. (2019). *Once Again, With Feeling: Pleasant Scenes Produce Repetition Suppression Whereas Unpleasant Scenes Produce*

- Repetition Enhancement*. Unpublished manuscript, Swarthmore College, Swarthmore, PA.
- 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*, 80(3), 381-396.
- O'Toole, L., & Dennis, T. A. (2012). Attention training and the threat bias: An ERP study. *Brain and cognition*, 78(1), 63-73.
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4), 296-320.
- Salimpoor, V. N., Chang, C., & Menon, V. (2010). Neural basis of repetition priming during mathematical cognition: repetition suppression or repetition enhancement?. *Journal of Cognitive Neuroscience*, 22(4), 790-805.
- Sapienza, P., Zingales, L., & Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences*, 106(36), 15268-15273.
- Smith, N. K., Cacioppo, J. T., Larsen, J. T., & Chartrand, T. L. (2003). May I have your attention, please: Electrocortical responses to positive and negative stimuli. *Neuropsychologia*, 41(2), 171-183.
- Sylvester, C. M., Hudziak, J. J., Gaffrey, M. S., Barch, D. M., & Luby, J. L. (2016). Stimulus-driven attention, threat bias, and sad bias in youth with a history of an anxiety disorder or depression. *Journal of Abnormal Child Psychology*, 44(2), 219-231.
- Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. *Psychological Bulletin*, 110(1), 67-85.

- Tottenham, N., Phuong, J., Flannery, J., Gabard-Durnam, L., & Goff, B. (2013). A negativity bias for ambiguous facial-expression valence during childhood: Converging evidence from behavior and facial corrugator muscle responses. *Emotion, 13*(1), 92.
- Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: the negativity bias in social-emotional development. *Psychological Bulletin, 134*(3), 383.
- Vaquero-Lorenzo, C., Baca-Garcia, E., Diaz-Hernandez, M., Perez-Rodriguez, M. M., Fernandez-Navarro, P., Giner, L., et al. (2008). Association study of two polymorphisms of the serotonin-2A receptor gene and suicide attempts. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 147B*(5), 645–649.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063.
- Wood, S., & Kisley, M. A. (2006). The negativity bias is eliminated in older adults: Age-related reduction in event-related brain potentials associated with evaluative categorization. *Psychology and Aging, 21*(4), 815.

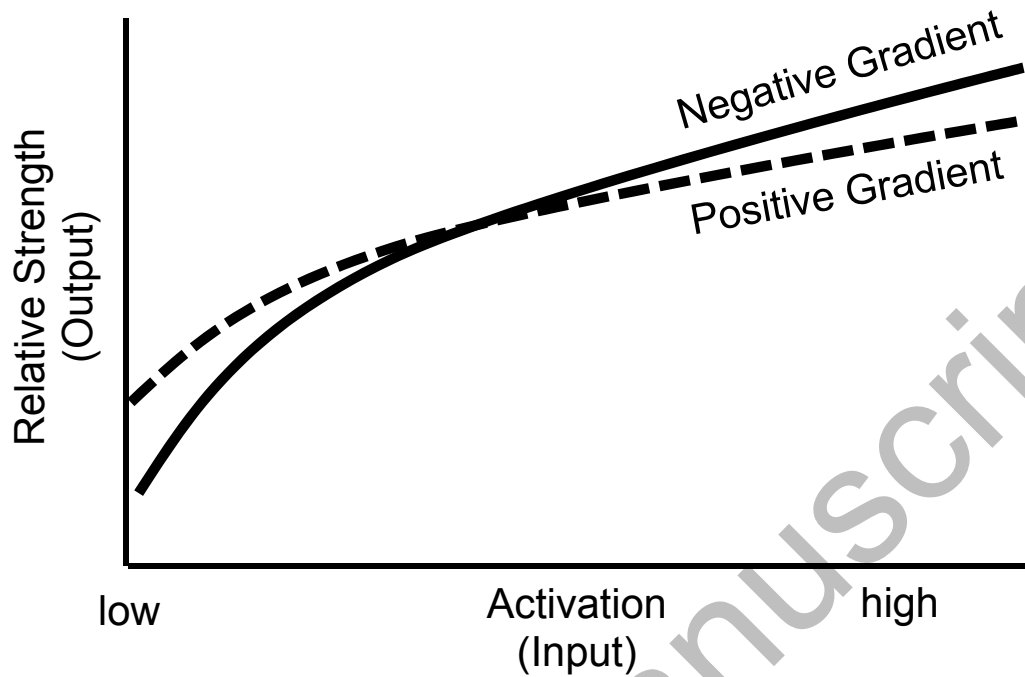


Figure 1. Activation functions for positive and negative dimensions of affective processing; the x -axis represents affective input, whereas the y -axis represents output of the system. The ESM proposes that there are two asymmetries in affective processing: the positivity offset is the result of greater positive than negative affect at low levels of emotional input; the negativity bias is the result of stronger responses to negative than to equally extreme positive input. Adapted from Cacioppo and Berntson (1994).

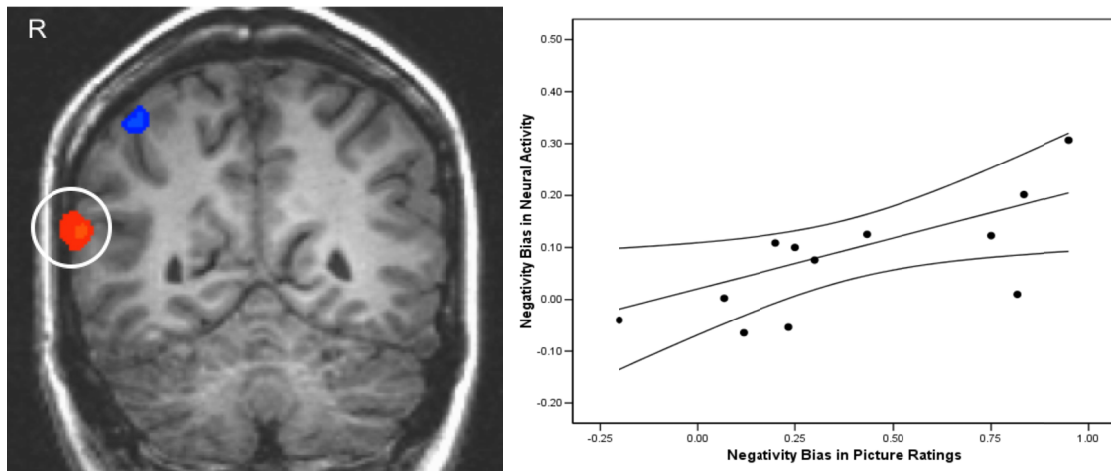


Figure 2. A negativity bias in neural activation (left) correlates with individual differences in the negativity bias in ratings (right). A whole brain contrast between negative images and positive images revealed a cluster of activation in right BA 19, visual association cortex, that indicated a negativity bias (Norris, 2004; Study 1). The bias in neural activation was correlated with the negativity bias in ratings, $r(11) = .72$, $p < .01$, $R^2 = .52$. Both findings were replicated in Study 2.