The expectancy bias

Proulx, Travis; Sleegers, W.W.A.; Tritt, Shona

Link to publication

The expectancy bias: Expectancy-violating faces evoke earlier pupillary dilation than neutral or negative faces

Travis Proulx [a,*], Willem Sleegers [b], Shona M. Tritt [c]

[a] School of Psychology, Cardiff University, United Kingdom
[b] Tilburg University, Netherlands
[c] New York University, United States

HIGHLIGHTS

- The negativity bias describes the preferential processing of negative stimuli relative to positive or neutral stimuli.
- We propose a superordinate *expectancy bias* for preferentially processing expectancy-violating stimuli.
- We measured pupillary dilation to assess attentional arousal to expectancy-violating, threatening and neutral faces.
- Expectancy-violating faces aroused earlier pupillary dilation, with Thatcherized faces arousing the earliest.
- These findings have implications for attentional bias and motivations to address incongruity relative to identifying threat.

ARTICLE INFO

ABSTRACT

Humans maintain a negativity bias, whereby they perceive threatening stimuli to be more salient than rewarding or neutral stimuli. Across 6 within-subject experimental comparisons, we tested the hypothesis that humans maintain an even stronger expectancy bias, preferentially processing stimuli that violate mental representations of expected associations. To assess this bias, we measured variations in pupillary dilation as a means of determining attentional arousal in response to neutral, negative and expectancy-violating versions of the same social stimuli: human faces. We conducted three baseline manipulation checks that directly compared neutral faces with threatening (angry) and expectancy-violating (upside-down and Thatcherized) faces, and three bias comparisons that directly compared threatening and expectancy-violating faces with one another. Across these experiments, we found evidence for a dominant expectancy bias in pupillary arousal for social stimuli, whereby expectancy-violating faces produced pupillary dilation earlier than neutral and threatening faces, with Thatcherized faces producing the greatest magnitude of dilation.

© 2016 Published by Elsevier Inc.

1. Introduction

"Bad is stronger than good" (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001), insofar as negative stimuli are processed preferentially relative to positive or neutral stimuli (Rozin & Royzman, 2001). This negativity bias is easy to understand from an adaptive perspective: overlook a potential reward, and new opportunities may arise; overlook a potential threat, and it may harken the end of opportunity. Nevertheless, there is a mode of stimuli that may trump the heightened salience of negativity: expectancy-violating stimuli, which are inconsistent with the expected associations that model our experiences.

In order to act effectively on our environment, we must understand our environment, which requires an adaptive mental model that assigns a construal of reward or threat (Peterson, 1999). Expectancy-violating stimuli defy this construal, and thereby thwart the assignment of an unambiguous motivational impetus. We are initially conflicted as to whether we should approach or avoid the source of the unexpected, and the aversive sense of anxious uncertainty (or dissonance; Festinger, 1957, or disequilibrium; Piaget, 1937) that often arises is a warning to address the motivational conflict quickly, by means of rapid attention and action (Gray & McNaughton, 2003; Harmon-Jones, Amodio, & Harmon-Jones, 2009). While acting prior to an unambiguous assessment of threat may come at potential cost, failing to act at all may come at an even greater cost.

Following from this account, we argue that expectancy-violation constitutes the most initially salient class of motivational stimuli,

* Corresponding author at: School of Psychology, Cardiff University, Tower Building, 70 Park Place, CF10 3AT Cardiff, United Kingdom.
E-mail address: saptp7@cardiff.ac (T. Proulx).

producing arousal indicative of increased attentional resources earlier than other salient experiences. To test this hypothesis, we measured variations in pupillary dilation (Bradley, Miccoli, Escrig, & Lang, 2008; Sleegers, Proulx, & van Beest, 2015) as a means of assessing arousal indicative of preferential attention to differentially salient social stimuli: human faces. We conducted three baseline manipulation checks that directly compared neutral faces with threatening (angry) and expectancy-violating (upside-down and "Thatcherized"; Thompson, 1980) faces, and three bias comparisons that directly compared angry, upside-down and Thatcherized faces with one another. Across these experiments, we expected evidence for a dominant expectancy bias, whereby expectancy-violating faces would evoke earlier attentional arousal (PD) than neutral and threatening faces.

2. Attentional salience of negativity

Human consciousness is bombarded by internally and externally generated stimuli—an array of information that would become overwhelming if not for attentional filters that render some stimuli inherently salient. This salience is determined by neuro-attentional mechanisms involving large-scale networks in the parietal and frontal cortices that modulate sensory processing to selectively enhance the visual representation of relevant stimuli and suppress irrelevant information (Kastner & Ungerleider, 2001). Much of the research on the allocation of attentional resources has focused on those qualities of stimuli that render them motivationally salient, more generally. To date, the predominant focus has been on the negativity bias, often understood as a propensity to attend (e.g., Cacioppo, Gardner, & Berntson, 1999; Carretie, Martin-Loeches, Hinojosa, & Mercado, 2001) and respond (Baumeister et al., 2001) more rapidly to threatening stimuli (i.e., punishments, e.g., Rozin & Royzman, 2001 or obstacles to reward, e.g., Blascovich, 2000) than to rewarding or neutral stimuli.

While the preponderance of this research demonstrates that negatively valenced stimuli are more salient than positive or neutral information, some work suggests that positive stimuli may be equally salient if it is of equivalent magnitude (e.g., Foti, Hajcak, & Dien, 2009; Schupp, Junghofer, Weike, & Hamm, 2003; Schupp, Junghofer, Weike, & Hamm, 2004; see Weinberg & Hajcak, 2010). Taken together, these findings suggest that it is cues to the overall motivational significance of stimuli that primarily determines the attentional salience of a given stimulus (in this case, cues to threat or reward), though it is likely that negatively valenced stimuli tend to be more salient overall (e.g., see Baumeister et al., 2001). However, a category of motivationally salient stimuli that has yet to be directly compared to negativity is expectancy-violation.

3. Attentional salience of expectancy violation

Humans take in sensory stimuli, store these experiences in memory and construct mental models that predict and interpret subsequent experiences. These mental models are a fundamental prerequisite for functioning in our changing environment, allowing us to understand experiences that range from naive physics to social interactions to our own internal states. Depending on the kind of experience organized by a given model, these representations have been called perceptual paradigms (Bruner, 1990), assumptive worlds (Janoff-Bulman, 1992) or schemata (Piaget, 2000), and to the extent that the mental model implies an a priori guide for action: meaning frameworks (Peterson, 1999; Proulx & Inzlicht, 2012). Inherent in any of these representations is a motivational significance that categorizes the stimuli as a potential reward or threat, which in turn cues the emotional valence of subsequent arousal (Hirsh, Mar, & Peterson, 2012).

Underlying all of these modes of representation is a common construct: expectation. We expect our experiences to conform to our representations, whether this involves an expectation that crows are black (Piaget, 1937), the world is just (Lerner, 1980) or that threatening

stimuli will cause harm. These models are mental representations of expected associations, and experiences that are inconsistent with these representations evoke an initial attentional salience and a proximal motivation to react to the discrepancy (Proulx, Inzlicht, & Harmon-Jones, 2012; see also Peterson, 1999).

Most often, stimuli violate our expectations insofar as they are different from what we have experienced in the past. These stimuli could be novel relative to a lifetime of prior experiences (i.e., conceptual novelty [Kagan, 2009], e.g., a white crow [Piaget, 1937], or a red Ace of Spades [Bruner & Postman, 1949]) or novel relative to a recently learned, arbitrary pattern (i.e., stimuli novelty [Kagan, 2009] e.g., a 'Y' preceded by a series of 'X's in an oddball paradigm [Dunchan-Johnson & Donchin, 1977]). In the face of these experiences, the absence of an initially applicable mental model makes us unsure as to whether the stimuli represent a reward we should approach, or a threat we should avoid (Peterson, 1999)—a state of motivational ambivalence that must be quickly resolved if a state of potentially paralyzing goal conflict is to be avoided (Gray & McNaughton, 2003).

This conflict may be particularly acute for stimuli that is not merely different from what we have experienced in the past, but which also activates two incongruous representations simultaneously, and as such, constitute neither (e.g., uncanniness following experiences that are both familiar and unfamiliar [Freud, 1990], or a duck/rabbit gestalt illusion [Köhler, 1929]). In an experimental setting, inherently incongruous experiences may be operationalized as conceptually inconsistent stimuli, such as the word "green" coloured blue (Stroop, 1992). Alternatively, it may involve stimuli that arouse attitudinal ambivalence, insofar as the experience simultaneously activates threatening and rewarding associations (Miller, 1944), or ambiguous evaluative feedback that forces us to maintain multiple outcomes (e.g., encountering threatening and rewarding associations [Holroyd, Hajcak, & Larsen, 2006]).

4. Neuroaffective outcomes

More recently, a theoretical perspective has emerged which frames these and other experiences in terms of prediction error (Montague, Dayan, & Sejnowski, 1996), where many of the cognitive conflict phenomena reported in cognitive, social, developmental and clinical psychology are hypothesized to be instantiations of expectancy violation (Jonas et al., 2014; Proulx et al., 2012). Specifically, phenomena associated with the aversive arousal state that follows from the violation of expected associations described in these literatures, whether it is the disequilibrium that follows from the violation of a developmental schema (Piaget, 2000), the dissonance that follows from the violation of a committed belief (Harmon-Jones et al., 2009), the uncertainty that follows from the violation of a mental representation (Van den Bos, 2001) or the anxiety we experience when our assumptive worlds fail to account for our experiences (Janoff-Bulman, 1992)—a state that we summarize as disanxiouuncertlibrium (Proulx & Inzlicht, 2012). From this perspective, all of these terms designate the same neuroaffective syndrome of brain activation and sympathetic nervous system arousal that follows from any violation of expectation, whether it represents a "high"- or "low"-level prediction error.

This syndrome of arousal is initiated by the orienting response (Courchesne, Hillyard, & Galambos, 1975; Kok, 1997; see Peterson, 1999), reflected in an ERP amplitude that is understood as the organism's initial reaction to changes in its environment (Sokolov, 2002). This response cues to any expectancy violation, whether the source is mere novelty (e.g., oddball paradigm, Johnson, 1988) or an incongruous stimulus (e.g., Stroop task, Liotti, Woldorff, Perez, & Mayberg, 2000). Following the orienting response, tonic inhibition of the reticular formation by hippocampal CA3 neurons is released and a lower brain circuit including the amygdala is disinhibited, which activates circuitry in the right hemisphere (Tucker & Frederick, 1989) and subsequently inhibits the prefrontal left cortical hemisphere that is associated with approach motivation (see Harmon-Jones & Harmon-Jones, 2011;

Harmon-Jones et al., 2009). This right-brain asymmetry is associated with activation of the behavioral inhibition system, more generally, and the release of cortical noradrenaline that underlies heightened attentional vigilance and scanning of the environment. Subsequently, there is a psychological threat response—a chain reaction of sympathetic nervous system arousal that involves cortisol and changes in cardiac activity (Blascovich, 2000; Fowles, 1980).

To the extent that a given stimulus violates activated expected associations, the arousal that is initiated by the orienting response plays a role in how the valence of subsequent stimuli is interpreted. Specifically, the emotional valence of threatening or rewarding stimuli is experienced as more intense if it is associated with expectancy-violating experiences. For example, in a meta-analysis of laboratory-based studies, Dickerson and Kemeny (2004) found that threats grounded in uncertain outcomes increase cortisol levels more dramatically than any other stressor. Similarly, Grupe and Nitschke (2011) in a series of skin conductance and mood experiments, discovered that uncertainty amplified the impact of negative events. In terms of positive experiences, Wilson and colleagues (Wilson, Centerbar, Kermer, & Gilbert, 2005) found that uncertainty about a positive event prolonged the pleasure that it caused. Additionally, animals have been found to eat more if their food was encountered in an unexpected location than if it was found in an expected location (Roitman, van Dijk, Thiele, & Bernstein, 2001).

5. Negativity vs. expectancy violation

People attend and react more strongly to unexpected stimuli compared to familiar/expected information. However, few studies have directly compared the effect of expectancy violation to other types of attentionally salient information. One exception is ERP work that has demonstrated greater ERN amplitudes in response to ambiguous compared to negative feedback. Holroyd et al. (2006) established, for example, that ambiguous feedback elicited heightened error-related negativity to an equivalent magnitude as negative feedback. Hirsh and Inzlicht (2008) further investigated individual differences in response to ambiguous vs. negative feedback (also see Gu, Ge, & Huang, 2010) and found that some individuals react more strongly to ambiguous feedback than failure feedback. However, no study has directly compared negatively valenced and expectancy-violating versions of the same stimuli with the expectation of a superordinate expectancy bias: a preferential processing of expectancy-violating stimuli relative to negative stimuli. Specifically, no study has compared these classes of salience in a manner that can distinguish negative and expectancy-violating versions of the same stimuli, both in terms of processing amplitude, the time course of arousal and the allocation of attentional resources. To assess these relevant dimensions, we will measure pupillary dilation, a neuroaffective marker that varies in size with the magnitude of attentional salience, and has been shown to index the differential time course of attentional arousal for both negative (e.g., Bradley et al., 2008) and expectancy-violating (Sleegers et al., 2015) stimuli.

6. Pupillary dilation as a measure of attentional salience and arousal

Pupillary reactivity (i.e., changes in pupil size) is a marker of preferential processing, and can serve as an index of both neurocognitive arousal in reaction to the attentional saliency of a stimulus, and an index of subsequent sympathetic nervous system arousal corresponding to the valence of the stimuli. Initially, pupillary dilation (PD) increases when salient stimuli activate the locus coeruleus-norepinephrine system (LC-NE), a neurocognitive substrate of the Behavioral Inhibition System (Amodio, Master, Yee, & Taylor, 2008; Aston-Jones & Cohen, 2005). LC-NE activation plays a fundamental role in engagement or withdrawal from a task by means of norepinephrine (NE) that is released by the locus coeruleus (LC) through projections in the forebrain (Aston-Jones & Cohen, 2005). In turn, this NE

directly modulates dilation of the pupil, such that PD correlates with LC activity in humans (Gilzenrat et al., 2003) and other primates (Raikowski, Kubiak, & Aston-Jones, 1993), and is consistent with LC phasic response to a variety of working memory concentration tasks, such as judgments of similarity (Beatty, 1982), oddness/evenness of recalled numbers (Smallwood et al., 2011) or relative probabilities (Jepma & Nieuwenhuis, 2011).

Of central relevance to our main hypothesis, concentration task errors evoke especially heightened PD (e.g., Smallwood et al., 2011), whereby experiences associated with surprise (i.e., prediction error) more generally increase PD due to the rapid release of NE (Preuschoff, Hart, & Einhauser, 2011). Subsequently, pupillary variability co-varies with measures of skin conductance and cardiac activity, suggesting that it also represents an additional component of downstream sympathetic nervous system arousal (Bradley et al., 2008). As such, PD modulations in task engagement reflect and differentiate arousal produced by the salient aspects of the relevant stimuli, whether they represent expectancy violation or emotional valence.

The pupillary manifestation of arousal associated with expectation and emotional valence has previously been explored in separate studies that correspond to these distinct time intervals. Expectancy violation PD has been primarily observed initially following the pupillary light reflex calibration to a presented stimulus, represented as an initial spike in PD, with the peak appearing between 500 ms–1500 ms following stimulus onset. This initial spike in PD is consistently evoked by a variety of expectancy-violating task stimuli, such as incongruent Stroop task trials (e.g., Laeng, Ørbo, Holmlund, & Miozzo, 2011; Rondeel, Van Steenbergen, Holland, & van Knippenberg, 2015), inconsistent Simon task trials (Van Steenbergen & Band, 2013) and incongruent playing card features—even when people have no conscious awareness of the incongruity (Sleegers et al., 2015). In contrast, emotional valence PD has been primarily observed at longer intervals following stimulus onset, represented as a gradually heightened, then sustained dilation following 1500 ms. This different and temporally latent pattern of PD is consistently evoked by a variety of negatively and positively valenced stimuli, for example, with unpleasant and pleasant photos (Bradley et al., 2008; Lang, Bradley, & Cuthbert, 2005) and sounds (Partala & Surakka, 2003).

Taken together, these phases of PD may correspond to two distinct time courses of preferential processing, whereby initial PD (500 ms–1500 ms) is evoked by the comparison of stimuli to relevant expected associations, and subsequent PD (1500 ms+) is a response to the assessment of emotional valence. Stimuli that violate expectations appear to cause the LC to release NE, initiating an initial PD spike that indexes heightened attentional arousal. The valence of stimuli may be determined subsequently, potentially resulting in heightened and sustained pupillary arousal indicative of sympathetic nervous system arousal whether the stimulus is negative (potential threat) or positive (potential reward). To date, no study has directly assessed such preferential processing by comparing the time course of PD in response to negatively-valenced and expectancy-violating versions of the same social stimuli.

7. Threatening and expectancy-violating faces

The stimuli we chose for this comparison are photos of human faces, insofar as they can discretely project salient threat (angry expressions), and can be altered in ways to discretely evoke salient expectancy violations (upside-down orientation and Thatcherized features, see Fig. 1).

Angry faces have been commonly used to assess a general negativity bias (Hansen & Hansen, 1988; Öhman, et al., 2001). Specifically, angry faces have been used as a means of evoking threat (Schutter, de Haan, & van Honk, 2004; Van der Schalk et al., 2011) without also constituting a violation of expectation (e.g., see Whalen & Taylor, 2014), and to the same extent as other especially salient threats (e.g., snakes; Öhman, 2009). While angry faces may be less common in everyday experience,
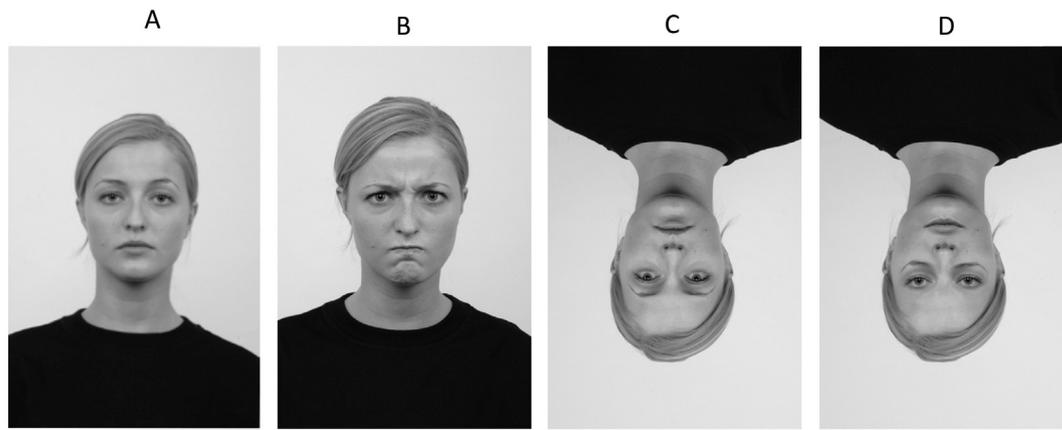
Fig. 1. An example of A) neutral, B) angry, C) upside-down and D) Thatcherized faces.

participants have presumably encountered and mentally mapped innumerable angry faces prior to entering the lab, where they may be especially expected in the initial context of an emotion recognition task (see Methods). As such, we expected angry faces to evoke the same pattern of PD produced by other negatively-valenced stimuli (gradually heightened, then sustained increase following 1500 ms), while showing no evidence of the pattern of PD produced by expectancy-violating stimuli (initial spike between 500 and 1500 ms).

To robustly examine the relative effect of expectancy-violating faces, we operationalized expectancy violation in a face-valid manner with two versions of expectancy-violating faces. The first of these faces presented a neutral expression on a face that was rotated 180°. These upside-down faces are novel, insofar as they represent an orientation seldom encountered in natural experience (Kagan, 2009), which may be particularly unexpected in the context of an emotion recognition task. Moreover, the emotion conveyed by the face conveys no threat, with nothing inherently negative about the inverted image. As such, we expected upside-down faces to evoke the same pattern of PD produced by other expectancy-violating stimuli (initial spike between 500 and 1500 ms), while showing no evidence of the pattern of PD produced by negative stimuli (gradually heightened, then sustained increase following 1500 ms).

The second of these expectancy-violating faces also presented a neutral expression on a face that was rotated 180°. However, in addition to the novel upside-down orientation, these faces contained an inherent incongruity, insofar as the eyes and mouth were oriented right-side-up. This mis-orientation represents a classic "Thatcher illusion" (Thompson, 1980), whereby individuals consciously reorient the facial features so that they appear congruent with the inverted face, even as they unconsciously register the incongruity (e.g., Carbon, Schweinberger, Kaufmann, & Leder, 2005). As with the merely upside-down faces, the Thatcherized faces should evoke the same pattern of PD produced by other expectancy-violating stimuli (initial spike between 500 and 1500 ms)—a PD pattern also observed for other expectancy–violating stimuli whose novel features are not consciously detected (e.g., reverse-coloured playing cards, Sleegers et al., 2015). As with the upside-down faces, the absence of inherently negative elements should preclude the pattern of PD produced by negative stimuli (gradually heightened, then sustained increase following 1500 ms). However, given the presence of both novel and inherently incongruent features, we expected that Thatcherized faces would be experienced as expectancy–violating to a greater extent, evoking even larger makers of attentional arousal relative to merely upside-down faces.

8. Study overview

Over 6 experimental comparisons, we tested for the presence of a dominant expectancy bias in the processing of differentially salient stimuli. Unlike other measures of attentional salience (e.g., late positive potential; Codispoti, Ferrari, & Bradley, 2007), we were unsure whether PD would differentiate multiple variations of a single class of stimuli (i.e., faces) in a single within-subjects setting. To ensure that we could clearly assess and interpret the comparisons relevant to our hypotheses, we conducted six separate within-subject experiments, allowing us the clearest means of directly contrasting pupillary responses to neutral, negative and expectancy-violating faces in direct comparison with one another, without the possibility of interfering assimilation/contrast effects (Charness, Gneezy, & Kuhn, 2012; Greenwald, 1976).

The first 3 of these experiments were baseline manipulation checks that would validate our stimuli as representations of threat and expectancy-violation, insofar as they should evoke the respective and differential patterns of arousal previously reported as associated with these stimuli. These experiments presented each of the attentionally salient faces (angry, upside-down and Thatcherized) and baseline neutral faces in direct paired comparisons. If our stimuli were operationally valid, angry, upside-down and Thatcherized faces would all evoke an attentional bias relative neutral faces. In the case of angry faces, this bias would manifest as gradually heightened, then sustained PD during the emotional valence time period (1500 + ms) (H1a). In the case of upside-down and Thatcherized faces, this bias would manifest as an initial spike primarily during the time period that PD responds to expectancy violation (500 ms–1500 ms) (H1b).

The next 3 experiments were bias comparisons that directly tested our expectancy bias hypothesis. These experiments presented each of the attentionally salient faces (angry, upside-down and Thatcherized) with one another in direct paired comparisons. If expectancy-violating faces evoke an attentional bias relative to threatening faces, then both of the expectancy-violating faces (upside-down and Thatcherized) would be processed preferentially relative to angry faces (H2a). This would manifest as heightened PD during the initial expectancy violation time period (500 ms–1500 ms) relative to the angry faces. If the Thatcherized faces, being both novel (upside-down) and inherently incongruent (right-side-up features) would be the most expectancy-violating, then they would evoke greater attentional bias even relative to the up-side-down faces (H2b), manifesting as a comparatively heightened PD during the initial expectancy violation time period (500 ms–1500 ms).

9. Method

9.1. Participants

237 total participants (76 males, 161 females) were recruited using the Tilburg University signup system. They participated either for course credit or €5. Their average age was 20.78 (min: 18, max: 32) and consisted largely of Tilburg University students. Participants were

randomly assigned to one of 6 within-subject experiments. A minimum sample size of 25 participants per experiment was determined by a power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), assuming a medium to large effect size as reported in prior research assessing PD in response to cognitive conflict in a within-subjects, repeated measures, paired comparisons design (e.g., Van Steenbergen & Band, 2013). Random assignment continued to the conclusion of the academic term, resulting in unequal Ns across conditions. Collection was not continued after data analysis.

9.2. Materials

All materials and measures are reported and are available upon request. All face images were taken from the Radboud Faces Database (Langner et al., 2010). The Radboud Faces Database (RaFD) is a set of pictures of 67 models (including Caucasian males and females, Caucasian children, both boys and girls, and Moroccan Dutch males) displaying 8 emotional expressions. Our neutral, up-side-down and Thatcherized faces were selected from Caucasian adult male and female database faces with neutral expressions. Neutral faces were correctly oriented. Upside-down faces were rendered expectancy-violating by presenting them rotated 180°. Thatcherized faces were rendered more expectancy-violating by altering them with image editing software (Pixelmator; Pixelmator Team, 2016), such that they incongruously represent a face that is upside-down, with eyes and mouth which are right-side-up. For the threatening faces, we selected Caucasian adult males and females with angry emotional expressions.

9.2.1. Faces task

Each faces task consisted of two types of faces being presented to the participant, with probe trials interspersed throughout the task at which point the participant had to indicate the negative or positive valence of the previous facial expression. Across all experiments, negative vs. positive was chosen as a fast, intuitive valence judgment relative to negative vs. neutral, given that neutrally valenced faces are generally perceived as negative (e.g., Lee et al., 2008). In addition to ensuring that participants maintained attention to the task, the probe trials served as a manipulation check to ensure that threatening faces were experienced as negative relative to the neutral, upside-down and Thatcherized faces, and to ensure that neutral, upside-down and Thatcherized faces were not experienced as negative relative to one another.

In addition to establishing distinct pupillary responses for threatening and expectancy–violating faces in the baseline condition, these judgments would provide further evidence that any evidence for an expectancy bias could not be taken as additional evidence for an emotionally-valenced negativity bias. Within each condition we counted the total number of probe trials for each kind of face and calculated what percentage of those trials each type of face was rated as negative. This allowed us to analyze each face-type pairing in terms of how often it was seen as more negative than positive, by means of paired t-tests.

Each face was 550 × 827 pixels in size and presented at the center of the screen for a duration of 5000 ms. Probe trials consisted of a question mark and required a response from the participant—a "1" to indicate that the previous face had a negative valence and a "2" for a positive valence. Before each presentation, a fixation cross was visible at the center of the screen. Participants had to remain focused at the fixation cross for a duration of 1000 ms for a stimulus to appear. The task consisted of 10 practice trials and two blocks of experimental trials. Each block consisted of 100 trials: 39 of each type of face, and 22 probe trials.

9.2.2. Pupillary dilation recording and processing

Pupil dilation was measured using the Tobii T60 eye tracker (Tobii Technology, Inc., Sweden) and E-Prime Professional 2.0 software (Psychology Software Tools, Pittsburgh, PA). The Tobii T60 eye tracker is integrated into a 17-inch TFT monitor and records a variety of pupil

characteristics, including pupil size, through screen based eye tracking at a rate of 60 Hz for each eye separately. Each measurement has a validity indication that ranges from 0 (the system is certain that all data belongs to the particular eye) to 4 (gaze data is missing or incorrect). Only recordings with a validity score of 0 were used. Pupil size from each eye were averaged together to create a single pupil size score and filtered with a modified repeated median filter (outer width: 25, inner width 15) using the robfilter package (Fried, Schettlinger, & Borowski, 2014) in R (R Core Team, 2016). Artifacts such as blinks were corrected using linear interpolation from the 'zoo' package in R (Zeileis, Grothendieck, Ryan, Andrews, & Zeileis, 2015). Baseline differences in pupil size were controlled for by subtracting the average pupil size during a 500 ms pre-trial period from the subsequent pupil measurements.

10. Results

To investigate the impact of the different types of faces on pupil diameter, we conducted six separate experimental comparisons, each one assessed with a 2 (Face type: [Baseline Manipulation Check] angry vs. neutral, upside-down vs. neutral, Thatcherized vs. neutral, [Bias Comparisons] upside-down vs. angry, Thatcherized vs. upside-down, Thatcherized vs. angry) × 19 (Time: 500 to 5000 s, in 250 ms bins) repeated-measures GLM analyses. The first 500 ms after stimulus onset was regarded as the light reflex period and was therefore not included in the analysis (Prehn et al., 2008). For each comparison, we tested the main effects for face type, time, and the interaction of face type and time on pupil diameter. We also separately tested the mean differences of pupil diameter at the time periods primarily associated with cognitive conflict (500 ms–1500 ms) and sympathetic nervous system arousal (1500 + ms). Sphericity assumptions were checked with the Mauchly's test of sphericity. In case of a violation of this assumption, we report Huynh-Feldt corrected statistics.

11. Baseline manipulation checks: angry, upside-down and Thatcherized vs. neutral

11.1. Angry vs. neutral faces

11.1.1. Negativity manipulation check

Angry faces were rated as more negative than neutral faces, $t(40) = 16.38, p < 0.001, dz = 2.56$. Angry faces were rated as negative 96.46% of the time, while neutral faces were rated as negative only 28.53% of the time.

11.1.2. Pupillary dilation

The comparison between angry and neutral faces revealed a significant main effect of face type, $F(1, 40) = 33.24, p < 0.001, \eta^2_p = 0.454$, a main effect of time, $F(1.85, 123.72) = 26.40, p < 0.001, \eta^2_p = 0.398$, as well as a significant interaction of face type and time, $F(3.18, 127.38) = 11.88, p < 0.001, \eta^2_p = 0.229$; see Fig. 2. During both the 500–1500 ms period and the 1500 + ms period there is a greater average pupil size for angry faces compared to neutral faces. In support of H1a, this difference is larger after 1500 + ms (angry faces: $M = 0.11$, $SD = 0.07$; neutral faces: $M = 0.07, SD = 0.08$), $t(40) = 5.82$, $p < 0.001, dz = 0.91$), compared to the initial period (angry faces: $M = 0.06, SD = 0.06$; neutral faces: $M = 0.04, SD = 0.06, t(40) = 3.96, p < 0.001, dz = 0.62$).

11.2. Upside-down vs. neutral faces

11.2.1. Negativity manipulation check

No difference was found in valence between upside-down and neutral faces, $t(40) = 1.91, p = 0.064, dz = 0.30$. Upside-down faces were rated as negative 52.98% of the time and neutral faces were rated as negative 57.46% of the time.
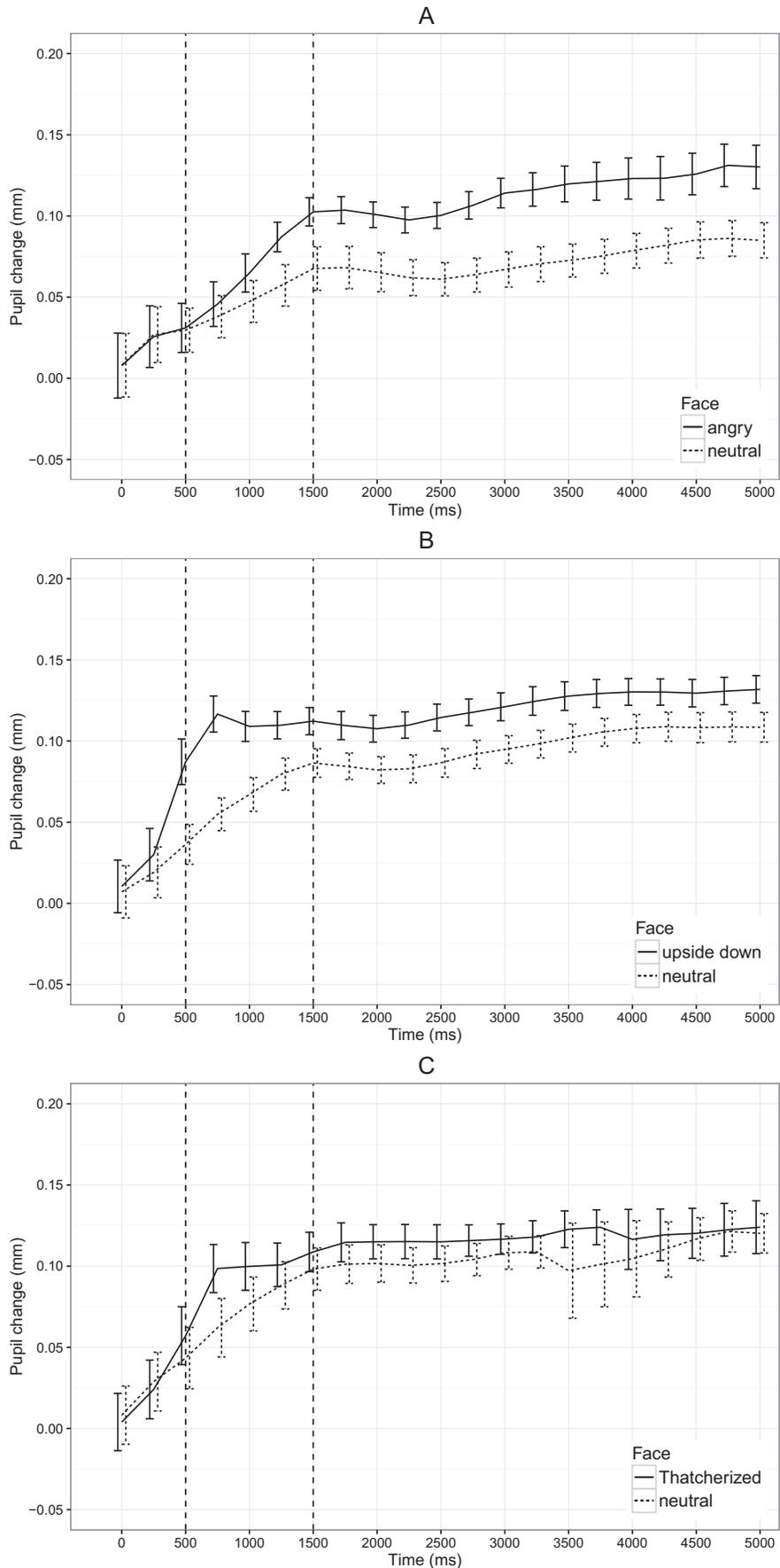
Fig. 2. Baseline manipulation checks of pupillary dilation in response to A) angry vs. neutral faces B) upside-down vs. neutral faces C) Thatcherized vs. neutral faces.

11.2.2. Pupillary dilation

Comparing upside-down faces to neutral faces, we found a significant main effect of face type, $F(1, 40) = 20.13$, $p < 0.001$, $\eta^2_p = 0.335$, a main effect of time, $F(2.57, 102.61) = 10.69$, $p < 0.001$, $\eta^2_p = 0.211$, as well as a significant interaction of face type and time, $F(3.52, 140.98) = 6.95$, $p < 0.001$, $\eta^2_p = 0.148$; see Fig. 2. During both the 500–1500 ms period and the 1500 + ms period there is greater average pupil size for upside-down faces compared to neutral faces. In support of H1b, this difference is larger in the initial 500 ms–1500 ms period (upside-down faces: $M = 0.10$, $SD = 0.05$, neutral faces: $M = 0.06$, $SD = 0.05$, $t(40) = 6.82$, $p < 0.001$, $dz = 1.06$) compared to the 1500 + ms period (upside-down faces: $M = 0.12$, $SD = 0.07$, neutral faces: $M = 0.09$, $SD = 0.07$, $t(40) = 3.55$, $p = 0.001$, $dz = 0.55$).

11.3. Thatcherized vs. neutral faces

11.3.1. Negativity manipulation check

No difference in valence was found between Thatcherized faces and neutral faces, $t(40) = 1.42$, $p = 0.162$, $dz = 0.22$. Thatcherized faces were rated as negative 66.88% of the time and neutral faces were rated as negative 62.70% of the time.

11.3.2. Pupillary dilation

Comparing Thatcherized faces to neutral faces, we found a marginal significant main effect of the type of face, $F(1, 40) = 3.53$, $p = 0.067$, $\eta^2_p = 0.081$, a main effect of time, $F(2.06, 82.42) = 19.12$, $p < 0.001$, $\eta^2_p = 0.323$, as well as a marginally significant interaction of face type and time, $F(2.46, 98.47) = 2.65$, $p = 0.064$, $\eta^2_p = 0.062$; see Fig. 2. During the 500–1500 ms period we observe a greater average pupil size for Thatcherized faces compared to neutral faces; this is not the case during the 1500 + ms period. In support of H1b, this difference during the initial 500–1500 ms is significant (Thatcherized faces: $M = 0.09$, $SD = 0.05$, neutral faces: $M = 0.07$, $SD = 0.05$; $t(41) = 3.50$, $p = 0.001$, $dz = 0.54$). There was no significant difference during the 1500 + ms period, $t(41) = 1.68$, $p = 0.10$, $dz = 0.26$.

12. Bias comparisons: angry vs. upside-down vs. Thatcherized

12.1. Upside-down vs. angry faces

12.1.1. Negativity manipulation check

Angry faces were rated as more negative than upside-down faces, $t(33) = 13.38$, $p = 0.003$, $dz = 2.09$. Angry faces were rated as negative 97.83% of the time, while upside-down faces were rated as negative 35.17% of the time.

12.1.2. Pupillary dilation

Comparing upside-down faces and angry faces revealed no main effect of face type, $F(1, 33) = 0.18$, $p = 0.673$, $\eta^2_p = 0.005$, but did show a main effect of time, $F(2.90, 95.53) = 10.64$, $p < 0.001$, $\eta^2_p = 0.244$, and a significant interaction of face type an time, $F(3.41, 112.48) = 13.44$, $p < 0.001$, $\eta^2_p = 0.289$, see Fig. 3. During the 500–1500 ms period we observe a greater average pupil size for upside-down faces compared to angry faces; this is not the case during the 1500 + ms period. In support of H2a, this difference during the initial 500–1500 ms is significant (upside-down faces: $M = 0.10$, $SD = 0.08$, angry faces: $M = 0.08$, $SD = 0.08$; $t(33) = 4.50$, $p < 0.001$, $dz = 0.77$). There was no significant difference during the 1500 + ms period, $t(33) = 0.408$, $p = 0.686$, $dz = 0.07$.

12.2. Thatcherized vs. angry faces

12.2.1. Negativity manipulation check

Angry faces were also rated as more negative than Thatcherized faces, $t(42) = 16.80$, $p < 0.001$, $dz = 2.56$. Angry faces were rated as negative 96.72% of the time, while Thatcherized faces were rated as negative only 31.21% of the time.

12.2.2. Pupillary dilation

The comparison between Thatcherized and angry faces revealed a significant main effect of face type, $F(1, 42) = 5.82$, $p = 0.02$, $\eta^2_p = 0.122$, a main effect of time, $F(2.70, 113.31) = 41.03$, $p < 0.001$, $\eta^2_p = 0.494$, as well as a significant interaction of face type and time, $F(2.95, 123.96) = 30.32$, $p < 0.001$, $\eta^2_p = 0.419$; see Fig. 3. During the 500–1500 ms period we observe a greater average pupil size for Thatcherized faces compared to angry faces; during the 1500 + ms period we observe a greater average pupil size for angry faces compared to Thatcherized faces. In support of H2a, this difference during the initial 500–1500 ms is significant (Thatcherized faces: $M = 0.07$, $SD = 0.06$, angry faces: $M = 0.05$, $SD = 0.06$ $t(42) = 2.87$, $p = 0.006$, $dz = 0.44$). The difference during the 1500 + ms period is also significant (Thatcherized faces: $M = 0.11$, $SD = 0.09$, angry faces: $M = 0.14$, $SD = 0.07$; $t(42) = 3.27$, $p = 0.002$, $dz = 0.50$).

12.3. Thatcherized vs. upside-down faces

12.3.1. Negativity manipulation check

Thatcherized faces were seen as less negative than upside-down faces, $t(35) = 2.05$, $p = 0.048$, $dz = 0.34$. Thatcherized faces were rated as negative 48.11% of the time while upside-down faces were rated as negative 59.17% of the time.

12.3.2. Pupillary dilation

Comparing Thatcherized faces to upside-down faces, we found a significant main effect of face type, $F(1, 35) = 16.66$, $p < 0.001$, $\eta^2_p = 0.322$, a main effect of time, $F(1.83, 64.21) = 17.05$, $p < 0.001$, $\eta^2_p = 0.328$, as well as a significant interaction of face type and time, $F(3.42, 119.82) = 3.43$, $p = 0.015$, $\eta^2_p = 0.089$; see Fig. 3. We observe a greater average pupil size for Thatcherized faces compared to upside-down faces during the 500–1500 ms period and the 1500 + ms period. In support of H2b, this difference is significant in the initial 500 ms–1500 ms period (Thatcherized faces: $M = 0.11$, $SD = 0.06$, upside-down faces: $M = 0.09$, $SD = 0.05$, $t(35) = 5.22$, $p < 0.001$, $dz = 0.87$), and produces a slighter larger effect size than the PD difference in the 1500 + ms period (Thatcherized faces: $M = 0.15$, $SD = 0.09$, upside-down faces: $M = 0.13$, $SD = 0.09$, $t(35) = 3.69$, $p = 0.001$, $dz = 0.61$).

13. Discussion

In three baseline manipulation check and three bias comparisons, we used relative PD to test our expectancy bias hypothesis. The baseline comparisons represented our primary manipulation check determining whether our threatening (angry faces) and expectancy-violating stimuli (upside-down and Thatcherized faces) were experienced as attentionally salient relative to neutral versions of the same stimuli (neutral faces), and whether the negative and expectancy-violating stimuli produced patterns of pupillary reactivity that corresponded with other negative (H1a) and expectancy-violating (H1b) stimuli. These comparisons confirmed the validity of our chosen stimuli, insofar as angry, upside-down and Thatcherized faces evoked greater pupillary dilation compared to neutral faces, with angry faces evoking the same distinctive pattern of PD associated with sympathetic nervous system arousal following other negative stimuli (heightened PD 1500 ms +), and upside-down and Thatcherized faces both evoking the same distinctive pattern of PD associated with conflict detection following other expectancy-violating stimuli (PD spike 500 ms–1500 ms).

The three bias comparisons represented the primary test of our expectancy bias hypothesis, determining whether our expectancy-violating stimuli would evoke heightened PD earlier than negatively-valenced versions of the same stimuli (H2a), and whether our incongruous expectancy-violating stimuli (Thatcherized faces) would evoke heightened PD relative to a merely novel version of the same stimuli (upside-down faces) (H2b). Consistent with our hypotheses, we found that both upside-down and Thatcherized faces evoke
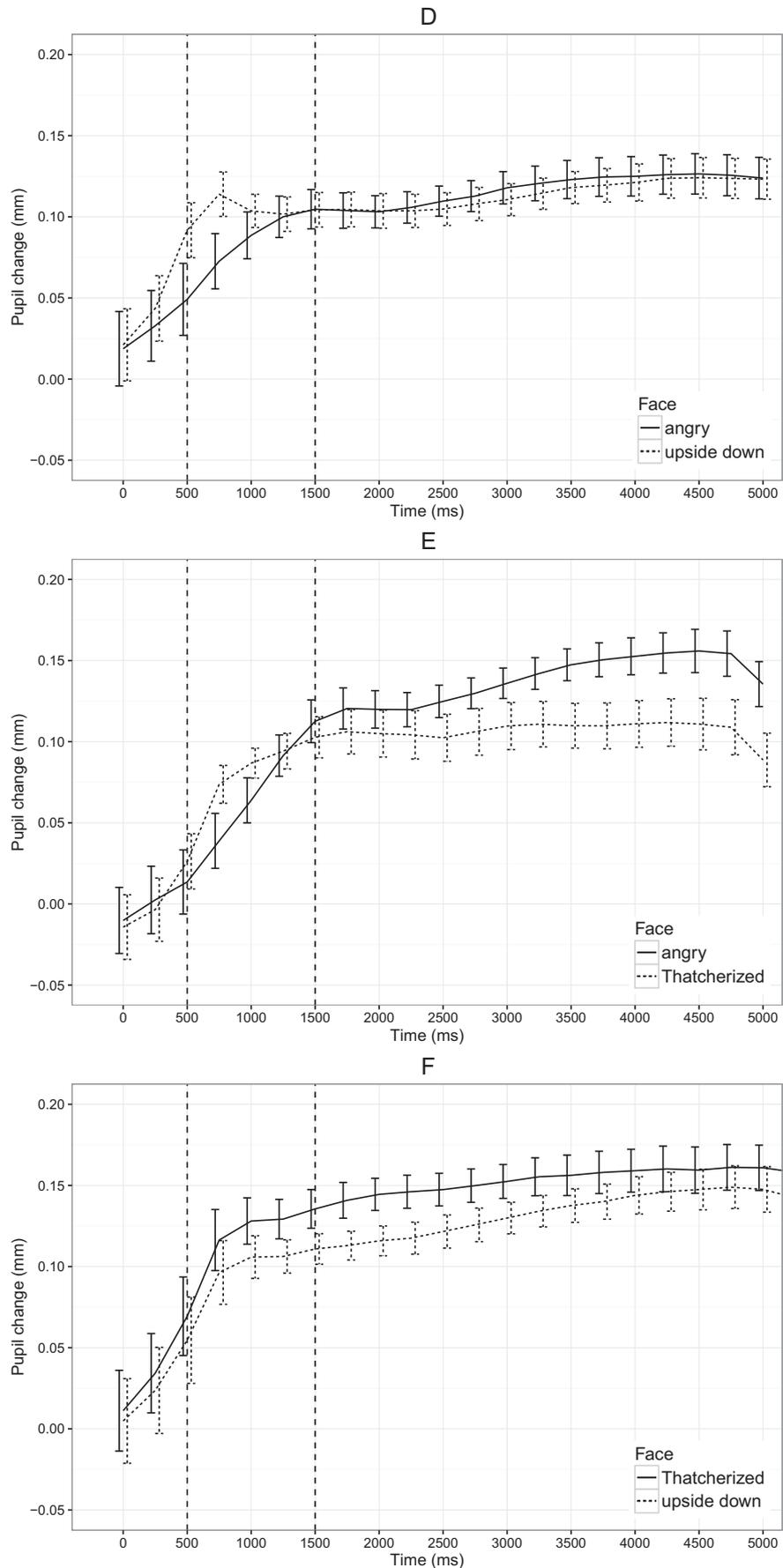
Fig. 3. Bias comparisons of pupillary dilation in response to D) upside-down vs. angry faces E) Thatcherized vs. angry face F) Thatcherized vs. upside-down faces.

heightened PD compared to angry faces, in a manner consistent with the baseline conditions (an initial spike between 500 ms–1500 ms vs. heightened PD following 1500 ms+). Furthermore, the degree of expectancy-violation impacted attentional arousal, such that Thatcherized upside-down faces evoked a heightened initial PD (500 ms–1500 ms) relative to upside-down faces.

Taken together, these findings are entirely consistent with an expectancy bias in arousal and attention, insofar as expectancy-violating stimuli are preferentially processed (i.e., heightened earlier PD) relative to neutral and negative versions of the same stimuli, where this effect is greatest for stimuli that is expectancy-violating to the greatest extent. None of these findings are consistent with a dominant negativity bias in pupillary arousal, insofar as the baseline manipulation checks validated and distinguished the patterns of pupillary dilation differentially evoked by expectancy-violating and negative versions of the same social stimuli, where these distinct patterns were manifest more generally across all six experiments—a pattern of results that would not be possible if the expectancy-violating stimuli were experienced in an emotionally negative manner, or vice versa. This validation of the stimuli was further confirmed by the additional negativity manipulation check, which established that neither of the expectancy-violating faces were experienced as more negative than neutral or angry faces.

14. Influence of Thatcherized faces

Rather than conducting a single within-subjects design, we chose to thoroughly test our expectancy bias hypothesis with six distinct paired comparisons out of concern that PD in response to one class of stimuli may be influenced by the presence of other classes of the same stimuli. While it did not alter the pattern of results relevant to our expectancy bias hypothesis, it appeared that the response to neutral and negative stimuli may have been heightened in the presence of the most expectancy-violating stimuli: Thatcherized faces.

14.1. Thatcherized and neutral

In the baseline comparison of Thatcherized and neutral faces, Thatcherized faces evoked the expected increase in pupillary dilation relative to neutral faces, however, this difference appeared relatively smaller than the increase evoked by upside-down faces. Inspection of the graph reveals that the relatively smaller difference between Thatcherized and neutral faces may have been due to an overall relative increase in pupillary dilation for the neutral faces, rather than a relative decrease in pupillary dilation for the Thatcherized faces. To test this possibility, we conducted exploratory analyses comparing PD for the neutral faces in the angry and upside-down baseline conditions to PD for neutral faces in the Thatcherized face condition.

When we compare the overall mean PD of neutral faces when presented with angry ($M = 0.07$, $SD = 0.07$) or Thatcherized ($M = 0.1$, $SD = 0.07$) faces, we find a near significant difference, $F(1, 81) = 3.96$, $p = 0.05$, $\eta^2_p = 0.047$, whereby neutral (vs. Thatcherized) faces evoke greater PD than neutral (vs. angry) faces. However, when we compare the overall mean PD of neutral faces when presented with upside-down ($M = 0.09$, $SD = 0.06$) or Thatcherized ($M = 0.1$, $SD = 0.07$) faces, the difference is not significant, $F(1, 81) = 0.51$, $p = 0.48$, $\eta^2_p = 0.01$, though the overall mean PD values for Thatcherized ($M = 0.1$, $SD = 0.06$) and upside-down ($M = 0.1$, $SD = 0.06$) faces are identical, and not significantly different $F(1, 81) = 0.02$, $p = 0.89$, $\eta^2_p < 0.01$.

In sum, both Thatcherized and upside-down faces produce the same magnitude of PD ($M = 0.1$, $SD = 0.06$) in their respective baseline conditions, and the magnitude of PD for neutral faces in the Thatcherized condition appears greater than neutral PD in the angry and upside-down conditions, albeit with near significance and non-significance, respectively. Future replications can determine whether the relative difference in PD is indeed greater for upside-down and neutral faces compared to Thatcherized and neutral faces, and if so, if this relative

difference is due to a general increase in PD for neutral faces in the presence of Thatcherized faces.

14.1.1. Thatcherized and angry

In the bias comparison of Thatcherized and angry faces, we found a "crossover" interaction between inconsistent and negative faces over time, whereby Thatcherized faces evoked relatively heightened PD during the initial expectancy-violation period (500 ms–1500 ms), and angry faces evoked relatively heightened PD during the period associated with sympathetic nervous system arousal (1500 ms+). These findings suggest that sympathetic nervous system arousal following the negative faces was amplified in the presence of inherently incongruent stimuli, relative to the presence of other expectancy-violating stimuli (e.g., merely novel). Indeed, contrasting the Thatcherized vs. angry and upside-down vs. angry comparisons, we find an interaction, whereas there is no such increase in sympathetic nervous system PD (1500 + ms) for negative faces in the presence of novel faces ($F(1, 75) = 4.89$, $p = 0.03$, $\eta^2_p = 0.061$; angry: $M = 0.11$, $SD = 0.09$ vs. upside-down: $M = 0.11$, $SD = 0.09$; angry: $M = 0.14$, $SD = 0.07$ vs. Thatcherized: $M = 0.11$, $SD = 0.09$).

These findings converge with prior research showing that the presence of incongruity both enhances and prolongs the sympathetic nervous system response to emotional stimuli (e.g., Dickerson & Kemeny, 2004; Grupe & Nitschke, 2011). More generally, it may be the case that the presence of highly expectancy-violating stimuli generally heightens arousal in response to all non-expectancy-violating stimuli, whether it is emotionally valenced or even neutral. Future studies can explore this hypothesis directly.

14.2. Expectancy bias vs. negativity bias

Prior research has shown evidence for a negativity bias, whereby humans respond earlier to negative than to positive or neutral stimuli (e.g., Baumeister et al., 2001; Cacioppo et al., 1999; Carretie et al., 2001). Over the course of six experimental comparisons, we offer initial support for the notion that humans maintain an even stronger expectancy bias, insofar as different examples of expectancy-violating stimuli captured a fundamental marker of attentional arousal—pupillary dilation—earlier than threatening versions of the same stimuli. While previous research has demonstrated that animals (e.g., Pearce & Hall, 1980) and humans (e.g., Dunchan-Johnson & Donchin, 1977; Hogarth, Dickinson, & Duka, 2010) are motivated to attend to unexpected relative to expected information, no research, to our knowledge, has directly compared attentional bias to un-valanced expectancy-violating stimuli relative to negatively and neutrally valanced unambiguous versions of the same stimuli. Moreover, our direct comparisons distinguish the differential impact of the degree to which stimuli are expectancy-violating, insofar as the most expectancy-violating stimuli evokes the greatest PD markers of attentional bias.

14.3. Limitations and future directions

14.3.1. Additional markers of attention

While our studies support an attentional bias towards expectancy-violating stimuli, they are limited by the use a specific measure of attentional arousal—pupillary dilation—which does not represent a complete and transparent window into the totality of attentional processing. This limitation is shared by other measures of attention and arousal to varying degrees (e.g., assessed with EEG, heightened N2 ERN is generally associated with incongruent stimuli (e.g., Stroop inconsistencies; Liotti et al., 2000), but may also result from positively valenced stimuli (e.g., pleasant images; Carretie, Hinojosa, Martin-Loeches, Mercado, & Tapia, 2004).

In the case of our six experiments, the consistent spike in PD recorded following both examples of expectancy-violating stimuli (Thatcherized and upside-down faces) is consistent with the same spike following

other examples of expectancy-violating stimuli (i.e., Stroop inconsistencies, Laeng et al., 2011), which evoke markers consistent with expectancy-violation using other measures (i.e., Stroop inconsistencies [Liotti et al., 2000] and mismatch paradigms [Folstein & van Petten, 2008] both evoke N2 ERPs with EEG). As well, the subsequent rise in PD recorded following the threatening stimuli (angry faces) is consistent with the same subsequent rise following other examples of threatening stimuli (i.e., threatening images, Bradley et al., 2008), which evoke subsequent markers consistent with threatening and otherwise emotionally negative stimuli using other measures (i.e., angry faces [Schupp et al., 2004] and unpleasant images [Hajcak, Dunning, & Foti, 2009] both evoke Late Positive Potential ERPs with EEG).

Subsequent research using simultaneous eye tracking and EEG can determine whether pupillary markers of inconsistency (spike in PD) that begin to peak 500 ms after stimulus onset are reliably preceded by markers of inconsistency (e.g., N2) that arrive 200 ms after stimulus onset. Furthermore, it can be determined whether pupillary markers of valenced emotional salience (gradually heightened PD) that peak 1500 ms after stimulus onset are reliably preceded by ERP markers of emotional regulation (e.g., Late Positive Potential) that arrive as late as 1000 ms after stimulus onset (Hajcak et al., 2009). If, indeed, markers of inconsistency and emotional salience assessed as PD were associated with corresponding markers assessed as ERPs—along the same sequential time course—this would provide robust convergent evidence for a general expectancy bias.

14.3.2. Additional sources of expectancy violation and emotional valence

These experiments operationalized threat in terms of a single class of salient social stimuli: human faces. Future research should assess the extent to which expectancy-violating and threatening examples of non-social stimuli (e.g., snakes) also produce effects consistent with an expectancy bias. Examining the extent of this bias should also involve comparing expectancy-violating and emotionally valenced versions of other negative stimuli, for example, sources of disgust. As well, the extent of any expectancy bias can be further assessed in terms of expectancy-violating and emotionally valenced version of non-negative stimuli: equally salient positive images. Finally, several current theorists propose an array of behaviors that should follow from the experience of expectancy-violation (e.g., Jonas et al., 2014; Proulx, 2012), and it remains to be seen whether these compensation behaviors uniquely follow from expectancy-violating (vs. threatening) experiences.

15. Conclusion

Whereas prior theorists have proposed an overriding negativity bias in human attention, the present study provides initial evidence for a superseding expectancy bias in pupillary arousal for social stimuli. Such a bias would have important implications for why—and how—people seek to regulate their emotions and guide their actions, insofar as thoughts, beliefs and behaviors may be proximally motivated by the reduction of feelings of uncertainty relative to potential threat. However, further research using a wider range of stimuli and attentional measures is required to specify the nature of any general expectancy bias, as well as to explore the diversity of its manifestation. In sum, though bad may be stronger than good, the eyes have it that incongruity may be strongest of all.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jesp.2016.12.003.

References

Amodio, D. M., Master, S. L., Yee, C. M., & Taylor, S. E. (2008). Neurocognitive components of the behavioral inhibition and activation systems: Implications for theories of self-regulation. *Psychophysiology, 45*, 11–19.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*, 403–450.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, D. (2001). Bad is stronger than good. *Review of General Psychology, 5*, 323–370.

Beatty, J. (1982). Task-evoked pupillary responses, processing workload, and the structure of processing resources. *Psychological Bulletin, 91*(2), 276.

Blascovich, J. (2000). Using physiological indexes of psychological processes in social psychological research. In H. T. Reis, C. M. Judd, H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 117–137). New York, NY US: Cambridge University Press.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602–607.

Bruner, J. (1990). *Acts of meaning.* Cambridge, MA: Harvard University Press.

Bruner, J., & Postman, L. (1949). On the perception of incongruity: A paradigm. *Journal of Personality, 18*, 206–223.

Cacioppo, J., Gardner, W., & Berntson, G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology, 76*, 839–855.

Carbon, C. -C., Schweinberger, S. R., Kaufmann, J. M., & Leder, H. (2005). The Thatcher illusion seen by the brain: An event-related brain potentials study. *Cognitive Brain Research, 24*, 544–555.

Carretie, L., Martin-Loeches, M., Hinojosa, J. A., & Mercado, F. (2001). Emotion and attention interaction studied through event related potentials. *Journal of Cognitive Neuroscience, 13*, 1109–1128.

Carretie, L., Hinojosa, J. A., Martin-Loeches, M., Mercado, F., & Tapia, M. (2004). Automatic attention to emotional stimuli: Neural correlates. *Human Brain Mapping, 22*, 290–299.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subjects and within-subject design. *Journal of Economic Behavior & Organization, 81*, 1–8.

Codispoti, M., Ferrari, V., & Bradley, M. M. (2007). Repetition and event-related potentials: Distinguishing early and late processes in affective picture perception. *Journal of Cognitive Neuroscience, 19*, 577–586.

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man Nouveauin du stimulus, adequation a la tache et potentiels evoques visuels chez l'homme. *Electroencephalography and Clinical Neurophysiology, 39*, 131–143.

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*, 355–391.

Dunchan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology, 14*, 456–467.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Folstein, J. R., & van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*, 152–170.

Foti, D., Hajcak, G., & Dien, J. (2009). Differentiating neural responses to emotional pictures: Evidence from temporal-spatial PCA. *Psychophysiology, 46*, 521–530.

Fowles, D. C. (1980). The three arousal model: Implications of Gray's two factor learning theory for heart-rate, electrodermal activity, and psychopathy. *Psychophysiology, 17*, 87–104.

Freud, S. (1990). The uncanny. In A. Dickson (Ed.), *Sigmund Freud: 14. Art and literature* (pp. 335–376). New York, NY: Penguin (Original work published 1919).

Fried, R., Schettlinger, K., & Borowski, M. (2014). robfilter: Robust Time Series Filters. [Computer software]. Retrieved from https://CRAN.R-project.org/package=robfilter

Gilzenrat, M. S., Cohen, J. D., Rajkowski, J., & Aston-Jones, G. (2003). Pupil dynamics predict changes in task engagement mediated by locus coeruleus. *Society for Neuroscience Abstracts, Vol. 515.* (pp. 19).

Gray, J. A., & McNaughton, N. (2003). *The neuropsychology of anxiety: An enquiry into the function of the septo-hippocampal system (no. 33).* Oxford University Press.

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin, 83*, 220–314.

Grupe, D. W., & Nitschke, J. B. (2011). Uncertainty is associated with biased expectancies and heightened responses to aversion. *Emotion, 11*, 413–424.

Gu, R. L., Ge, Y., & Huang, Y. (2010). Anxiety and outcome evaluation: The good, the bad, and the ambiguous. *Biological Psychiatry, 85*, 200–206.

Hajcak, G., Dunning, J. P., & Foti, D. (2009). Motivated and controlled attention to emotion: Time-course of the late positive potential. *Clinical Neurophysiology, 120*(3), 505–510.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology, 54*(6), 917.

Harmon-Jones, E., & Harmon-Jones, C. (2011). Social neuroscience of asymmetrical frontal cortical activity: Considering anger and approach motivation. In A. Todorov, S. Fiske, & D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind* (pp. 173–187). Oxford University Press.

Harmon-Jones, E., Amodio, D., & Harmon-Jones, C. (2009). Action-based model of dissonance: A review, integration, and expansion of conceptions of cognitive conflict. Advances in Experimental Social Psychology (vol. 41, pp. 119–166). San Diego, CA: Academic Press.

Hirsh, J. B., & Inzlicht, M. (2008). The devil you know: Neuroticism predicts neural response to uncertainty. *Psychological Science, 19*, 962–967.

Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychological Review, 119*, 304–320.

Hogarth, L., Dickinson, A., & Duka, T. (2010). Selective attention to conditioned stimuli in human discrimination learning: Untangling the effect of outcome prediction, value and uncertainty. In C. J. Mitchell, & M. E. Le Pelley (Eds.), *Attention and associative learning: From brain to behaviour* (pp. 71–97). Oxford: Oxford University Press.

Holroyd, C. B., Hajcak, G., & Larsen, J. T. (2006). The good, the bad and the neutral: Electrophysiological responses to feedback stimuli. *Brain Research, 1105*, 93–101.

Janoff-Bulman, R. (1992). *Shattered assumptions.* New York, NY: The Free Press.

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration–exploitation trade-off: Evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience, 23*(7), 1587–1596.

Johnson, R. (1988). The amplitude of the P300 component of the event-related potential: Review and synthesis. *Advances in Psychophysiology, 3*, 69–137.

Jonas, E., McGregor, I., Klackl, J., Agroskin, D., Fritsche, I., Holbrook, C., ... Quirin, M. (2014). Threat and defense: From anxiety to approach. In J. M. Olson, & M. P. Zanna (Eds.), *Advances in experimental social psychology, vol. 49.* (pp. 219–286). San Diego, CA: Academic Press.

Kagan, J. (2009). Categories of novelty and states of uncertainty. *Review of General Psychology, 13*, 290–301.

Kastner, S., & Ungerleider, L. G. (2001). The neural basis of biased competition in human visual cortex. *Neuropsychologia, 39*(12), 1263–1276.

Köhler, W. (1929). *Gestalt psychology.* New York: Liveright.

Kok, A. (1997). Event-related-potential (ERP) reflections of mental resources: A review and synthesis. *Biological Psychology, 45*, 19–56.

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary stroop effects. *Cognitive Processing, 12*, 13–21.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *Technical report A-6.* Gainesville, FL: University of Florida.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud faces database. *Cognition and Emotion, 24*, 1377–1388.

Lee, E., Kang, J. I., Park, I. H., Kim, J. J., & An, S. K. (2008). Is a neutral face really evaluated as being emotionally neutral? *Psychiatry Research, 157*(1), 77–85.

Lerner, M. J. (1980). *The belief in a just world: A fundamental delusion.* New York: Plenum Press.

Liotti, M., Woldorff, M. G., Perez, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia, 38*, 701–711.

Miller, N. E. (1944). *Experimental studies of conflict.* New York: Ronald.

Montague, P., Dayan, P., & Sejnowski, T. J. (1996). A framework for mescencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience, 16*, 1936–1947.

Öhman, A. (2009). Of snakes and faces: An evolutionary perspective on the psychology of fear. *Scandinavian Journal of Psychology, 50*(6), 543–552.

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.

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies, 59*, 185–198.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 106*, 532–552.

Peterson, J. (1999). *Maps of meaning: The architecture of belief.* New York: Routledge.

Piaget, J. (1937/1954). *The construction of reality in the child.* New York: Basic Books.

Piaget, J. (2000). Piaget's theory. In K. Lee, & K. Lee (Eds.), *Childhood cognitive development: The essential readings* (pp. 33–47). Malden: Blackwell Publishing.

Pixelmator [Computer software] (2016). Retrieved from http://www.pixelmator.com/

Prehn, K., Heekeren, H. R., Blasek, K., Lapschies, K., Mews, I., & van der Meer, E. (2008). Neuroticism influences pupillary responses during an emotional interference task. *International Journal of Psychophysiology, 70*, 40–49.

Preuschoff, K. T., Hart, B. M., & Einhauser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience, 5*, 115.

Proulx, T. (2012). Threat-compensation in social psychology: Is there a core motivation? *Social Cognition, 30*, 643.

Proulx, T., & Inzlicht, M. (2012). Moderated disanxiousuncertlibrium: Specifying the moderating and neuroaffective determinants of violation-compensation effects. *Psychological Inquiry, 23*, 386–396.

Proulx, T., Inzlicht, M., & Harmon-Jones, E. (2012). Understanding all inconsistency compensation as a palliative response to violated expectations. *Trends in Cognitive Sciences, 16*, 285–291.

Raikowski, J., Kubiak, P., & Aston-Jones, G. (1993). Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of LC in attention. *Society for neuroscience annual meeting abstract, Vol. 19.* (pp. 974).

R Core Team (2016). R: A language and environment for statistical computing [Computer software]. Retrieved from https://www.R-project.org/

Roitman, M. F., van Dijk, G., Thiele, T. E., & Bernstein, I. L. (2001). Dopamine mediation of the feeding response to violations of spatial and temporal expectancies. *Behavioural Brain Research, 122*, 193–199.

Rondeel, E. W., Van Steenbergen, H., Holland, R. W., & van Knippenberg, A. (2015). A closer look at cognitive control: Differences in resource allocation during updating, inhibition and switching as revealed by pupillometry. *Frontiers in Human Neuroscience, 9*.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Osychology Review, 5*, 296–320.

Schupp, H. T., Junghofer, M., Weike, A. I., & Hamm, A. O. (2003). Attention and emotion: An ERP analysis of facilitated emotional stimulus processing. *Neuroreport, 14*, 1107–1110.

Schupp, H. T., Junghofer, M., Weike, A. I., & Hamm, A. O. (2004). The selective processing of briefly presented affective pictures: An ERP analysis. *Psychophysiology, 41*, 441–449.

Schutter, D. J., de Haan, E. H., & van Honk, J. (2004). Functionally dissociated aspects in anterior and posterior electrocortical processing of facial threat. *International Journal of Psychophysiology, 53*(1), 29–36.

Sleegers, W. W., Proulx, T., & van Beest, I. (2015). Extremism reduces conflict arousal and increases values affirmation in response to meaning violations. *Biological Psychology, 108*, 126–131.

Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., ... Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PloS One, 6*, e18298. http://dx.doi.org/10.1371/journal.pone.0018298.

Sokolov, E. N. (2002). *The orienting response in information processing.* New Jersey: Mahwah, Erlbaum.

Stroop, J. R. (1992). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General, 121*, 15.

Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception.*

Tucker, D. M., & Frederick, S. L. (1989). Emotion and brain lateralization. In H. Wagner (Eds.), *Handbook of social psychophysiology* (pp. 27–70). Chichester, UK: Wiley & Sons.

Van den Bos, K. (2001). Uncertainty management: The influence of uncertainty salience on reactions to perceived procedural fairness. *Journal of Personality and Social Psychology, 80*, 931–941.

Van der Schalk, J., Fischer, A., Box, B., Wigboldus, D., Hawk, S., Rotteveel, M., & Hess, U. (2011). Convergent and divergent responses to emotional displays or ingroup and outgroup. *Emotion, 11*, 286–298.

Van Steenbergen, H., & Band, G. P. (2013). Pupil dilation in the Simon task as a marker of conflict processing. *Frontiers in Human Neuroscience, 7*, 215.

Weinberg, A., & Hajcak, G. (2010). Beyond good and evil: The time-course of neural activity elicited by specific picture content. *Emotion, 10*, 767–782.

Whalen, P. J., & Taylor, J. M. (2014). Fearful, but not angry, expressions diffuse attention to peripheral targets in an attentional blink paradigm. *Emotion, 14*, 462–468.

Wilson, T. D., Centerbar, D. B., Kermer, D. A., & Gilbert, D. T. (2005). The pleasures of uncertainty: Prolonging positive moods in ways people do not anticipate. *Journal of Personality and Social Psychology, 88*, 5–21.

Zeileis, A., Grothendieck, G., Ryan, J. A., Andrews, F., & Zeileis, M. A. (2015). *Package 'zoo'.*