Loving yourself more than your neighbor: ERPs reveal online effects of a self-positivity bias

Eric C. Fields[1,2] and Gina R. Kuperberg[1,2,3]

[1]Department of Psychology, Tufts University, Medford, MA, USA, [2]MGH/HST Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA, USA, and [3]Department of Psychiatry, Massachusetts General Hospital, Charlestown, MA, USA

A large body of social psychological research suggests that we think quite positively of ourselves, often unrealistically so. Research on this 'self-positivity bias' has relied mainly on self-report and behavioral measures, but these can suffer from a number of problems including confounds that arise from the desire to present oneself well. What has not been clearly assessed is whether the self-positivity bias influences the processing of incoming information as it unfolds in real time. In this study, we used event-related potentials to address this question. Participants read two-sentence social vignettes that were either self- or other-relevant. Pleasant words in self-relevant contexts evoked a smaller negativity between 300 and 500 ms (the N400 time window) than the same words in other-relevant contexts, suggesting that comprehenders were more likely to expect positive information when a scenario referred to themselves. This finding indicates that the self-positivity bias is available online, acting as a general schema that directly influences real-time comprehension.

INTRODUCTION

Most of us think well of ourselves, often unrealistically so. For example, in one classic study, 94% of college instructors thought that they were better than average in teaching ability and 68% placed themselves in the top 25%—obviously statistical impossibilities (Cross, 1977; for many other examples, see reviews by Taylor and Brown, 1988; Dunning et al., 2004; Alicke and Govorun, 2005). This 'self-positivity bias' has been widely studied and there are now numerous studies showing that we tend to evaluate ourselves more positively than others (the 'better-than-average effect'; Alicke and Govorun, 2005), and that we believe good things are more likely (and bad things less likely) to happen to ourselves than to others (the 'optimistic bias'; Armor and Taylor, 2002).

Positively biased self-views are argued to be a key component of healthy psychological functioning, influencing self-esteem, motivation, and determination (Taylor and Brown, 1988). Indeed, a lack of a self-positivity bias (or even a self-negativity bias) may contribute to mood and anxiety disorders (Beck et al., 1979; Taylor and Brown, 1988; Shestyuk and Deldin, 2010; Goldin et al., 2013). However, there are also negative effects of unrealistic self-assessment: for example, underestimating the likelihood of future health problems can stop us from taking preventative measures, and students' unrealistic views of how well they understand material can undermine effective studying (see Dunning et al., 2004 for a review).

Despite its importance and practical implications, there remains controversy about measurement of the self-positivity bias. The self-report questionnaires that are traditionally used often require subjects to explicitly compare themselves with others (either with a specific person or an 'average person'), and it has been argued that

self-positivity effects could be artifacts of this judgment process. For example, when rating a trait such as honesty, it may simply be easier to think of specific instances of honesty relating to ourselves than to the comparison target, leading to artificially high ratings (for a comprehensive review of such 'non-motivated' accounts of self-positivity effects, see Chambers and Windschitl, 2004). Perhaps even more importantly, responses on such questionnaires may not necessarily reflect our true self-views, but rather our desire to present ourselves well to others or even ourselves. Supporting this idea, measures that are designed to index self-presentation (such as impression management and self-deception) correlate with measures of self-positivity (Farnham et al., 1999).

In an effort to bypass conscious deliberation and tap into more automatic processes, implicit measures of self-esteem such as the Implicit Association Test (IAT; Greenwald and Farnham, 2000) and the Name-Letter Test (Koole and Pelham, 2003) have been developed. Self-positivity effects have been described using both these paradigms, and this has been taken as evidence for the existence of an automatic or implicit self-positivity bias that is unconfounded by controlled aspects of responding (for discussion, see Farnham et al., 1999; Greenwald and Farnham, 2000; Fazio and Olson, 2003; Olson et al., 2007). However, concerns have also been raised about these measures. One concern is that by attempting to avoid motivated responses, implicit measures may reduce access to important aspects of the self-concept itself. For example, the IAT's emphasis on speedy responding may 'deprive [participants] of the time they need to access and reflect upon autobiographical knowledge . . . that is potentially relevant to the associations they are making' (Buhrmester et al., 2011, p. 375). In a comprehensive review of the literature, Buhrmester et al. (2011) take this argument a step further: they conclude that currently existing implicit measures do not actually measure the self-concept at all and emphasize that these measures have repeatedly failed to correlate with socially important phenomena that self-esteem should predict. Finally, there is evidence that implicit measures may not be as successful as originally hoped in avoiding self-presentational confounds, possibly because they still require a behavioral response and self-enhancement tendencies can be automatized (see discussion in Buhrmester et al., 2011; see also Paulhus, 1993).

Thus, neither the explicit or implicit measures currently available have proven fully satisfactory in providing a measure with full access to the self-concept, while also avoiding self-presentational confounds. Put simply, it remains unclear whether the self-positivity bias emerges only through the process of making judgments or behavioral decisions about the self, or whether it acts as a schema that reflects a basic, implicit aspect of the way we comprehend the world. A measure showing that the self-positivity bias can directly influence the way we make sense of incoming events as they unfold in real time, in the absence of a behavioral response, would address this question and avoid many of the concerns raised in the previous paragraphs. Event-related potentials (ERPs), a direct measure of neural activity with excellent temporal resolution, are an ideal technique for this purpose (Luck, 2014).

Our focus in this work was on the N400, a negative-going, often centro-parietally distributed ERP component that peaks at around 400 ms after the onset of meaningful stimuli such as words or pictures. The N400 is thought to reflect semantic processing of a stimulus in relation to expectations set up by the preceding context and multiple types of information (including schemas) stored within memory (Kutas and Federmeier, 2011). A stimulus that is expected evokes a smaller N400 than a stimulus that is unexpected because the context pre-activates (or leads people to predict) upcoming relevant features (Kuperberg, 2013; Lau et al., 2013). For example, in a sentence like 'The children went outside to play/look', 'play' elicits a smaller N400 than 'look' (example from Federmeier et al., 2007). Although both these sentences are fully plausible, greater expectations for 'play' (as can be shown via the cloze procedure) lead to an attenuated N400. If positively biased aspects of the self-concept are available online, then people should expect more positive outcomes in sentences about themselves than about others, predicting a smaller N400 to positive words in self-relevant than other-relevant contexts. Testing this hypothesis was the goal of this research.

A few ERP studies have asked how self-relevance influences the processing of incoming emotional information (Li and Han, 2010; Shestyuk and Deldin, 2010; Herbert et al., 2011a,b; Fields and Kuperberg, 2012). However, these studies primarily focused on how self-relevance influences a later allocation of attentional and/or processing resources to emotional stimuli, as indexed by a fairly late ERP component,[1] a late positivity that generally peaks after 500 ms (Hajcak et al., 2010; Citron, 2012).[2] In addition, in many of these studies the late positivity began before 500 ms, overlapping spatially and temporally with the N400 and therefore making it difficult to discern independent effects within the N400 time window.

This study

In this study, we presented two-sentence social vignettes with a neutral, pleasant, or unpleasant critical word. To manipulate self-relevance, we exploited the fact that grammatical person directly influences the perspective of the mental model developed by comprehenders. Although second person leads to the engagement of a self-perspective, third person leads to the adoption of an 'other' perspective (Brunyé et al.,

2009). Thus, our study was a 3 (Emotion: neutral, pleasant, unpleasant) × 2 (Self-Relevance: self, other) design, e.g.: 'A man knocks on Sandra's/your hotel room door. She/You see(s) that he has a tray/gift/gun in his hand.'

We recently carried out a study with these same stimuli using a different, more active task (Fields and Kuperberg, 2012), but in that study were unable to examine modulation on the N400 because the late positive component began within the N400 time window (~400 ms). Here we used a comprehension task, which did not draw attention to the emotional aspects of stimuli. Our lab (Holt et al., 2009) and others (e.g. Fischler and Bradley, 2006) have shown that such comprehension tasks (compared with more explicit tasks such as emotional categorization) can reduce and delay the late positivity evoked by emotional words enough that independent effects on the earlier N400 can be observed.

If positively valenced aspects of the self-concept are available online and influence our expectations about upcoming information, this should produce an interaction between Emotion and Self-Relevance on the N400. Specifically, we hypothesized that participants' implicit expectations of positive outcomes in sentences about themselves would lead to facilitated processing of pleasant words in self-relevant versus other-relevant contexts reflected by an attenuated N400. In contrast, we predicted no N400 effects of Self-Relevance on the neutral words or unpleasant words. This is because the amplitude of the N400 is generally facilitated by words whose semantic features match prior expectations; it is not a direct measure of incongruence or implausibility per se (Kutas and Federmeier, 2011; Paczynski and Kuperberg, 2012). Thus, an effect of Self-Relevance on unpleasant stimuli would only be predicted if the other-relevant contexts led to expectations for unpleasant outcomes. In fact, all else being equal, we tend to evaluate other people positively (Sears, 1983), and the self-positivity bias is thought to reflect even more positive views of ourselves than of others (Alicke et al., 1995).

MATERIALS AND METHODS

Participants

Twenty-eight Tufts University students originally participated in the ERP study. Four were excluded from analysis due to excessive artifact in the electroencephalogram (EEG), leaving 24 participants (17 females) in the final analysis. Self-reported race and ethnicity were non-Hispanic White ($n = 21$), mixed Asian/White ($n = 1$), Hispanic (race not otherwise indicated, $n = 1$), and unreported ($n = 1$). All participants were right-handed native English speakers (age 18–23 years, $M = 19.3$, s.d. = 1.6) who reported no history of psychiatric or neurological disorders and were not currently taking psychoactive medications. We also administered the Beck Depression Inventory (BDI-II) and the State-Trait Anxiety Inventory in order to rule out participants with symptoms of mood and anxiety disorders (none scored outside the lower range of each scale, and so none were excluded for this reason). Participants were paid for their participation and provided informed consent in accordance with the procedures of the Institutional Review Board of Tufts University.

Stimuli

Stimuli were the same as those used in Fields and Kuperberg (2012). In total, 222 sets of two-sentence scenarios were developed with Emotion (pleasant, neutral, and unpleasant) and Self-Relevance (self and other) conditions crossed in a 3 × 2 factorial design. The scenarios were written to include a broad range of situations that would be familiar and/or plausible to our subject population (e.g. many were about school or professional jobs). The first sentence introduced a situation involving one or more people, only one of which was specifically named (evenly

---

[1] One previous study has reported what the authors suggest may be a self-positivity bias on the N400 (Watson et al., 2007). However, because the N400 is a negative-going component with a fairly stable latency around 400 ms (Kutas and Federmeier, 2011) and the component in question was a positive-going component peaking after 500 ms, we think it is more likely that Watson et al.'s results reflect a modulation of the late positivity.

[2] It is not clear how a self-positivity bias would influence this late positive component. On the one hand, one might predict a larger late positivity to stimuli that are consistent with a self-positivity bias because they are preferred; on the other hand, one might argue that the positivity should be larger to incoming stimuli that are inconsistent with a self-positivity bias because they are unexpected (and may therefore require more processing resources to integrate).

Table 1 Examples of two-sentence scenarios in each of the six conditions

Self			Other		
Pleasant	Neutral	Unpleasant	Pleasant	Neutral	Unpleasant
A man knocks on Sandra's hotel room door. She sees that he has a gift in his hand	A man knocks on Sandra's hotel room door. She sees that he has a tray in his hand	A man knocks on Sandra's hotel room door. She sees that he has a gun in his hand	A man knocks on your hotel room door. You see that he has a gift in his hand	A man knocks on your hotel room door. You see that he has a tray in his hand	A man knocks on your hotel room door. You see that he has a gun in his hand
Fletcher writes a poem for a class. His classmates think it is a very beautiful composition	Fletcher writes a poem for a class. His classmates think it is a very intricate composition	Fletcher writes a poem for a class. His classmates think it is a very boring composition	You write a poem for a class. Your classmates think it is a very beautiful composition	You write a poem for a class. Your classmates think it is a very intricate composition	You write a poem for a class. Your classmates think it is a very boring composition
Vince spends time with his relatives over the vacation. This turns out to be a wonderful experience for him in many ways	Vince spends time with his relatives over the vacation. This turns out to be a characteristic experience for him in many ways	Vince spends time with his relatives over the vacation. This turns out to be a disastrous experience for him in many ways	You spend time with your relatives over the vacation. This turns out to be a wonderful experience for you in many ways	You spend time with your relatives over the vacation. This turns out to be a characteristic experience for you in many ways	You spend time with your relatives over the vacation. This turns out to be a disastrous experience for you in many ways
After dinner, Lydia is involved in a discussion. Many of her remarks impress people	After dinner, Lydia is involved in a discussion. Many of her remarks surprise people	After dinner, Lydia is involved in a discussion. Many of her remarks hurt people	After dinner, you are involved in a discussion. Many of your remarks impress people	After dinner, you are involved in a discussion. Many of your remarks surprise people	After dinner, you are involved in a discussion. Many of your remarks hurt people
Carmelo has been in his current job for over a year. He learns that he is getting a bonus this month	Carmelo has been in his current job for over a year. He learns that he is getting a transfer this month	Carmelo has been in his current job for over a year. He learns that he is getting a pay-cut this month	You have been in your current job for over a year. You learn that you are getting a bonus this month	You have been in your current job for over a year. You learn that you are getting a transfer this month	You have been in your current job for over a year. You learn that you are getting a pay-cut this month

Notes: The critical word is underlined (but did not appear underlined in the actual stimulus lists). Additional examples can be found in Table S1 (supplementary material).

split between male and female names), and which was always neutral or ambiguous in valence. The second sentence continued the scenario and was the same across all Emotion conditions except for the critical word, which was pleasant, neutral, or unpleasant. To create the self condition, the named protagonist was changed to 'you'. See Table 1 for examples.

A series of norming studies, summarized below, were carried out via the internet, with inclusion criteria for participants that were the same as for the ERP study. All means and standard deviations are reported in Table 2.

Length, concreteness, cloze and constraint

Stimuli were matched across conditions for critical word length (number of letters) and HAL word frequency (Lund and Burgess, 1996; Balota *et al.*, 2007) [$Fs < 2.9$, $Ps > 0.05$]. Critical words differed in concreteness across conditions [$F(2, 660) = 7.13$, $P = 0.001$], with neutral critical words being rated as slightly, but significantly, more concrete than pleasant or unpleasant critical words, which did not differ from one another. In a cloze study, scenarios were cut off before the critical word and 29 participants per scenario gave the word they thought most likely to come next. Stimuli were matched across conditions for cloze probability (percentage of responses matching the critical word) and constraint (frequency of the modal response) [$Ps > 0.5$].

Valence and arousal

Valence and arousal ratings were gathered for both the critical words in isolation and the scenarios up until the critical word. Valence ratings were as expected for both critical words and scenarios: the pleasant condition was rated as more positive than the neutral condition, which was rated as more positive than the unpleasant condition [$Fs > 1000$, $Ps < 0.001$]. In the scenarios, Self-Relevance amplified these differences, making pleasant scenarios more positive and unpleasant scenarios more negative [Emotion × Self-Relevance interaction: $F(2, 442) = 26.50$, $P < 0.001$].

As expected, there was a main effect of arousal for the both the critical word and scenario ratings [$Fs > 70$, $Ps < 0.001$], with pleasant and unpleasant stimuli being rated as more arousing than neutral stimuli. For the critical word ratings pleasant were rated as more arousing, but for the scenario ratings unpleasant were rated as more arousing. For the scenario ratings, there was no Emotion by Self-Relevance interaction [$F(2, 442) = 0.02$, $P = 0.980$], but there was a main effect of Self-Relevance [$F(1, 221) = 162.71$, $P < 0.001$] due to self-relevant scenarios being rated as more arousing than other-relevant scenarios.

Procedure
Stimulus presentation

All stimuli were counterbalanced such that each of the 222 scenarios appeared in a different condition in each of six lists (thus appearing in all conditions across lists). Participants were randomly assigned to a list (with the provision that there were an equal number of participants for each list). Trials were presented in random order, both within and across lists.

Trials were self-paced: they each began with the word 'READY' until the participant pressed a button to begin the trial. In each trial, the first sentence then appeared in full until the participant pressed a button to advance to the second sentence. The second sentence began with a fixation cross displayed for 500 ms, followed by an interstimulus interval (ISI) of 100 ms, followed by each word presented individually for 400 ms with an ISI of 100 ms. The final word of the scenario appeared on the screen for a longer duration of 750 ms, 400 ms ISI. Participants were asked to refrain from blinking during the second sentence of each scenario (which contained the critical word), but no restrictions were given for other parts of the trial.

Task

To ensure that participants were attending to the scenarios and comprehending them for meaning, 40 scenarios (randomly interspersed in

Table 2 Stimuli ratings and characteristics

	Other			Self		
	Pleasant	Neutral	Unpleasant	Pleasant	Neutral	Unpleasant
Cloze probability	3% (9%)	3% (7%)	3% (9%)	3% (8%)	3% (8%)	3% (7%)
Constraint	22% (13%)	22% (13%)	22% (13%)	22% (12%)	22% (12%)	22% (12%)
(log) HAL frequency[a]	8.39 (2.04)	8.47 (1.89)	8.28 (1.72)	—	—	—
Number of letters	7.67 (2.38)	7.48 (2.20)	7.14 (2.47)	—	—	—
Concreteness	3.45 (0.85)	3.72 (0.92)	3.54 (0.84)	—	—	—
Valence (word)	5.69 (0.55)	4.32 (0.56)	2.34 (0.57)	—	—	—
Arousal (word)	4.48 (0.80)	3.38 (0.64)	3.80 (0.63)	—	—	—
Valence (scenario)	5.25 (0.48)	4.12 (0.51)	2.37 (0.48)	5.40 (0.52)	4.17 (0.55)	2.24 (0.53)
Arousal (scenario)	3.61 (0.77)	3.22 (0.66)	3.84 (0.74)	3.87 (0.79)	3.49 (0.75)	4.11 (0.75)

Notes: Means are shown with standard deviations in parentheses. Cloze probability and constraint are represented as the percentage of total responses from 29 subjects. Concreteness, valence and arousal were all rated on seven-point scales from least concrete (most abstract), very unpleasant and least arousing, to most concrete, very pleasant and most arousing, respectively. '—' indicates that the values were the same in the self conditions as in the other conditions as the identical critical words were used, except for in six scenarios in which the verb was conjugated differently.
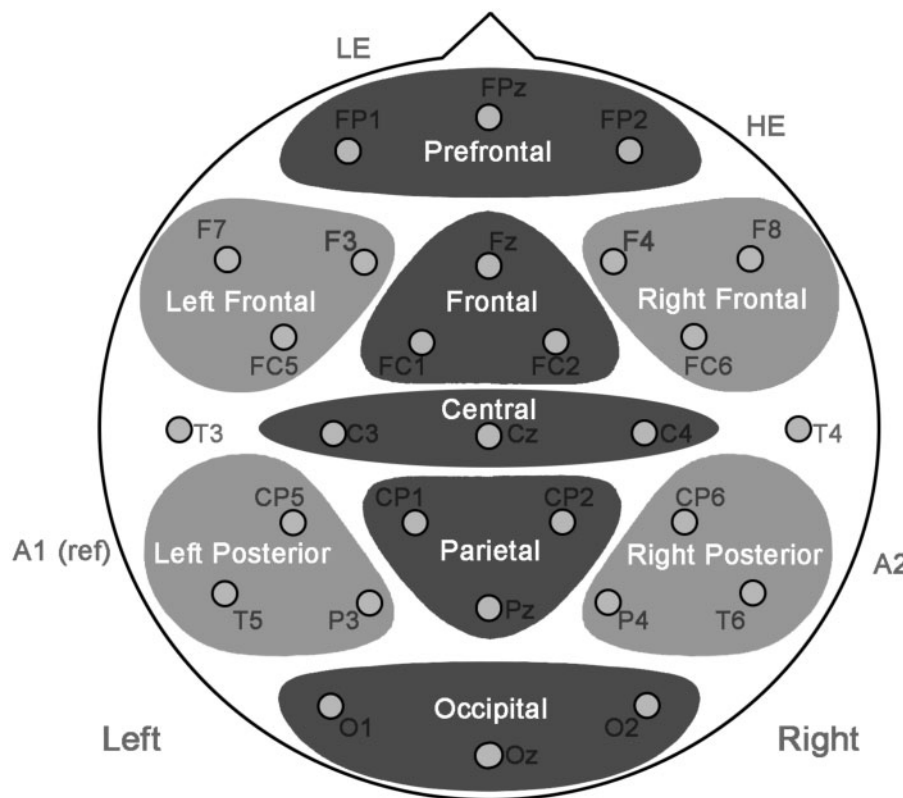
Fig. 1 Electrode montage with regions used for analysis. For the purposes of statistical analyses, the scalp was divided into three-electrode regions. Regions in dark gray were part of the mid-regions omnibus ANOVA and regions in light gray were part of the peripheral regions omnibus ANOVA.

each list) were followed by a yes-or-no comprehension question that stayed on the screen until the participant gave his/her answer via button press. Each question and its correct answer were the same across all conditions of a particular scenario except where the self-relevance manipulation required changes to names/pronouns and verb conjugations and no question directly referred to the valenced aspects of the scenarios. For example, the scenario 'Casper/You is/are new on campus. Everyone thinks he/you is/are quite idiosyncratic/ clever/dumb compared with most people' was followed by the question 'Did Casper/you go to this school last year?' with the correct answer being 'no'.

ERP acquisition and processing

The EEG response was recorded from 29 tin electrodes in an elastic cap (Electro-Cap International, Inc., Eaton, OH; Figure 1). Additional electrodes were placed below the left eye and at the canthus of the right eye to monitor vertical and horizontal eye movements. The impedance was kept below 2.5 kΩ for mastoid electrodes, 10 kΩ for electrooculogram (EOG) electrodes and 5 kΩ for all other electrodes. The EEG signal was amplified by an Isolated Biometric Amplifier (SA Instrumentation Co., San Diego, CA), band pass filtered online at 0.01–40 Hz and continuously sampled at 200 Hz. ERPs were referenced online to the left mastoid. The results reported below were derived using this reference for

Table 3 Effects of Self-Relevance at each level of Emotion within the N400 (300–500 ms) time window

	Effect	df	Neutral			Pleasant			Unpleasant		
			F	p	η^2	F	p	η^2	F	p	η^2
Mid-regions omnibus ANOVA	S	1,23	0.35	0.562	0.015	5.51	**0.028**	0.193	0.02	0.893	0.001
	S×R	4,92	0.26	0.755	0.011	3.26	**0.040**	0.124	0.68	0.526	0.029
Prefrontal	S	1,23	—	—	—	3.76	0.065	0.141	—	—	—
Frontal	S	1,23	—	—	—	7.66	**0.011**	0.250	—	—	—
Central	S	1,23	—	—	—	5.43	**0.029**	0.191	—	—	—
Parietal	S	1,23	—	—	—	4.39	**0.047**	0.160	—	—	—
Occipital	S	1,23	—	—	—	1.45	0.241	0.059	—	—	—
Peripheral regions omnibus ANOVA	S	1,23	0.79	0.383	0.033	4.45	**0.046**	0.162	0.03	0.872	0.001
	S×R	1,23	0.45	0.510	0.019	0.23	0.634	0.010	2.56	0.124	0.100
	S×H	1,23	0.11	0.742	0.005	0.57	0.457	0.024	0.02	0.877	0.001
	S×R×H	1,23	0.05	0.828	0.002	0.65	0.428	0.027	0.00	0.970	0.000
Frontal	S	1,23	—	—	—	—	—	—	—	—	—
Parietal	S	1,23	—	—	—	—	—	—	—	—	—

Notes: We resolved the significant Emotion × Self-Relevance interaction by examining effects of Self-Relevance at each level of Emotion. Follow-up analyses were carried out in individual regions only when the omnibus ANOVA showed a significant or marginally significant (<0.10) interaction between Self-Relevance and Region. Effects significant at an alpha of 0.05 are shaded gray and shown in a bold font. S, Self-Relevance; R, Region; H, Hemisphere.

comparability with our previous studies of emotional language, which also used a left mastoid reference (Holt *et al.*, 2009; Fields and Kuperberg, 2012; Delaney-Busch and Kuperberg, 2013). However, all analyses were repeated after re-referencing to average mastoids and the pattern of effects in the N400 time-window was the same (see Footnote 4).

The EEG was collected and processed using in-house software (available at: http://neurocoglaboratory.org/ERPSystem.htm). Segments from 100 ms before onset to 1100 ms after onset of each event were obtained. The 100 ms period immediately preceding stimulus onset was used as the baseline for all amplitude measurements. Trials with muscular and ocular artifact were identified and discarded using three algorithms: the first returns the number of time points within a given amplitude range of the minimum or maximum point of an epoch and is used to monitor for amplifier blocking or signal loss (i.e., a flat line), the second returns the difference between the maximum and minimum point of an epoch at the vertical and horizontal eye channels (independently) to monitor for eye movement and other large deflections unlikely to be neural activity, the third returns the difference of the mean difference and maximum difference between the electrode under the left eye and the electrode on the forehead above this eye and is used to identify blinks (which are characterized by opposite polarity shifts in these two channels). Appropriate thresholds for each of these algorithms were determined for each subject via visual inspection of the raw data, but were the same across all events within each subject (i.e. they were the same across all experimental conditions due to the within-subjects design). Overall, 7.5% of critical word trials were rejected for artifact with at least 28 trials averaged for each condition for every subject, and the rejection rate did not differ across conditions [Fs < 1, Ps > 0.4]. Averaged ERPs were calculated from trials remaining after artifact rejection.

ERP statistical analysis

The scalp was subdivided into three-electrode regions along its anterior–posterior distribution, at both mid-line and peripheral sites. Two omnibus analyses of variance (ANOVAs), one covering mid-regions (dark gray in Figure 1) and another covering peripheral regions (light gray in Figure 1), were conducted using SPSS 21 (IBM) with Emotion, Self-Relevance, Region and Hemisphere (peripheral regions only) as within-subjects factors. For all tests of significance, the Greenhouse

and Geisser (1959) estimation of ε was used to correct the degrees of freedom (the original degrees of freedom are reported in the text along with the corrected P values). A significance level of $\alpha = 0.05$ was used for all analyses.

RESULTS

Participants were quite accurate in answering the comprehension questions, indicating that they were engaged in reading and comprehending the scenarios. The average accuracy was 94% and no participant was below 85%, except for one participant who expressed confusion during the experiment about which buttons corresponded to yes and no (as there were no other problems with this participant's data, it was included in subsequent ERP analyses).

The N400 was defined a priori as the mean amplitude between 300 and 500 ms. Analyses of this time-window revealed significant interactions between Emotion and Self-Relevance in both the mid-regions [$F(2, 46) = 4.50$, $P = 0.018$, $\eta^2 = 0.164$] and peripheral regions [$F(2, 46) = 3.87$, $P = 0.029$, $\eta^2 = 0.144$] ANOVAs. In line with our a priori hypothesis, we broke down this interaction by examining the effect of Self-Relevance at each level of Emotion. Results are reported in Table 3 and Figure 2. There were no significant differences between the self and other contexts on the N400 evoked by neutral or unpleasant critical words. However, pleasant critical words evoked a smaller negativity in the self-relevant contexts than the other-relevant contexts.[3] This effect was significant in the mid-regions and marginally significant in the peripheral regions ANOVA. In the mid-regions ANOVA, there was a further interaction with Region, and follow-ups showed that the effect of Self-Relevance on pleasant words was significant in frontal, central and parietal regions, was marginally significant in the prefrontal region, but was not significant in the occipital region.[4]

[3] This effect cannot be attributed to an overlapping late positivity to the self-relevant (*vs* other-relevant) words, as it was not significant either in the full late positivity time window (500–800 ms) or in the 500–600, 600–700, or 700–800 ms time windows all Fs < 4.3, Ps > 0.05. In addition, the self-other contrast for pleasant words showed a different (more posterior) scalp distribution in these later time windows than in the 300–500 ms time window.

[4] Results remained the same when the EEG was re-referenced to an average of the two mastoid electrodes. In the mid-regions ANOVA, there was a significant interaction of Emotion and Self-Relevance [$F(2, 46) = 3.22$, $P = 0.049$, $\eta^2 = 0.123$] on the N400 and follow-ups showed a significant effect of Self-Relevance for pleasant words [$F(1, 23) = 4.66$, $P = 0.041$, $\eta^2 = 0.169$], but not neutral or negative words Fs < 1, Ps > 0.5. The distribution of the effect was nearly identical and all interactions with the hemisphere factor remained non-significant Fs > 2, Ps > 0.1.
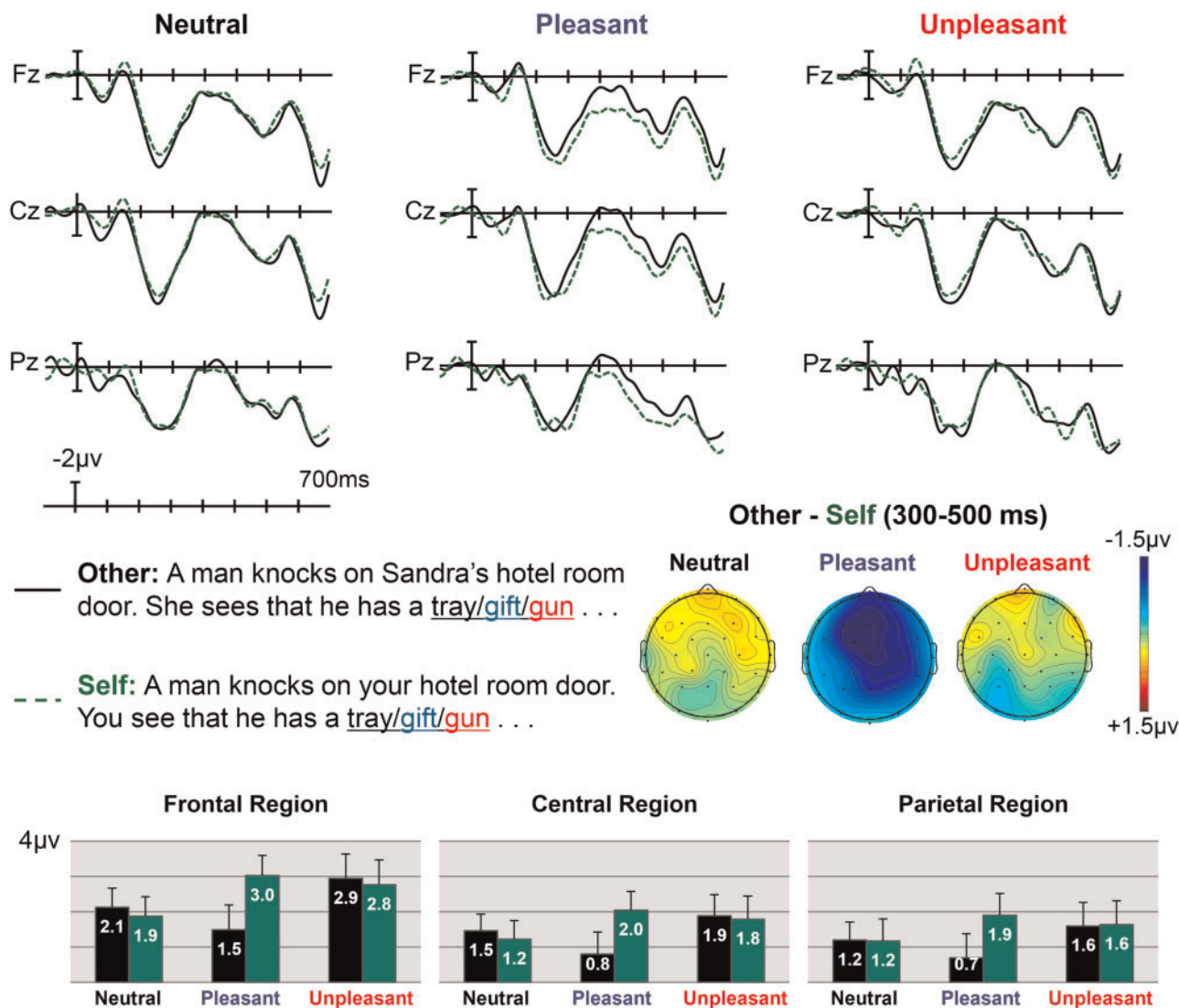
Fig. 2 Effects of Self-Relevance at each level of Emotion. Waveforms are time-locked to the critical word (underlined). They are low-pass filtered with a half-amplitude cutoff at 15 Hz for viewing purposes. Voltage maps show mean voltage across the N400 (300–500 ms) time window. Tables show mean voltage for the midline Frontal, Central and Parietal regions in the N400 (300–500 ms) time window and error bars represent the standard error of the mean. Difference waves are available in the supplementary material.

Analyses of other time windows are available in the supplementary material.

DISCUSSION

We used ERPs to determine whether positively biased aspects of the self-concept are available online and whether they influence the processing of incoming stimuli as they unfold in real time. We found that pleasant words elicited a reduced negativity between 300 and 500 ms in self-relevant contexts compared with other-relevant contexts.

This effect cannot be explained by any differences in the critical words themselves, as these were identical across the self-relevant and other-relevant conditions within each Emotion condition. Cloze probability was also the same across the six conditions and the contexts were not highly constraining. It is therefore unlikely that the effect seen here was driven by a pre-activation (or prediction) of the specific words that were presented. Rather, we suggest that a self-relevant context activated subjects' (generally positive) self-schema, along with its associated positive features. This made it easier to retrieve the positively valenced properties of the pleasant critical words, reflected by

activity within the N400 time window (cf. Federmeier and Kutas, 1999; Paczynski and Kuperberg, 2012; for discussion, see Kuperberg, 2013).

As noted in the Introduction, some theorists have argued that the better-than-average effect and optimistic bias could be artifacts of certain types of judgments, including the process of explicitly comparing oneself with others (Chambers and Windschitl, 2004). Others have proposed that self-positivity effects could be primarily a function of impression management—that is, attempts to present oneself in a good light (Farnham et al., 1999; Paulhus, 1993, 2002; Alicke and Govorun, 2005). Under either account, the self-positivity bias should only emerge when making a judgment or behavioral response with regard to the self, as is done in traditional explicit measures and implicit measures (although designed to avoid impression management, implicit measures may be sensitive to automated self-enhancement tendencies, see Paulhus, 1993; Buhrmester et al., 2011). The current findings show that this is not the case: in our paradigm, participants were not asked to make any kind of self-related judgment and nothing in our instructions indicated that we were assessing self-views. Nonetheless, we saw evidence of a self-positivity bias within a few hundred milliseconds of processing incoming words.

Our paradigm also avoids another potential drawback of implicit measures of self-esteem, which, by attempting to bypass conscious processing altogether, may fail to index important aspects of the self-concept (see Buhrmester *et al.*, 2011 for discussion). Because self-relevance was introduced in the first sentence of the context in our stimuli, participants had sufficient time to access and activate scenario-relevant self-knowledge. Thus, we would expect ERP effects in our paradigm to be sensitive to any relevant stored information about the self.

With these characteristics of the paradigm in mind, there are two important implications of our findings. First, our results suggest that the self-positivity bias quickly influences how we make sense of incoming information at early stages of comprehending meaning. This provides evidence that a positively biased self-concept is a basic, relatively automatic aspect of the way we view the world. Thus, whatever the role of impression management in unrealistic self-positivity, it appears the self-positivity bias cannot be fully attributed to socially desirable responding, whether conscious or automated (Paulhus, 1993, 2002). Second, if our interpretation is correct, the ERPs in this paradigm covertly measure whatever expectations subjects naturally generate about themselves as they comprehend the scenarios, and these expectations are not necessarily explicitly, intentionally, or even fully consciously generated by participants. Our paradigm may therefore provide a more accurate measure of what subjects fundamentally believe and feel about themselves than traditional explicit or implicit measures.

Open questions and future directions

Of note, the scalp distribution of the effect within the N400 time window was fairly widespread, but it had a more anterior focus than the more central-posterior (and sometimes right-lateralized) N400 effect that is typically associated with lexically expected *vs* unexpected words in sentences. Some researchers have argued that emotional words elicit a more frontal N400 than non-emotional words (Egidi and Nusbaum, 2012; see also Kanske and Kotz, 2007; Pratt and Kelly, 2008; Delaney-Busch and Kuperberg, 2013). The N400 is likely generated by multiple underlying sources (Federmeier and Laszlo, 2009). It is therefore possible that words that match (*vs* mismatch) emotional expectations may elicit additional contributing sources, or that such words fail to generate activity in some of the sources that are active in response to words that match on other types of semantic features. Indeed, other factors, including imageability and concreteness, are also known to shift the N400 anteriorly and these effects have been explained in similar ways (Holcomb *et al.*, 1999; Swaab *et al.*, 2002; West and Holcomb, 2002; see also Sitnikova *et al.*, 2003; Lee and Federmeier, 2008).[5] However, it is important to note that N400 effects with a more canonical posterior distribution have also been seen to emotional words in discourse contexts (e.g. Holt *et al.*, 2009; Leon *et al.*, 2010). Future research will be needed to more fully understand this pattern.

More generally, it will be important for future work to test the properties of our effect as a measure of the self-concept (and social schemas more generally, see below). For example, how does it correlate with common explicit and implicit measures (and does this correlation increase under situations where participants are expected to be more honest?). Regardless of its precise interpretation, however, for all the reasons discussed above, the pattern of results suggest the observed effect is related to emotionally valenced aspects of the self-concept.

To the extent that the present paradigm proves a useful measure of the self-concept, it may be helpful in determining how the self-positivity bias varies both within and between individuals and populations. In this study, our sample was a group of Western, largely white students at a private university who were screened to exclude psychiatric disorders. These participants are probably particularly likely to show evidence of a self-positivity bias (Heine *et al.*, 1999; Henrich *et al.*, 2010). It is important to recognize, however, that a self-positivity bias may not be universal for all people at all times, and it will be important for follow-up studies to examine variability in self-positivity both within and between individuals and populations. For example, our paradigm could be used to determine which aspects of the self are most immediately accessible in a given situation, mental state, or mood (Markus and Kunda, 1986; Helweg-Larsen and Shepperd, 2001). Relatedly, it may be useful for examining how the self-concept is altered in psychiatric disorders, including whether negative cognitions and attitudes about the self reflect implicit and relatively automatic effects or whether they result from effortful and elaborative rumination (Shestyuk and Deldin, 2010). Finally, it could be used to address debates in cultural psychology about whether a self-positivity bias is universal *vs* specific to Western cultures, which, for many of the same reasons discussed in this article, current self-report and implicit measures have not been able to fully resolve (Heine *et al.*, 1999, 2007; Sedikides *et al.*, 2003, 2007; see also Kitayama and Park, 2010, 2014).

Our findings also have some more general methodological implications. Social psychologists have long sought methods to measure attitudes in ways that are not affected by social norms and impression management, and we are not the first to think that neurophysiological techniques (because they are a direct measure of mental activity not relying on a subject response) may be one way to do this. In some previous ERP work, researchers have attempted to use the late positivity as such a measure (e.g. Crites *et al.*, 1995). However, the late positivity is highly sensitive to task and context, and generates less clear predictions for the effects of social schemas. The N400 is a relatively implicit neural measure of online comprehension, and it has now been shown to be sensitive to various types of social schemas and biases (e.g. gender stereotypes: White *et al.*, 2009; moral beliefs: Van Berkum *et al.*, 2009; and class and culture based differences in trait inference: Na and Kitayama, 2011; Varnum *et al.*, 2012). This study adds to this evidence by suggesting the N400 may also be sensitive to the self-positivity bias. Taken together, these studies suggest that ERPs and the N400 can prove a valuable addition to current implicit and explicit behavioral measures in the social psychologist's toolbox.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

Conflict of Interest

None declared.

REFERENCES

Alicke, M.D., Govorun, O. (2005). The better-than-average effect. In: Alicke, M.D., Dunning, D., Krueger, J., editors. *The Self and Social Judgement*. New York: Psychology Press, pp. 85–106.

Alicke, M.D., Klotz, M.L., Breitenbecher, D.L., Yurak, T.J., Vredenburg, D.S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology*, 68(5), 804–25.

Armor, D.A., Taylor, S.E. (2002). When predictions fail: the dilemma of unrealistic optimism. In: Gilovich, T., Griffin, D., Kahneman, D., editors. *Heuristics and Biases: The Psychology of Intuitive Judgement*. New York: Cambridge University Press, pp. 334–47.

Balota, D.A., Yap, M.J., Cortese, M.J., et al. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3), 445–59.

Beck, A.T., Rush, A.J., Shaw, B.F., Emery, G. (1979). *Cognitive Therapy of Depression*. New York: Guilford Press.

---

[5] Another possibility is that emotion (and self-relevance) may make readers more likely to form mental images of the scenarios they are reading, leading to a more frontal distribution. However, this explanation remains speculative.

Brunyé, T.T., Ditman, T., Mahoney, C.R., Augustyn, J.S., Taylor, H.A. (2009). When you and I share perspectives: pronouns modulate perspective taking during narrative comprehension. *Psychological Science*, 20(1), 27–32.

Buhrmester, M.D., Blanton, H., Swann, W.B., Jr (2011). Implicit self-esteem: nature, measurement, and a new way forward. *Journal of Personality and Social Psychology*, 100(2), 365–85.

Chambers, J.R., Windschitl, P.D. (2004). Biases in social comparative judgments: the role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin*, 130(5), 813–38.

Citron, F.M.M. (2012). Neural correlates of written emotion word processing: a review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, 122(3), 211–26.

Crites, S.L., Cacioppo, J.T., Gardner, W.L., Bernston, G.G. (1995). Bioelectrical echoes from evaluative categorization: II. A late positive brain potential that varies as function of attitude registration rather than attitude report. *Journal of Personality and Social Psychology*, 68(6), 997–1013.

Cross, P. (1977). Not can but will college teachers be improved. *New Directions for Higher Education*, 17, 1–15.

Delaney-Busch, N., Kuperberg, G.R. (2013). Friendly drug-dealers and terrifying puppies: affective primacy can attenuate the N400 effect in emotional discourse contexts. *Cognitive Affective & Behavioral Neuroscience*, 13(3), 473–90.

Dunning, D., Heath, C., Suls, J.M. (2004). Flawed self-assessment: implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3), 69–106.

Egidi, G., Nusbaum, H.C. (2012). Emotional language processing: how mood affects integration processes during discourse comprehension. *Brain and Language*, 122(3), 199–210.

Farnham, S.D., Greenwald, A.G., Banaji, M.R. (1999). Implicit self-esteem. In: Abrams, D., Hogg, M.A., editors. *Social Identity and Social Cognition*. London: Blackwell Publishing, pp. 230–48.

Fazio, R.H., Olson, M.A. (2003). Implicit measures in social cognition research: their meaning and use. *Annual Review of Psychology*, 54, 297–327.

Federmeier, K.D., Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4), 469–95.

Federmeier, K.D., Laszlo, S. (2009). Time for meaning: electrophysiology provides insights into the dynamics of representation and processing in semantic memory. *Psychology of Learning and Motivation*, 51, 1–44.

Federmeier, K.D., Wlotko, E.W., De Ochoa-Dewald, E., Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, 1146, 75–84.

Fields, E.C., Kuperberg, G.R. (2012). It's all about you: an ERP study of emotion and self-relevance in discourse. *NeuroImage*, 62(1), 562–74.

Fischler, I., Bradley, M. (2006). Event-related potential studies of language and emotion: words, phrases, and task effects. *Progress in Brain Research*, 156, 185–203.

Goldin, P.R., Jazaieri, H., Ziv, M., Kraemer, H., Heimberg, R.G., Gross, J.J. (2013). Changes in positive self-views mediate the effect of cognitive-behavioral therapy for social anxiety disorder. *Clinical Psychological Science*, 1(3), 301–10.

Greenhouse, S.W., Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24(2), 95–112.

Greenwald, A.G., Farnham, S.D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, 79(6), 1022–38.

Hajcak, G., MacNamara, A., Olvet, D.M. (2010). Event-related potentials, emotion, and emotion regulation: an integrative review. *Developmental Neuropsychology*, 35(2), 129–55.

Heine, S.J., Kitayama, S., Hamamura, T. (2007). Which studies test whether self-enhancement is pancultural? Reply to Sedikides, Gaertner, and Vevea, 2007. *Asian Journal of Social Psychology*, 10(3), 198–200.

Heine, S.J., Lehman, D.R., Markus, H.R., Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review*, 106(4), 766–94.

Helweg-Larsen, M., Shepperd, J.A. (2001). Do moderators of the optimistic bias affect personal or target risk estimates? A review of the literature. *Personality and Social Psychology Review*, 5(1), 74–95.

Henrich, J., Heine, S.J., Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83.

Herbert, C., Herbert, B.M., Ethofer, T., Pauli, P. (2011a). His or mine? The time course of self-other discrimination in emotion processing. *Social Neuroscience*, 6(3), 277–88.

Herbert, C., Pauli, P., Herbert, B.M. (2011b). Self-reference modulates the processing of emotional stimuli in the absence of explicit self-referential appraisal instructions. *Social Cognitive and Affective Neuroscience*, 6(5), 653–61.

Holcomb, P.J., Kounios, J., Anderson, J.E., West, W.C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *Journal of Experimental Psychology: Learning Memory and Cognition*, 25(3), 721–42.

Holt, D.J., Lynn, S.K., Kuperberg, G.R. (2009). Neurophysiological correlates of comprehending emotional meaning in context. *Journal of Cognitive Neuroscience*, 21(11), 2245–62.

Kanske, P., Kotz, S.A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research*, 1148, 138–48.

Kitayama, S., Park, J. (2010). Cultural neuroscience of the self: Understanding the social grounding of the brain. *Social Cognitive and Affective Neuroscience*, 5(2–3), 111–29.

Kitayama, S., Park, J. (2014). Error-related brain activity reveals self-centric motivation: culture matters. *Journal of Experimental Psychology: General*, 143(1), 62–70.

Koole, S.L., Pelham, B.W. (2003). On the nature of implicit self-esteem: the case of the name letter effect. In: Spencer, S., Fein, S., Zanna, M.P., editors. *Motivated Social Perception: The Ontario Symposium*, Vol. 9, Mahwah, NJ: Erlbaum, pp. 93–116.

Kuperberg, G.R. (2013). The proactive comprehender: what event-related potentials tell us about the dynamics of reading comprehension. In: Miller, B., Cutting, L., McCardle, P., editors. *Unraveling Reading Comprehension: Behavioral, Neurobiological, and Genetic Components*. Baltimore, MD: Paul H Brookes Publishing.

Kutas, M., Federmeier, K.D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–47.

Lau, E.F., Holcomb, P.J., Kuperberg, G.R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, 25(3), 484–502.

Lee, C.L., Federmeier, K.D. (2008). To watch, to see, and to differ: an event-related potential study of concreteness effects as a function of word class and lexical ambiguity. *Brain and Language*, 104(2), 145–58.

Leon, I., Diaz, J.M., de Vega, M., Hernandez, J.A. (2010). Discourse-based emotional consistency modulates early and middle components of event-related potentials. *Emotion*, 10(6), 863–73.

Li, W., Han, S. (2010). Perspective taking modulates event-related potentials to perceived pain. *Neuroscience Letters*, 469(3), 328–32.

Luck, S.J. (2014). *An Introduction to the Event-Related Potential Technique*, 2nd edn. Cambridge, MA: The MIT Press.

Lund, K., Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods: Instruments & Computers*, 28(2), 203–8.

Markus, H., Kunda, Z. (1986). Stability and malleability of the self-concept. *Journal of Personality and Social Psychology*, 51(4), 858–66.

Na, J.Y., Kitayama, S. (2011). Spontaneous trait inference is culture-specific: behavioral and neural evidence. *Psychological Science*, 22(8), 1025–32.

Olson, M.A., Fazio, R.H., Hermann, A.D. (2007). Reporting tendencies underlie discrepancies between implicit and explicit measures of self-esteem. *Psychological Science*, 18(4), 287–91.

Paczynski, M., Kuperberg, G.R. (2012). Multiple influences of semantic memory on sentence processing: distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, 67(4), 426–48.

Paulhus, D.L. (1993). Bypassing the will: the automatization of affirmations. In: Wegner, D.M., Pennebaker, J.W., editors. *Handbook of Mental Control*. Hillsdale, NJ: Psychology Press, pp. 573–87.

Paulhus, D.L. (2002). Socially desirable responding: the evolution of a construct. In: Braun, H.I., Jackson, D.N., Wiley, D.E., editors. *The Role of Constructs in Psychological and Educational Measurement*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers, pp. 49–69.

Pratt, N.L., Kelly, S.D. (2008). Emotional states influence the neural processing of affective language. *Social Neuroscience*, 3(3–4), 434–42.

Sears, D.O. (1983). The person-positivity bias. *Journal of Personality and Social Psychology*, 44(2), 233–50.

Sedikides, C., Gaertner, L., Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality and Social Psychology*, 84(1), 60–79.

Sedikides, C., Gaertner, L., Vevea, J.L. (2007). Evaluating the evidence for pancultural self-enhancement. *Asian Journal of Social Psychology*, 10(3), 201–3.

Shestyuk, A.Y., Deldin, P.J. (2010). Automatic and strategic representation of the self in major depression: trait and state abnormalities. *American Journal of Psychiatry*, 167(5), 536–44.

Sitnikova, T., Kuperberg, G., Holcomb, P.J. (2003). Semantic integration in videos of real-world events: an electrophysiological investigation. *Psychophysiology*, 40(1), 160–4.

Swaab, T.Y., Baynes, K., Knight, R.T. (2002). Separable effects of priming and imageability on word processing: an ERP study. *Cognitive Brain Research*, 15(1), 99–103.

Taylor, S.E., Brown, J.D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin*, 103(2), 193–210.

Van Berkum, J.J.A., Holleman, B., Nieuwland, M., Otten, M., Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science*, 20(9), 1092–9.

Varnum, M.E.W., Na, J.Y., Murata, A., Kitayama, S. (2012). Social class differences in N400 indicate differences in spontaneous trait inference. *Journal of Experimental Psychology: General*, 141(3), 518–26.

Watson, L.A., Dritschel, B., Obonsawin, M.C., Jentzsch, I. (2007). Seeing yourself in a positive light: brain correlates of the self-positivity bias. *Brain Research*, 1152, 106–10.

West, W.C., Holcomb, P.J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, 13(3), 363–75.

White, K.R., Crites, S.L., Jr, Taylor, J.H., Corral, G. (2009). Wait, what? Assessing stereotype incongruities using the N400 ERP component. *Social Cognitive and Affective Neuroscience*, 4(2), 191–8.