

Electrophysiological time course and brain areas of spontaneous and intentional trait inferences

Marijke Van Duynslaeger,¹ Frank Van Overwalle,¹ and Edwin Verstraeten²

¹Vrije Universiteit Brussel, Belgium, and ²Swansea University, UK

This study measured event-related potentials during spontaneous and intentional trait inferences. Participants read sentences describing the behavior of a target person from which a strong moral trait could be inferred. The last word of each sentence determined the consistency with the trait induced during an introductory paragraph. In comparison with behaviors that were consistent with the implied trait, a P300 waveform was obtained when the behaviors were evaluative inconsistent with that trait. This dependency on behavioral consistency indicates that trait inferences were made previously while reading the preceding behaviors, irrespective of the participants' spontaneous or intentional goals. Overall, the P300 shows considerable parallels between spontaneous and intentional inferences, indicating that the type and timing of the inconsistency process is very similar. In contrast, source localization (LORETA) of the event-related potentials suggest that spontaneous inferences show greater activation in the temporo-parietal junction compared to intentional inferences following an inconsistency. Memory measures taken after the presentation of the stimulus material involved sentence completion and trait-cued recall, and supported the occurrence of trait inferences associated with the actor. They also showed significant correlations with the neural components (i.e. P300 and its current density at the temporo-parietal junction) predominantly following spontaneous instructions, indicating that these components are valid neural indices of spontaneous inferences.

Keywords: spontaneous trait inferences; ERP; LORETA; temporo-parietal junction; medial prefrontal cortex

Forming impressions about others is a never ending story. When we meet novel people, we often have to make corrections and revisions of our first impressions. Unfortunately, even after a long-lasting romantic relationship or friendship, we are sometimes struck by unexpected and disappointing behaviors that oblige us to re-evaluate our beloved others. What are the cognitive processes underlying the detection of such discrepancies? Do they reflect automatic responses or more controlled thought to resolve the inconsistency with what we know about the other person? A major goal of this research is to explore whether the processing of such inconsistencies is similar for initial trait expectations made spontaneously while performing other tasks or intentionally when forming explicitly an impression about a person (Bargh, 1989), and how electrophysiological measures can shed some more light on the neural components associated with each. If the brain responds differentially to information that confirms or violates expectancies, this neural response can be used as a novel index for the occurrence of trait expectancies developed previously, either spontaneously or intentionally. Moreover, research into such neural responses can increase our understanding of

the interplay between top-down expectancies and bottom-up processing of novel behavioral information.

From the broad perspective of dual-process models which posit that information processing involves either automatic associative processes or controlled symbolic reasoning in general (Sloman, 1996; Smith and DeCoster, 2000) and in person perception in particular (Brewer, 1988; Fiske and Neuberg, 1988; Gilbert *et al.*, 1988; Van Overwalle and Labiouse, 2004; Satpute and Lieberman, 2006), many authors have argued that the evolutionary older and more spontaneous associative processes subserve and guide intentional thoughts. It has been proposed that automatic associative processes occur first and that symbolic reasoning follows optionally, or that both processing modes occur simultaneously. If the latter is true, one should expect substantial parallels not only in the type of processes but also in their timing and localization. Event-related brain potential (ERP) measures are ideally suited to answer these questions, as they provide millisecond accuracy in the timing of brain processes and recent techniques allow localizing the source of their activity (e.g. LORETA, Pascual-Marqui *et al.*, 2002).

INTENTIONAL AND SPONTANEOUS TRAIT INFERENCE

In contrast to intentional trait inferences (ITI) that are made with the explicit goal to form an impression about a target, spontaneous trait inferences (STI) are formed without intention or awareness (Uleman *et al.*, 1996; Uleman, 1999; Uleman *et al.*, 2005). STI are relatively automatic in the sense that they require little mental effort, are difficult to suppress

Received 11 April 2007; Accepted 11 April 2007

Advance Access publication 17 May 2007

This research was supported by an FWO Grant of the Research Foundation - Flanders to Frank Van Overwalle and a VUB-OZR grant to Edwin Verstraeten. We are very grateful to Bruce Bartholow for providing us his experimental stimulus material, and to Els Somers for conducting the experiment.

Correspondence should be addressed to Frank Van Overwalle, Department of Psychology, Vrije Universiteit Brussel, Pleinlaan 2, B - 1050 Brussel, Belgium. E-mail: Frank.VanOverwalle@vub.ac.be.

and are hard to interfere with, although limited mental resources can reduce STI for counter-stereotypical behaviors (Wigboldus *et al.*, 2004) and some goals can reduce STI, such as deciding whether the communicated information is false or not (Skowronski *et al.*, 1998). Inducing ITI is straightforward, as it simply involves giving the participants the instruction to make impressions about a target, and measuring trait inferences using traditional rating scales. However, measuring STI involves more covert and implicit tasks, because participants should not be aware that impressions are to be formed and are measured.

Various psychological measures have been developed to demonstrate the occurrence of STI. One of earliest paradigms was *cued recall*, developed by Winter and Uleman (1984). When people make STI while observing or reading about a behavior, the inferred traits are assumed to be stored in memory together with the behavioral information from which they are inferred. As a result, these traits are effective retrieval cues for the behavioral information. Winter and Uleman (1984) found that trait cues are stronger aids to recall than semantic cues that were a priori related with the actor or other sentence parts, or when no cue is given (non-cued recall). A disadvantage of the cued recall measure is that it is unclear whether the trait reflects an impression of the actor or only an interpretation of the behavior (McKoon and Ratcliff, 1986; Carlston and Skowronski, 1994). This latter limitation was largely overcome in recent memory tasks, such as *relearning* (Carlston and Skowronski, 1994) and *false recognition* (Todorov and Uleman, 2002), which measure the link between a photo of the actor and the implied trait, using more implicit measures (i.e. facilitation while relearning and recognition errors, respectively). This research confirmed that trait-cued recall is indeed a valid measure of STI.

In addition to STI research, it has been shown that behavior that is inconsistent with one's expectations (e.g. impressions, stereotypes) about a person is often recalled better than consistent behavior (Hastie and Kumar, 1979; Hastie, 1980; Srull & Wyer, 1989; Stangor and McMillan, 1992). It is generally assumed that such unexpected information receives more cognitive processing and is therefore recalled better than expected behavior. Research has, however, established that if inconsistencies are not tied to an expectation about a person, then no inconsistency resolution processes emerge. That is, when an impression or stereotype is formed on a group of loose individuals rather than a single individual, there seems to be decreased rather than enhanced memory for stereotype-inconsistent behaviors (for a review, see Fyock and Stangor, 1994). This implies that recall of inconsistent behaviors can be considered a measure of person inferences established previously, if perhaps not in terms of specific traits, at least in terms of one's general valenced impression of a person (e.g. positive or negative).

It is important to note that earlier research has documented that impression judgments are more influenced by an actor's negative behaviors than positive behaviors

(Cacioppo *et al.*, 1999; Ybarra, 2002). Thus, negative behaviors that disconfirm a trait expectancy are more effective at changing impressions than positive disconfirming behaviors. The most dominant explanation in the literature rests on the asymmetric diagnosticity of negative *vs* positive social behaviors (Reeder and Brewer, 1979; Skowronski and Carlston, 1989; Eiser *et al.*, 2003). In social domains, when making inferences about honesty and kindness, negative behaviors are more diagnostic because moral actors are constrained to perform only moral behaviors (e.g. an honest person never cheats), while immoral persons are free to display either moral or immoral behaviors. The reverse is true for achievement domains. For inferences about intelligence or athleticism, positive behaviors are more diagnostic. Actors high in ability can exhibit either low or high performance (e.g. even the best athlete does not win all competitions), while actors low in ability are constrained to low performance only (Reeder and Fulks, 1980; Reeder and Spores, 1983; Skowronski and Carlston, 1987; Reeder, 1997; Lupfer *et al.*, 2000). An additional question of this research is whether this positive-negative asymmetry will be also revealed for ERP components of trait inferences.

ERP COMPONENTS

Traditional implicit memory measures of STI and self-report measures of ITI are limited in that they do not allow exploring the type and timing of process differences related to the detection and resolution of inconsistencies with traits inferred previously. Do such inconsistencies induce immediate automatic reactions or are they resolved given more controlled thought? Are they operating at early stages or at later semantic comprehension and resolution stages, and do these timings differ between STI and ITI? These issues can be more closely examined by using ERPs. As far as we are aware, such measures have not been used previously for detecting STI, although they have been applied under ITI instructions by Bartholow and his colleagues (Bartholow *et al.*, 2001; Bartholow *et al.*, 2003). Likewise, reliable techniques to localize the source of ERPs have recently emerged and applied to social phenomena (Esslen *et al.*, 2004), but not yet on STI. Although the spatial resolution of ERP waves is poorer than functional magnetic resonance imaging (fMRI), a prime advantage is that their high time resolution allows localizing and imaging the electric activity in a millisecond range, unlike fMRI which offers a time resolution in the range of a few seconds.

P300 and evaluative inconsistency

ERPs are waveforms that reflect electric activity of the brain during responses to specific stimuli. They are manifestations of information processing activities, and different types of ERP components are associated with different functions in this process. Research has revealed two main types of ERPs which index inconsistency detection. The first type contains the N200 which react to early and automatic response

conflict (Botvinick *et al.*, 2004). The second type contains the P300 which occurs later and responds to inconsistencies in comprehension. This characteristic makes the P300 ideally suited for exploring the neural correlates of inconsistency resolution during trait inferences.

The P300 is a late positive peak that typically initiates around 300 ms after the critical stimulus and continues till 600 or 1000 ms, and attains the highest amplitudes mainly at parietal and central scalp locations (Andreassi, 2000). It is suggested that the P300 originates from working memory activity in the middle frontal and parietal lobe as seen with brain imaging techniques (McCarthy *et al.*, 1997). Research has documented that there is a relation between the P300 and the processing of anomalous, inconsistent or infrequent stimuli presented in a context of otherwise normal or frequent information, as long as this information is relevant for the task. The amplitude of the P300 increases as a function of the amount of discrepancy between the stimulus and the preceding context, and correlates with later recall of the discrepant stimuli, especially when elaborate rehearsal strategies are minimized (Fabiani *et al.*, 1986; Fabiani and Donchin, 1995; Andreassi, 2000). These findings have led to the view that the P300 is an index of on-line updating of working memory after inconsistency detection.

P300 in person perception

In social research, Cacioppo and coworkers (Cacioppo *et al.*, 1993; Cacioppo *et al.*, 1994) found that evaluative inconsistency between a trait word and previously presented trait words (e.g. a negative trait after a sequence of positive traits) elicited a large P300 between approximately 500 and 1000 ms at central and parietal scalp locations. However, in this research traits were given rather than inferred. Of more interest for our research, Bartholow and colleagues (Bartholow *et al.*, 2001, 2003, 2006) instructed their participants to form impressions about actors engaged in several behaviors depicted in short sentences. The behaviors were either consistent with traits implied during preceding behaviors (e.g. a friendly act after a sequence of courteous behaviors), or were opposite in valence with the implied traits (e.g. an impolite act). Meanwhile ERPs were recorded in order to track the neural activity associated with inconsistency resolution and to examine how this activity relates to later recall. Bartholow *et al.* (2001, 2003) found greater P300 activation at 300–800 ms after presenting the critical word for trait-inconsistent sentences as opposed to trait-consistent sentences. Because a P300 indicates the detection of a violation of an expectancy generated by the previous stimulus sequence history, the enhanced P300 amplitude for the trait-inconsistent sentences can be interpreted as indicating that traits had been inferred earlier. Of interest is to note that Bartholow *et al.* (2003) found this P300 effect only after discrepant negative behaviors following a positive trait expectancy, consistent with earlier research on the higher diagnosticity of negative social behaviors

(Reeder and Brewer, 1979; Skowronski and Carlston, 1989; Ybarra, 2002; Eiser *et al.*, 2003) and stronger ERPs given negative inconsistencies embedded in series of positive trait words (for a review, see Cacioppo *et al.*, 1999). The increase in P300 was paralleled by enhanced memory performance for trait-inconsistent sentences in comparison with trait-consistent sentences on a sentence completion task. As noted earlier, a memory advantage for inconsistent information in person perception is often explained in terms of deeper processing and greater cognitive activity required to reconcile the inconsistent information with an already formed person impression (Hastie and Kumar, 1979; Hastie, 1980; Srull and Wyer, 1989; Stangor and McMillan, 1992).

Scalp locations of the P300

Recent fMRI evidence suggests that two brain areas involved in the understanding and attribution of mental states (i.e. goals and traits) of others are the *temporo-parietal junction* (TPJ) and the *medial prefrontal cortex* (mPFC; see Frith and Frith, 2001; Harris *et al.*, 2005; Saxe, 2006). Research seems to indicate that the TPJ is mainly involved in the attribution of temporary intentions, desires and beliefs by others (Saxe and Wexler, 2005; Saxe, 2006; Saxe and Powell, 2006), while the mPFC seems to be more essential in attributing enduring traits (e.g. Mitchell *et al.*, 2005; Todorov *et al.*, 2006). It should be noted, however, that these trait studies mostly involve intentional judgments. A recent study by Mitchell *et al.* (2006) compared intentional with spontaneous (i.e. memory) instructions while participants were scanned using fMRI, but these instructions were alternated between trials so that it is unlikely that the 'spontaneous' trait inferences were made without any awareness and intention. To summarize, previous ERP research suggests that a P300 following inconsistent information is most likely to be found in parietal and central scalp regions, and earlier fMRI imaging research suggests that the main brain activity during trait inferences is localized in the TPJ and mPFC.

PRESENT RESEARCH

For the present research, we borrowed Bartholow *et al.*'s (2001, 2003) paradigm in which short behavioral descriptions were provided that were either consistent or inconsistent with the traits implied during preceding behaviors. We extended this paradigm to explore intentional and spontaneous impression formation, by instructing our participants either to form impressions about each target (ITI) or to read the stimulus material carefully, without mentioning anything about person traits or impressions (STI). We also made other modifications in order to pursue a number of additional theoretical questions.

First, all the traits implied in Bartholow *et al.*'s (2001, 2003) research involved moral and immoral behaviors, which captures only one (i.e. social) dimension of personality. To study the breadth of the inconsistency process,

we additionally investigated behaviors involving a descriptively different dimension of personality (Stangor and McMillan, 1992), in particular, high or low competence. To the extent that spontaneous or intentional inferences generalize to evaluative inconsistencies across different domains of personality traits (e.g. a moral person performing badly in an exam or in sports), these competence-related behaviors should also elicit a P300.

Second, we took several memory measures as behavioral validation of the ERP measures, taken after the presentation of all stimulus material so that they did not interfere with the ongoing EEG measures. It is important to realize that these memory measures take a somewhat different interpretation than in earlier STI research because of some essential differences. The first, a sentence completion task, borrowed from Bartholow *et al.* (2001, 2003), consists of completing the last, trait-implicating word of the original sentence. This task has not been used earlier in STI studies, because each behavior in these studies typically involved a different actor so that trait inconsistencies were absent in the material. In contrast, in the present paradigm, the behaviors are not independent but form a collection performed by the same actor. Hence, increased memory in the sentence completion task suggests a violation with an expectation formed earlier about the actor (Hastie and Kumar, 1979; Hastie, 1980; Srull & Wyer, 1989; Stangor and McMillan, 1992), although this expectation might reflect a general person evaluation rather than a specific and long-lasting trait inference. The second, a trait-cued recall task (Winter and Uleman, 1984), additionally verifies whether traits are inferred. Enhanced recall cued by the implied trait suggests that trait interpretations were made following consistent behaviors, rather than being merely evaluative interpretations. This interpretation should be strengthened if these recall scores are equally strong for STI and ITI, as in this latter condition explicit trait impressions are made about the actor.

HYPOTHESES

Our general position is that both spontaneous and intentional instructions lead to trait inferences about the actor's prevailing behaviors, if not about the actor him or herself. Moreover, in line with dual-process models, we expect many parallels between STI and ITI, so that trait inferences will lead to activity in the same broad neural network, although we also anticipate some smaller differences between STI and ITI. In particular, we make the following predictions.

First, based on earlier ERP research and in particular recent work by Bartholow and colleagues (2001, 2003), we hypothesize that evaluative inconsistencies generate P300 waveforms mainly at parietal or central scalp locations. In addition, because of the positive-negative asymmetry, we also predict that negative inconsistencies given a positive trait expectancy will lead to more discrepancies and larger P300 waveforms than positive inconsistencies. Based on earlier research (Reeder and Brewer, 1979; Skowronski and

Carlston, 1989), this bias is expected for inconsistencies of the induced traits that are of moral nature, and not for competence-related inconsistencies (because they contain a mixture of both, e.g. a moral person failing an exam). In addition, based on earlier fMRI research, we expect that trait inferences activate several distinct brain areas such as the TPJ and the mPFC at different periods of time. Given that the mPFC is part of the frontal cortex involved in controlled processing and action monitoring (Miller and Cohen, 2001; Satpute and Lieberman, 2006) and because this brain region has been implicated mainly in research on ITI (Mitchell *et al.*, 2005; Todorov *et al.*, 2006), we expect it to be particularly active during intentional trait inferences (see also Saxe and Powell, 2006). In contrast, because there is some fMRI evidence that the TPJ is involved in social attributions of intention and thoughts made spontaneously while reading stories about persons (Saxe and Kanwisher, 2003; Saxe and Powell, 2006) we expect that the TPJ will be more active during spontaneous inferences.

Second, with respect to the behavioral memory measures in support of the occurrence of trait inferences, we expect better memory for trait-inconsistent behaviors at the sentence completion task, as this would indicate that a prior expectation is built about the actor's prevailing (trait-consistent) behaviors. In addition, we expect better memory for trait-consistent behaviors after recall cued with the induced trait, indicating that these expectations are represented in terms of traits following consistent behaviors. Note that our memory predictions differ between cued recall (better for consistent information) and sentence completion (better for inconsistent information). Because sentence retrieval is aselective, it benefits from general enhanced memory due to inconsistency detection. In contrast, cued recall is selective with the aid of the trait as a cue, and therefore should be enhanced when that trait was actually inferred during information uptake.

Finally, if the P300 reflects inconsistency processing, we expect a correlation between the P300 and sentence completion, because this memory task specifically tracks the process of inconsistency resolution. Research suggests that this correlation is probably apparent mainly for STI where elaborate rehearsal is minimized, because the individual differences in thoughts during ITI may mask ERP measures (which are averages across individuals; see Fabiani *et al.*, 1986; Fabiani and Donchin, 1995; Andreassi, 2000). In contrast, we expect LORETA current density activation (Pascual-Marqui *et al.*, 2002) in the TPJ or mPFC to correlate with trait-cued recall following trait-consistent information, because trait-cued recall measures the strength or consolidation of a trait.

To recapitulate, we expect few essential timing differences between STI and ITI. For both instructions, we expect P300 waveforms at parietal or central scalp locations, especially following negative inconsistencies. We expect, however,

some differences in brain localizations, so that STI involve mainly the TPJ whereas ITI involve the mPFC. We predict better sentence completion memory for *inconsistent* behavioral information and a correlation with the P300, and better trait-cued recall for consistent information and a correlation with LORETA activation.

METHOD

Participants

Participants were 50 students at the Vrije Universiteit Brussel (VUB), without prior history of any neurological dysfunction. Due to equipment failure, three participants were removed and another participant was removed because she failed to attend to the whole experiment. Of the remaining 46 participants, there were 37 women and 13 men, with an age varying between 19 and 37. All participants were recruited via a university-wide electronic mailing system for all university students. In exchange for their participation, they were paid 20 euros. Among all participants, 23 received a typical STI-instruction to read the sentences (16 women and 7 men), while another 23 participants received a typical ITI instruction to form a trait impression about the actor (18 women and 5 men).

Stimulus material

The design and stimulus material were borrowed from Bartholow *et al.* (2001, 2003) with some important modifications. Participants read 20 introductory paragraphs that described the general behavior of a fictitious target person and from which a strong trait could be inferred. The paragraphs involved 10 positive and 10 negative moral traits, and each paragraph was shown for 30 s on the computer screen. To avoid association with a familiar and/or existing name, fictitious 'Star Trek'-like names were used (Bartholow *et al.*, 2001, 2003). For example, the next paragraph described the general behavior of target person 'Tolvan': *Tolvan smiles at everyone on the way to work. Whenever it snows, Tolvan shovels her elderly neighbor's walk. Tolvan always stops to help when she sees someone with car trouble. Tolvan's coworkers are all quite fond of her.* This paragraph implies that 'Tolvan' is a friendly person. After each paragraph, a series of 12 behavioral sentences was presented, each consisting of six words shown in the center of the computer screen. Every 350 ms a word was presented for 300 ms. The last word of each sentence was the critical one, because it determined the degree of consistency with the previously inferred trait: trait-consistent (TC), trait-inconsistent (TI), competence-inconsistent (CI) and irrelevant (IRR). TC-sentences describe moral behaviors that are consistent with the inferred trait with respect to valence (for example 'Tolvan gave her sister a *hug*'). TI-sentences (for example 'Tolvan dared the stranger to *fight*') are inconsistent with the inferred trait with respect to valence (here 'friendly'). The CI-sentences describe

competence-related behaviors that are inconsistent with the inferred trait in regard to valence and descriptive content of the trait (for example 'Tolvan obtained for math an *F*'). The IRR-sentences describe neutral behaviors (for example 'Tolvan gave her mother a *bottle*').

After each introductory paragraph, a series of 12 behavioral sentences was presented. These consisted of four filler sentences, always ending with TC-behavior, followed by the eight experimental sentences, consisting of 2 TC-, 2 TI-, 2 CI- and 2 IRR-behaviors presented in a random order. All the material was borrowed from Bartholow *et al.* (2001, 2003) and translated from English into Dutch while keeping the same amount of words (which sometimes required us to develop different sentences implying the same trait), except for the CI sentences which were all developed in Dutch. All Dutch sentences were pilot tested ($n=199$) to check if they reflected the expected degree of consistency with the inferred trait, and results showed the expected rank order of the ratings (on a 0–10 scale) from TC, TI, CI to IRR-sentences for all traits. A similar pilot test ($n=83$) was performed for the novel CI sentences to check whether they reflected the intended low or high competence, and the results revealed that CI-sentences reflecting high competence received higher ratings ($M=8.0$ on a 0–10 scale) than those reflecting low competence ($M=3.7$).

Electrophysiological registration and analysis

An average reference EEG was recorded with the Easy-Cap electrode system (Falk Minow Services, Munich) from 19 sites according to the international 10–20 electrode system. A bipolar horizontal electrooculogram (EOG) was recorded from the epicanthus of each eye, and a bipolar vertical EOG was recorded from supra- and infra-orbital positions of the right eye. The EEG and the EOG were recorded with Ag/AgCl electrodes. Prior to the placement of electrodes, the expected electrode sites were gently abraded. All impedances of the EEG electrodes were below 10 k Ω , and the differences in impedance between sites were below 2 k Ω . The EEG was sampled at a rate of 256 Hz. The 250 ms just before the presentation of the last critical word in each sentence served a prestimulus baseline, and the recording till 1250 ms after the presentation of the critical word served for analysis. Recording and analyses were done with software from Advanced Neuro Technology (EEvoke and ASA). The raw EEG data were filtered by a 0.03–30 Hz band-pass, EOG artifacts were corrected using the SOBI algorithm (Belouchrani *et al.*, 1993) and remaining artifacts beyond -75 and $75 \mu\text{V}$ were removed before analysis. To identify ERPs, the EEG data were averaged per participant, channel and condition. A grand average was then computed on the ERP data across all participants.

Procedure

The participant was seated in a dimly lit EEG room. EEG and EOG were recorded during participants' reading.

Table 1 Mean amplitude (in μ Volts) of the maximum ERP peaks as a function of Instruction, location, consistency and time interval

Location	Positive trait context									Negative trait context								
	Consistent			Inconsistent			P-level			Consistent			Inconsistent			P-level		
	300	450	650	300	450	650	300	450	650	300	450	650	300	450	650	300	450	650
<i>Evaluative TC-TI comparisons</i>																		
<i>Spontaneous trait inferences</i>																		
Cz	1.88	2.49	3.88	2.44	4.45	4.77	—	0.000	0.051	1.87	3.58	4.56	2.75	4.08	5.03	—	—	—
Pz	2.90	2.94	4.39	3.88	4.06	5.65	0.010	0.029	—	3.25	3.07	5.22	3.29	3.72	4.85	—	—	—
<i>Intentional trait inferences</i>																		
Cz	3.44	4.43	5.21	3.74	5.60	5.99	—	0.006	—	2.85	3.90	4.68	3.65	4.48	5.93	—	—	—
C4	1.83	1.95	3.11	2.65	3.53	3.86	0.048	0.001	—	2.90	3.20	4.26	2.52	2.70	3.87	—	—	—
<i>Descriptive TC-CI comparisons</i>																		
<i>Spontaneous trait inferences</i>																		
Cz	1.88	2.49	3.88	2.98	3.72	3.43	0.041	0.045	—	1.87	3.58	4.56	3.89	4.03	4.58	0.011	—	—
C4	2.14	3.06	3.77	2.38	3.09	2.79	—	—	—	1.10	1.73	2.47	3.91	4.13	4.24	0.000	0.004	0.039
T6	2.07	0.84	1.03	0.59	-0.16	0.58	—	—	—	2.02	0.43	0.40	2.32	1.90	2.36	—	—	0.001
<i>Intentional trait inferences</i>																		
C4	1.83	1.95	3.11	3.22	3.71	3.96	0.003	0.014	—	2.90	3.20	4.26	4.25	4.24	4.09	—	—	—

Note: Intervals are denoted by their starting ms. Only P-levels (of *t*-tests) for positive peaks are shown. Channels are ordered from left over central to right hemisphere and from anterior to posterior sites.

The instructions were presented on a computer screen. The participants were informed that they would read stories about several persons and that each story would start with a paragraph followed by different sentences about it. Because participants were shown each word at a fixed pace of 300 ms, they were also told to pay as much attention as possible to each word, because they would get questions about them afterwards. For the STI instruction, participants were told to read as attentively as possible (see also Todorov and Uleman, 2002). For the ITI instruction, participants were asked to form a trait impression about the actor in the introductory paragraph and the sentences (see Bartholow *et al.*, 2001). It was also emphasized to move and eye-blink as little as possible to limit artifacts in the EEG (Stern *et al.*, 2001).

After the computer reading task was finished, the electrodes for the EEG and EOG were removed. Next, the participants were given the cued recall and the sentence completion task in the same order for all participants. In the cued recall task, participants had to write as much behavioral sentences as possible with the aid of words that consisted of the implied traits. There were trait cues for all 20 series of behavioral sentences. In the sentence completion task, participants were presented with incomplete TC-, TI- and CI-sentences and had to complete the last word.

RESULTS

ERP data

To analyze the time course of positive and negative peaks in the ERP waveforms and associated cognitive activity, we divided the ERP data of each participant in several time intervals (50–300, 300–450, 450–600 and 600–1000 ms).

The largest positive and negative peaks (with maximum and minimum amplitude, respectively) in each interval were identified and statistically analyzed for each of the 19 channels separately by means of a Repeated Measures Analysis of Variance (ANOVA) with Trait Context (positive, negative), Consistency (TC, TI, CI, IRR) and Interval (50–300, 300–450, 450–600, 600–1000) as within-participants factors and Instruction (read, impression) as between-participants factor.

The ANOVA revealed for many channels a main effect for Interval, $F(3, 132) = 4.23-36.51$, $P_s < 0.01$, as well as significant main or interaction effects with Consistency, $P_s < 0.01$. To gain further insight in these effects, and in order to test our specific hypothesis concerning inconsistency resolution, we conducted simple *t*-tests for evaluative (TC vs TI) and descriptive (TC vs CI) inconsistencies separately for a positive and negative trait context, and for STI and ITI. In order to control for multiple comparisons, our analytical strategy was as follows. To test our specific hypotheses with respect to the P300, we used a conventional 0.05 α -level for the central (Cz) and parietal (Pz) scalp locations at the midline. The results are shown in Table 1 (No differences were found in the 50–300 interval at the 0.05 level for none of the channels and peaks, so that these results are further ignored and not reported in Table 1). To explore other possible locations, we adjusted the α -level to 0.000045 using the Bonferroni correction, which takes into account all cells of the ANOVA (64) and all remaining channels (17). None of the differences at other scalp locations survived this stringent α -level. For illustrative purposes, however, these differences exceeding the non-significant 0.01 level are reported in Table 1, together with adjacent time intervals that exceed the non-significant 0.05 level.

Evaluative trait inconsistencies. We first focus on the differences between TC and TI sentences, which reflect evaluative inconsistencies with the inferred trait. Our hypothesis for evaluative inconsistencies is that they generate P300 waveforms mainly at parietal or central scalp locations. Given the positive–negative asymmetry in trait attributions, we also suggested that negative inconsistencies in a positive trait context probably generate more evaluative discrepancies and therefore more P300 waveforms than positive inconsistencies in a negative trait context. The results were generally in accordance with our predictions. Given that our hypotheses focused on inconsistency resolution, the results were first explored using *t*-tests comparing TC and TI sentences, separately for STI and ITI.

For STI embedded in a positive trait context, as predicted, *t*-tests at the Cz scalp location revealed that TI sentences generated a greater P300-like positivity in comparison with TC sentences, which peaked at the 450–600 ms interval and continued until 1000 ms. Likewise, at the Pz location, TI sentences generated a similar P300-like waveform that peaked earlier at the 300–450 ms interval and that was further sustained for the 450–600 ms interval. Figure 1A shows that inconsistent behaviors (dark line) elicited a larger positive ERP at the Cz scalp location than consistent behaviors (light line). In a negative trait context, no such differences were found, as predicted.

For ITI embedded in a positive trait context, we found similar results. The *t*-tests revealed that TI sentences generated a greater P300-like positivity than TC sentences at the Cz scalp location (but not at the Pz location) that peaked at the 450–650 ms interval. Table 1 shows that at the right hemispheric C4 site, there was a similar P300 waveform that, however, did not exceed the α -level after Bonferroni correction. Figure 1B shows that inconsistencies (dark line) elicited a larger positive ERP at the Cz scalp location than consistent behaviors (light line). In a negative trait context, again no significant differences were found. To confirm that the ERP inconsistency effects were similar for STI and ITI instructions, we conducted additional analyses on the ERP amplitudes at the predicted Cz and Pz scalp locations. Corroborating the prior analyses, none of TC–TI differences in ERP amplitude was significant between STI and ITI, all $t(44) < 1.87$, *ns*.¹

¹ These results were further statistically analyzed by an ANOVA on each of the Cz and Pz channels with Instruction (read, impression) as between-participants factor and Consistency (TC, TI), Context (positive, negative) and Interval (300–450, 450–650, 650–1000) as within-participants factors, and the results generally confirmed the prior analyses and also revealed some additional effects of interest. For the Cz channel, the ANOVA confirmed the predicted main effect of Consistency indicating that inconsistent information resulted in more positive ERP amplitudes than consistent information, $F(1, 44) = 11.49$, $P < 0.01$, and an additional main effect of Instruction indicating that STI resulted in smaller ERP amplitudes than ITI, $F(1, 44) = 5.75$, $P < 0.05$. The critical interaction between Consistency and Instruction was, however, not significant. There was also a robust main effect of Interval that appeared in all subsequent analyses, and which indicated that the ERP amplitudes were generally higher at larger intervals, $F(2, 88) = 32.70$, $P < 0.001$. A significant triple interaction of these two factors with Context, $F(2, 88) = 5.79$, $P < 0.01$, confirmed the pattern of *t*-tests for the Cz channel. Likewise, for the Pz channel, the ANOVA revealed an analogous main effect of Interval, $F(2, 88) = 23.87$, $P < 0.001$, and an analogous triple interaction, $F(2, 88) = 3.28$, $P < 0.05$. No other effects were significant.

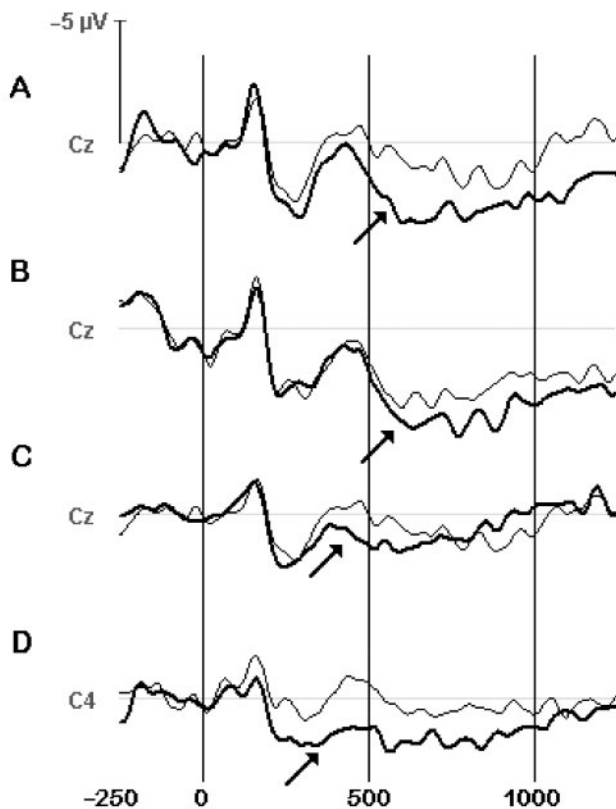


Fig. 1 Effects of inconsistency on grand-averaged ERP waveforms showing a P300 in a positive trait context at the Cz and C4 scalp sites for evaluative trait inconsistencies (TC vs TI) given (A) STI and (B) ITI; and for descriptive trait inconsistencies (TC vs CI) given (C) STI and (D) ITI. Dark lines denote inconsistencies and light lines denote consistencies. The timeline is given in ms. A positive amplitude is shown downward. The arrow indicates the onset of the P300.

Taken together, these findings seem to suggest that prior trait expectations were developed spontaneously and intentionally, before inconsistent information was presented. Once presented, negative inconsistencies with (positive) trait expectations prompted an inconsistency detection process as indexed by P300, while positive inconsistencies with (negative) trait expectations generated little change in brain activity. This is in line with earlier research and our hypothesis on the asymmetry between negative and positive moral behaviors. Thus, only immoral behaviors that violated a positive trait expectancy seem to prompt a reevaluation of the actor, either spontaneously or intentionally.

Descriptive trait inconsistencies. We now turn to the differences between TC and CI sentences, which besides evaluative inconsistencies also reflect a shift in personality domain (see also Table 1). Our main hypothesis is that because they involve evaluative inconsistencies, they also should generate a P300. Given that descriptive trait inconsistencies contain a mixture of moral and achievement-related behaviors (in the TC and CI conditions, respectively), less positive–negative asymmetry was expected.

For STI, as predicted, we found that CI sentences elicited P300 waves that were more positive than TC sentences at the

Cz site, and significantly so in a positive trait context at the 300–650 interval, and at the earlier 300–450 interval in a negative trait context. Table 1 depicts similar P300 components at the C4 location (right central cortex) and T6 location (at the right posterior temporal lobe) which did not survive the Bonferroni correction for significance. Figure 1C shows that negative inconsistencies (dark line) in a positive context elicited a larger positive ERP at the Cz scalp location than consistent behaviors (light line).

For ITI, we did not find the predicted P300-like waveform in the Cz or Pz locations. Of interest, however, is that given a positive context, the CI sentences generated a greater P300-like positivity than TC sentences at the C4 scalp location (Figure 1D), which did not exceed the significance level after Bonferroni adjustment. There were no differences given a negative trait context. To confirm that the ERP inconsistency effects were revealed only for STI and not for ITI, we conducted additional analyses on the ERP amplitudes at the predicted Cz scalp location. Corroborating the prior analyses, the TC–TI differences in ERP amplitude that were significant in Table 1, now showed also significance between STI and ITI given a positive context, $t(44) = 2.20$ – 2.81 , $P < 0.05$, but not given a negative context, $t < 1$.²

To summarize, these results suggest that compared to intentional inferences, spontaneous inferences are more susceptible to violations when they involve a different trait. When a moral (or immoral) actor performed an evaluatively inconsistent low (or high) achievement-related behavior, a P300 component was found following a spontaneous inference, for both positive and negative violations, as predicted. Contrary to our predictions, this P300 was not found for ITI, although we found a similar P300 pattern at an adjacent C4 scalp location in the right hemisphere, but only for negative violations.

Localization of ERP

The localization of electric activity in cortical brain areas was computed with the LORETA inverse solution method (Pascual-Marqui *et al.*, 2002; Esslen *et al.*, 2004), which is a reliable method of correct EEG localization with fairly low errors. In this analysis, a spatial resolution of 7 mm is used per voxel, resulting in a three-dimensional image consisting of 2394 voxels total. We focused our analyses on the (in)consistency conditions that showed the most robust ERP differences, that is, following a positive trait expectancy. Visual inspection of the ERP waves in Figure 1 suggests that

the timing of inconsistency differences at the predicted Cz scalp site is robust around 550 ms, so that we analyzed the LORETA solution in the 400–700 ms time interval. Before the LORETA analysis, for each condition, we first subtracted the ERP of the irrelevant condition to eliminate electric activation due to mere sentence reading and comprehension.

Figure 2 displays the time lines in the 400–700 ms interval of the maximum activation for two predicted brain areas (TPJ and mPFC), and for another area that was also frequently activated and identified as the *medial paracentral* (mPC) or superior parietal lobule which reacts to unpredictable stimuli (Berns *et al.*, 2001). This segmentation of brain activity was based on a visual inspection of the LORETA solutions. The time frame at 600 ms (small line in red) provides a representative piece of brain activity in each condition that is also close to the onset of ERP differences between conditions (Figure 1), so that all further localization analyses were conducted for this time segment. Figure 2 displays the LORETA brain maps above the corresponding time lines, for STI (left panels) and ITI (middle panels). The activation of the TPJ was at the right hemisphere and is therefore shown on the right lateral brain surface, whereas the activation of the mPFC and mPC was medial and therefore shown on the medial surface. To test our hypotheses, we calculated differences between conditions by a voxel-by-voxel *t*-test of these LORETA images (Esslen *et al.*, 2004). The *t*-values are displayed via LORETA images in the right panels of Figure 2, in red for *t*-values that are stronger in STI ($t > 1.96$) and in blue for *t*-values that are weaker ($t < -1.96$). We consider differences in the predicted TPJ and mPFC areas at the conventional 0.05 significance level ($|t| > 1.96$), while differences at other areas are assessed non-parametrically with a randomization test (Nichols and Holmes, 2002) which corrects for multiple comparisons. None of the other brain areas survived this more stringent criterion.

We predicted that the TPJ and mPFC brain areas would be most strongly activated under different conditions. Specifically, we predicted that the TPJ would be relatively more involved in STI while the mPFC would be more involved in ITI. In general, these predictions were corroborated. Figure 2 shows that the right TPJ was most strongly activated during spontaneous processing, and this almost uniquely so during the whole 400–700 ms time interval for all three conditions (TC, TI and CI). In contrast, during the intentional inferences, the TPJ activation was weaker and maximum activation switched from one brain area to the other across time. The mPFC was most strongly activated when the behavioral information was consistent and thus allowed a trait inference to be drawn, whereas the mPC was most strongly activated when the information was inconsistent. As can be seen from the *t*-values in Figure 2, the stronger activation of the right TPJ during spontaneous processing is reliable in the TI condition, but not in the other consistency conditions. In the same TI condition,

² The same ANOVA as in footnote 1, now comparing the Consistency between TC vs CI sentences, with the Cz and Pz channels as dependent variable, again corroborated these statistical results. There was a main effect of Interval indicating that the ERP amplitudes were generally higher at larger intervals, $F(2, 88) = 11.55$, $P < 0.001$, and an interaction of this factor with Consistency, $F(2, 88) = 15.58$, $P < 0.001$, confirming that inconsistencies generally resulted in higher ERP amplitudes. Likewise, for the Pz channel, the ANOVA revealed an analogous main effect of Interval, $F(2, 88) = 9.65$, $P < 0.001$, and an analogous interaction, $F(2, 88) = 20.92$, $P < 0.001$, although this was further qualified by the quadruple interaction involving all factors, $F(2, 88) = 6.33$, $P < 0.01$, including two of its underlying triple interactions, that is further analyzed in the text.

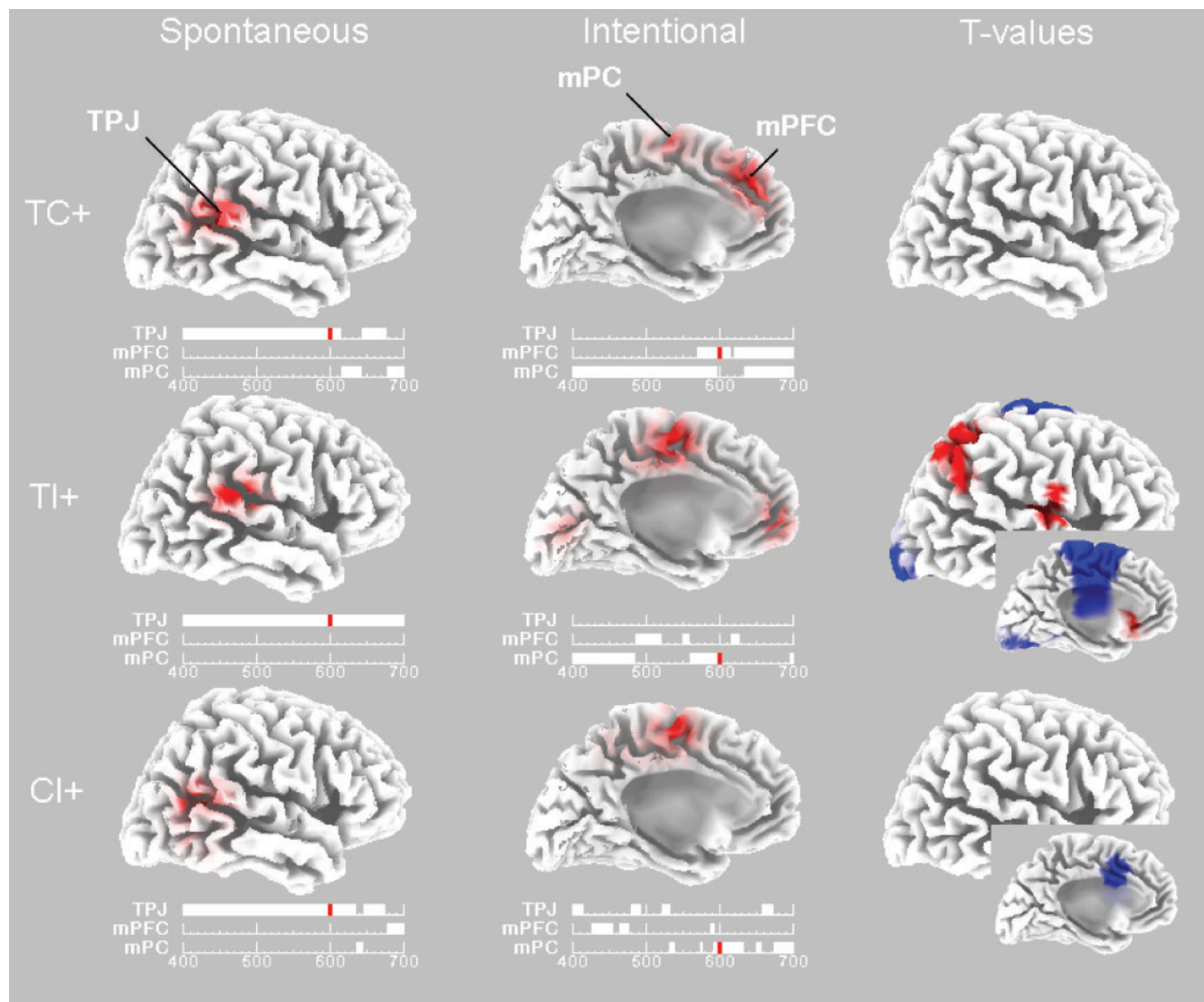


Fig. 2 LORETA source analysis. The first two columns depict the amplitudes during TC, TI and CI conditions (all minus the irrelevant condition) given a positive trait expectancy and under spontaneous and intentional instructions. The maps are scaled with respect to their minimum and maximum amplitude. The last column depicts maps of t -values, $P < 0.05$ uncorrected, for the statistical comparison between spontaneous and intentional LORETA maps, with red (vs blue) indicating greater activation under spontaneous (vs intentional) instructions. The brain maps are oriented from the posterior to the anterior cortex and display either a lateral view of the right hemisphere (e.g. maps on the left and right) or a medial view of the left hemisphere (e.g. maps in the middle and insets). Below each voltage map are time segments indicating in millisecond the largest activation of the most relevant cortical regions (ignoring occasional occipital and inferior temporal activations). TPJ, Temporo-Parietal Junction; mPFC, medial Prefrontal Cortex; mPC, medial ParaCentral area. All maps are taken at 600 ms post-stimulus, as indicated by the small red line on the time segments. Maximum activation and x, y, z Talairach-coordinates for STI: TC = $3.67 \mu\text{V}$ at 53, $-53, 15$; TI = $2.63 \mu\text{V}$ at 60, $-32, 22$; CI = $2.20 \mu\text{V}$ at 53, $-53, 15$ and for ITI: TC = $2.77 \mu\text{V}$ at $-3, 38, 29$; TI = $1.86 \mu\text{V}$ at $-3, -11, 50$ and CI = $2.14 \mu\text{V}$ at $-3, -11, 64$.

the stronger activation of the mPC during intentional processing approached significance, $P < 0.10$, corrected. There were no other reliable differences between STI and ITI. There were no differences within STI and ITI conditions, except in the STI condition where there was a significant decrease in a small portion of the mPC region following TI in comparison with TC, $t < 4.096$, $P < 0.05$ corrected (at coordinates 0, $-4, 57$). Figure 3 compares the local maximum of activation in all these three areas. As can be seen, although some areas dominate, substantial activation is observed for all areas.

Memory measures

The responses on the cued recall and sentence completion tasks were scored on the basis of verbatim accuracy of the sentence (without the actor's name) although synonyms were allowed. The proportion of correct responses is shown in Table 2.

Sentence completion. This memory measure was included to test the hypothesis that a prior impression was formed about the actor, and requires higher recall for TI sentences than for TC sentences. For sentence completion we used an ANOVA with Instruction (read, impression) as

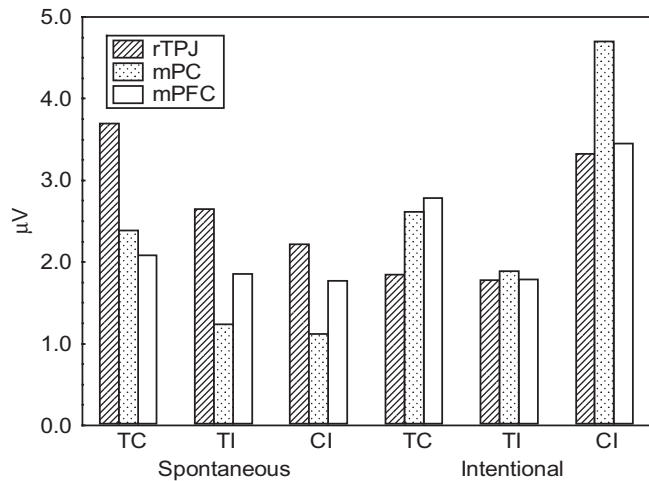


Fig. 3 Local maxima at each area of interest from the LORETA source analysis. rTPJ, right Temporo-Parietal Junction; mPC, medial ParaCentral area; mPFC, medial Prefrontal Cortex.

Table 2 Proportion sentence completion and correct recall as a function of instruction and trait consistency

Instruction	Trait consistency of the behavior			
	Consistent	Inconsistent	Competence	Irrelevant
Sentence completion				
STI	5.2 _c	7.6 _b	15.9 _a	—
ITI	8.3 _b	9.5 _b	15.7 _a	—
Trait-cued recall				
STI	6.6 _b	0.3 _c	1.1 _c	0.4 _c
ITI	8.7 _a	0.8 _c	1.2 _c	0.0 _c

Note: Means with different subscripts differ significantly from each other according to a Fisher LSD test, $P < 0.05$.

between-participants factor and Consistency (TC, TI, CI) as within-participants factors (without the IRR cell; in analogy with Bartholow *et al.*, 2001, 2003, the latter was not included in the sentence completion task). We found the predicted main effect of Consistency, $F(2, 88) = 73.61$, $P < 0.001$, $\eta^2 = 0.626$. Supporting the prediction that a prior impression was built, planned comparisons showed that TI sentences were remembered better than TC sentences, $F(1, 44) = 6.41$, $P < 0.05$. Sentence completion was best for CI sentences, and significantly so than TC sentences, $F(1, 44) = 103.48$, $P < 0.05$. However, this latter effect is probably due to a methodological limitation, as in fact there was less variation across the CI sentences, in that they often described a somewhat limited range of grades/results obtained so that participants could more easily guess the correct answer. No other effects were significant.

Cued recall. This memory measure was taken to verify the hypothesis that trait inferences were made on the basis of trait-consistent behavioral descriptions, and requires higher trait-cued recall for TC as compared to TI behaviors.

The recall data were analyzed with a similar ANOVA. Only the more interesting TC and TI conditions were analyzed because CI and IRR mean recall was zero in most STI or ITI conditions due to a floor effect, so that statistical tests were not possible. The ANOVA revealed the predicted main effect of Consistency, $F(1, 44) = 146.21$, $P < 0.001$, $\eta^2 = 0.768$, indicating that the trait cues elicited stronger recall of TC sentences than TI sentences. This suggests that the implied trait was associated with consistent behaviors.

Correlations between neural components and memory measures

The results thus far reveal that the processing of inconsistent information is associated with a P300 waveform, whereas the processing of all types of (in)consistent information is associated with activation in two major brain areas. In order to validate these neural effects as potential indicators of trait consolidation and inconsistency resolution given STI and ITI, we calculated several correlations between these memory and neural measures.

P300. Our prediction was that the P300 is indicative of inconsistency updating (as measured by sentence completion), and this was corroborated especially following STI. To track inconsistencies, for each participant, we computed a difference score between the TC and TI/CI conditions on both memory tasks as well as on the maximum P300 in the 300–450 ms interval at the Cz midline scalp location (because that location revealed the strongest effects). We then computed a Pearson correlation across all participants between the P300 and memory scores, separately for a positive and a negative trait context. As can be seen in Table 3, there was a significant correlation of the P300 with sentence completion across STI and ITI, for three out of the four consistency scores. Only inconsistency in a negative trait context (TC vs TI) did not show a significant correlation, paralleling the P300 results. Separate analysis revealed that these significant correlations were mainly due to STI ($r_s = 0.28-41$, $P < 0.11$, one-sided), while none of the ITI reached significance (Table 3). However, these differences between STI and ITI correlations were not significant. In contrast, cued recall was not significantly associated with the P300.

Loreta. Our prediction was that the LORETA activation is indicative of trait adjustment and consolidation (as measured by trait-cued recall), and this was corroborated particularly in the TPJ area following STI. To confirm this statistically, we computed a Pearson correlation across all participants between both memory tasks and the activation at all LORETA voxels in each of the TC, TI and CI conditions presented earlier (i.e. following a positive trait expectancy). To protect against multiple comparisons, we used a conventional α -level for the hypothesized brain areas (TPJ and mPFC), and adjusted levels for other areas. Trait-cued recall showed significant correlations in the TC condition, $P < 0.05$ (i.e. $r > 0.25$ across all instructions and

Table 3 Pearson correlations (across participants) between memory measures and the P300 at Cz.

Instruction	Consistency of the behavior			
	Evaluative (TC – TI)		Descriptive (TC – CI)	
	Positive Cxt	Negative Cxt	Positive Cxt	Negative Cxt
Sentence completion				
All	0.25*	0.11	0.27*	0.32**
STI	0.39*	0.05	0.28	0.41**
ITI	0.05	0.19	0.08	0.12
Trait-cued recall				
All	0.06	–0.01	0.07	–0.01
STI	0.03	–0.05	0.17	0.10
ITI	0.07	0.01	–0.11	–0.17

Note: Cxt=Context. The P300 refers to the largest positive peak within a 300–450 ms interval.

* $P < 0.05$; ** $P < 0.01$ (one-sided).

$r > 0.35$ for STI or ITI separately, one-sided), and their location is displayed in Figure 4. As predicted, trait-cued recall of consistent sentences revealed a significant correlation with a broad brain region including the TPJ and a smaller region including the mPFC following trait consolidation (TC; left panel) with a maximum of 0.39 (at the TPJ). Interestingly, a similar correlation was found between recall for consistent sentences and activation following trait inconsistencies (TI; left panel), with a maximum of 0.37. However, these reliable correlations were only generated following STI (right panels) and not during ITI, with a maximum of 0.62 for TC and 0.67 for TI, paralleling the results for the P300-sentence completion correlation. Many of these differences reached significance, $P < 0.05$. No other brain areas showed significant correlations (at the adjusted level), nor was the sentence completion task.

DISCUSSION

The aim of this study was to shed more light on the underlying neural processes involving STI and ITI, during inconsistency detection and resolution. To this end, we used besides traditional implicit memory measures typically used in STI research, also state of the art EEG measures. In support of the view inspired by dual-process models in social perception (Fiske and Neuberg, 1988; Gilbert *et al.*, 1988; Smith and DeCoster, 2000; Van Overwalle and Labiouse, 2004; Satpute and Lieberman, 2006) that STI are relatively automatic components of the inference process that guide and subserve the more controlled ITI, and in contrast to the view that these two goals of inference making proceed largely independently, our main hypothesis was that there would be substantial parallels in the memory and timing of STI and ITI processes, although we expected some differences in the main brain area of activity. Although a clear-cut confirmation of our hypothesis is premature, it seems fair to say that our study provides many interesting and suggestive findings.

Let us first look at the memory measures. Because there were no reliable differences in trait-cued recall and sentence completion between STI and ITI, this suggests that trait inferences were made also under spontaneous processing and that these involved a similar memory process as making explicit trait inferences (Carlston and Skowronski, 1994; Uleman *et al.*, 1996, 2005; Van Overwalle *et al.*, 1999; Todorov and Uleman, 2002). It is unlikely that the absence of differences is due to a lack of statistical power, as the number of participants in each group ($n = 23$) was considerably larger than earlier ERP research (i.e. Bartholow *et al.*, 2001, 2003, 2006). That trait-cued recall attained the same level under both spontaneous and intentional instructions suggests that a trait was associated with the actor, irrespective of processing goal, because the explicit instruction required making trait impressions about the actor. Moreover, the higher memory in the sentence completion task for inconsistent behaviors supports the idea that an impression was developed previously about the actor. Similar results have been reported in earlier research on inconsistency resolution (Hastie and Kumar, 1979; Hastie, 1980; Srull and Wyer, 1989; Stangor and McMillan, 1992) but this work has not explored the underlying neural correlates of these differences. Although this study did not provide direct behavioral evidence under spontaneous instructions that a trait expectancy was built about the actor, taken together, it seems save to conclude that the inferred traits referred to the actor rather than his or her behaviors.

The ERP results revealed considerable overlap between instructions and the neural timing of inconsistency processing. Consistent with the idea that the P300 component reflects the brain's response to potentially threatening and diagnostic social interaction, negative behaviors that violated an actor's morality implied previously, evoked a larger P300 than positive moral acts, in line with earlier research (Reeder and Brewer, 1979; Skowronski and Carlston, 1989; Cacioppo *et al.*, 1999; Ybarra, 2002). Despite the large number of behavioral sentences, the impoverished nature of the behavior information, and the poor memory for the information, a P300 waveform was clearly evident at the central midline scalp in the 450–650 ms interval, and irrespective of whether the trait was developed spontaneously or intentionally. Earlier research documented that the P300 is an index of a context-updating process by which an interpretation of the environment is updated and consolidated in long-term memory (Fabiani *et al.*, 1986; Fabiani and Donchin, 1995). Our findings suggest that the memory advantage of expectancy-discrepant information may stem from differences in the early engagement of underlying neural mechanisms irrespective of the intentions of the perceiver. Note that this interpretation refers to the initial stages of inconsistency detection only. It is possible that the P300 does not reflect a resolution of these discrepancies and that this is a more controlled process

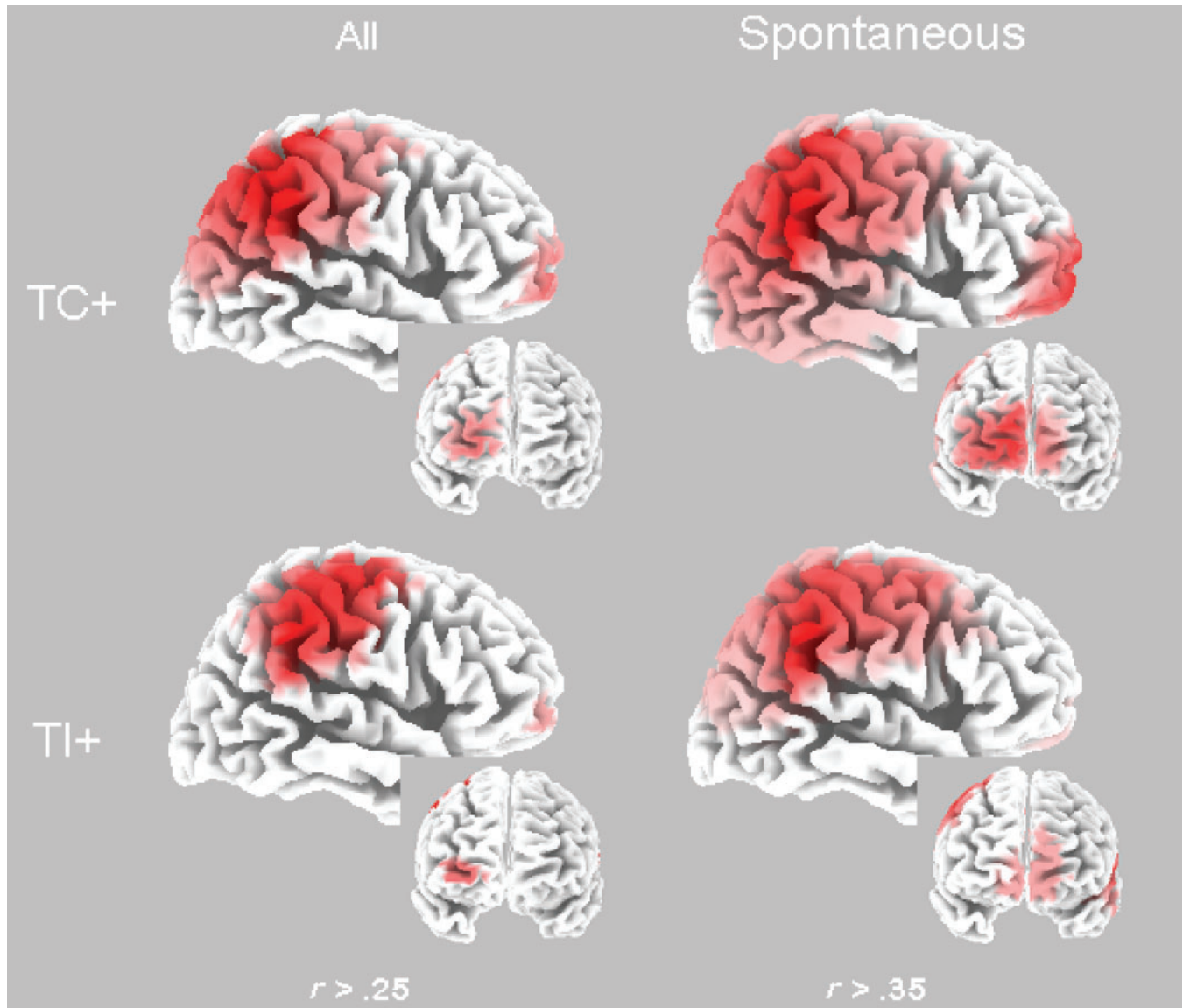


Fig. 4 LORETA maps of significant correlations (across participants) between trait-cued recall of consistent sentences and source amplitude given TC and TI following a positive trait expectancy. The maps are scaled from minimum to maximum correlation, with a minimum of > 0.25 (all; left) and > 0.35 (STI only; right) which is significant at $P < 0.05$, one-sided uncorrected. The brain maps are oriented from the posterior to the anterior cortex and display a lateral view of the right hemisphere. The insets display a frontal view oriented from right to left hemisphere. All maps are taken at 600 ms post-stimulus. The brain regions are as labeled in Figure 2.

involved during ITI only. Such complementary reasoning may remain undetected by ERP measures because the individual differences in controlled thoughts are blurred by the averaging computations of the ERP.

The ERP results pointed to interesting differences when these inconsistencies additionally involved a shift in personality domain of the trait. This is a novel contribution because earlier STI research largely ignored the scope of implied inferences. These so-called descriptive trait inconsistencies induced several P300 waveforms at the central midline scalp site in a similar 300–650 interval, for negative and positive violations during STI (as they involve a mixture of moral and behavioral information), and also for negative

violations during ITI at a nearby C4 scalp site. These results suggest that when discrepancies involve a shift of personality domain, slightly different processes are generated under intentional inferences. This may suggest either (i) that the underlying automatic processes are identical, but that increased controlled thought is elicited by the additional shift in personality domain or (ii) that these two processes are independent.

A more dramatic difference between processing goals is seen when the activation in different brain areas is considered via LORETA source analysis. Although localization of EEG activity is less precise and limited to the cortex, its millisecond time precision offers more insight in the time

course of brain activation. The results show that spontaneous thoughts activate the right TPJ during a 400–700 ms interval, whereas intentional inferences generate activation either at the mPFC following trait consistencies or at the medial paracentral lobule (mPC) following (negative) inconsistencies. This difference was significant for the TPJ following evaluative trait inconsistencies. Earlier research has documented that the TPJ and mPFC are implicated in mentalizing about others (Frith and Frith, 2001; Harris *et al.*, 2005; Saxe, 2006) and that the mPC responds to unexpected stimuli (Berns *et al.*, 2001). To our knowledge, this is the first time that reliable differences are observed between brain areas under spontaneous and intentional social inference goals. However, more research including fMRI imaging is needed to corroborate the present distinction. In a single fMRI study that addressed this issue (Mitchell *et al.*, 2006), although the TPJ and mPFC were activated, they did not reveal reliable differences between memory (STI) and explicit impression formation (ITI) instructions (only a marginal difference at the mPFC was found for trait-implicating ‘diagnostic’ information). As noted earlier, this lack of differences may be due to the fact that the memory *vs* explicit instruction was manipulated across trials within the same participants, so that they were not completely unaware of the goal of the experiment. Alternatively, it may be due to the fact that LORETA localizes electric activity at a limited representative time segment of <4 ms, whereas fMRI localizes blood de-oxygenation for several seconds. Still another possibility is that in our design, consistent sentences merely required to confirm a trait inference generated earlier during the introductory paragraph, while in typical fMRI and STI research traits need to be inferred anew given a single behavioral sentence. In another relevant fMRI study by Saxe and Wexler (2005), it was found that inconsistencies resulted in higher activation of the TPJ, in contrast to the present findings. It should be realized, however, that the ‘inconsistencies’ in Saxe and Wexler’s (2005) work involved deviations from social norms, and as such are especially diagnostic for making inferences about a person. In contrast, in our work, the inconsistencies involve contradictions to prior trait expectancies and are thus more disrupting for ongoing trait inferences.

Correlations established the reliability of our EEG measures as index of STI. Correlations between the P300 and sentence completion suggest that the higher recall of inconsistent ‘critical’ behaviors is associated with an increased P300. The correlations between LORETA brain activation (at the TPJ and mPFC) and trait-cued recall indicate that better memory for the traits implied by behavioral information is associated with increased brain activation in these two areas known to be critical in mentalizing. That no such correlation was found with trait-cued memory for inconsistent sentences may be due to the low memory performance for cued recall in these conditions, and suggests that few traits were inferred

following these inconsistencies. Remarkably, only reliable correlations were revealed during spontaneous processing. This is in line with earlier research that reported a correlation between the P300 and inconsistent material when mental elaboration is minimized (Fabiani and Donchin, 1995), and strengthens our suspicion that intentional processes are more difficult to correlate with EEG measures. Perhaps, individual differences in conscious thought may be masked by the averaging procedures for computing ERP and LORETA, or perhaps intentional reasoning is an additional process that disturbs these correlations. Many fMRI studies in this field do not corroborate the psychological validity of their neural findings with memory measures and so they cannot speak to this issue.

Some limitations

This research presented a large amount of behavioral sentences (120 TC, 40 TI, 40 CI and 20 irrelevant, not including the introductory paragraphs) to minimize strategic retrieval strategies given spontaneous instructions, in line with recent STI research using memory tasks (e.g. Todorov and Uleman, 2002). Nevertheless, future research might consider the use of potentially better alternatives of implicit memory measures of trait-actor associations that can be applied after the presentation of all stimulus material to avoid interruption of EEG measurement. Promising alternatives used in earlier research appear to be *false recognition* (Todorov and Uleman, 2002) and *relearning* (Carlston and Skowronski, 1994), which measure the link between a photo of the actor and the implied trait. The false recognition task (Todorov and Uleman, 2002) involves an explicit trait recognition based on the actor photo, and therefore requires a large set of behaviors, including distractors, to minimize controlled retrieval. The relearning task (Carlston and Skowronski, 1994) involves the relearning of photo-trait pairs that is facilitated by the trait spontaneously inferred earlier when reading about the behaviors. However, these alternative procedures have been typically used for only 12–24 trait-implicating sentences embedded in roughly a similar amount of non-implicating sentences. Scaling these numbers up to 120 trait-implicating sentences required for precise ERP measurement and taking into account that this lead to serious floor effects of memory in the present study, this runs the risk of obtaining floor effects in these alternative implicit measures as well.

CONCLUSION

The ERP analyses in this study provided novel intriguing evidence on the neural processes involved in the detection of violations of trait expectations, and more in particular on the differences between spontaneous *vs* intentional processing goals in these expectations and violations. We found a substantial overlap between memory performance and

ERP measures when an implied trait was contradicted by behaviors opposite in valence, which suggests that the timing and result of cognitive responses to trait-violations are similar irrespective of whether the trait was inferred spontaneously or intentionally. There was somewhat less overlap when the inconsistency also involved a shift in personality domain. In contrast, robust differences were observed in the brain areas that were active during spontaneous and intentional trait inferences. Unlike recent research using fMRI imaging techniques (e.g. Mitchell *et al.*, 2006; Todorov *et al.*, 2006), our results showed reliably greater activation of the TPJ during spontaneous inferences, in line with fMRI research on the role of the TPJ in spontaneous attributions of intentions and beliefs (Saxe and Kanwisher, 2003; Saxe and Powell, 2006). Nevertheless, both the TJP and mPFC were correlated with behavioral measures of trait inference, consistent with their role in mentalizing of others (Frith and Frith, 2001). Taken together, this seems to suggest that spontaneous and intentional social inferences run largely in synchrony, but with some preference for different areas in the brain. Although our memory measures did not reveal any differences in the resulting inference, it may well be that this differential brain activation leads to other qualitative differences that are too subtle to be picked up by these measures, especially during intentional thoughts (as demonstrated by a lack of correlation). Pathologies of person inference, such as in autism and paranoia, might benefit greatly from the insights gained from this and similar studies on the neural correlates of spontaneous and intentional impression formation.

Conflict of Interest

None declared.

REFERENCES

- Andreassi, J.L. (2000). *Psychophysiology: Human Behavior and Physiological Response*. Mahwah, NJ: Erlbaum.
- Bargh, J.A. (1989). Conditional automaticity: varieties of automatic influence in social perception and cognition. In: Uleman J.S., Bargh J.A., editors. *Unintended Thought*. New York: Guilford, pp. 3–51.
- Bartholow, B.D., Bushman, B.J., Sestir, M.A. (2006). Chronic violent video game exposure and desensitization to violence: Behavioral and event-related brain potential data. *Journal of Experimental Social Psychology*, 42, 532–539.
- Bartholow, B.D., Fabiani, M., Gratton, G., Bettencourt, B.A. (2001). A psychophysiological examination of cognitive processing of and affective responses to social expectancy violations. *Psychological Science*, 12, 197–204.
- Bartholow, B.D., Pearson, M.A., Gratton, G., Fabiani, M. (2003). Effects of alcohol on person perception: a social cognitive neuroscience approach. *Journal of Personality and Social Psychology*, 85, 627–38.
- Belouchrani, A., Meraim, K.A., Cardoso, J.-F., Moulines, E. (1993). Second-order blind separation of correlated sources. In *proceedings of the international conference on digital signal processing*, Cyprus, 346–51.
- Berns, G.S., McClure, S.M., Pagnoni, G., Montague, P.R. (2001). Predictability modulates human brain response to reward. *The Journal of Neuroscience*, 21, 2793–8.
- Botvinick, M.M., Cohen, J.D., Carter, C.S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, 8, 539–46.
- Brewer, M.B. (1988). A dual process model of impression formation. In: Wyer, R.S., Srull, T.K., editors. *Advances in Social Cognition*, Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 1–36.
- Cacioppo, J.T., Crites, S.L., Berntson, G.G., Coles, M.G.H. (1993). If attitudes affect how stimuli are processed, should they not affect the event-related brain potential? *Psychological Science*, 4, 108–12.
- Cacioppo, J.T., Crites, S.L., Jr., Gardner, W.L., Berntson, G.G. (1994). Bioelectrical echoes form evaluative categorizations: I. A late positive brain potential that varies as a function of trait negativity and extremity. *Journal of Personality and Social Psychology*, 67, 115–25.
- Cacioppo, J.T., Gardner, W.L., Berntson, G.G. (1999). The affect system has parallel and integrative processing components: form follows function. *Journal of Personality and Social Psychology*, 76, 839–55.
- Carlston, D.E., Skowronski, J.J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, 66, 840–80.
- Eiser, J.R., Fazio, R.H., Stafford, T., Prescott, T.J. (2003). Connectionist simulation of attitude learning: asymmetries in the acquisition of positive and negative evaluations. *Personality and Social Psychology Bulletin*, 29, 1221–35.
- Esslen, M., Pascual-Marqui, R.D., Hell, D., Kochi, K., Lehmann, D. (2004). Brain areas and time course of emotional processing. *NeuroImage*, 21, 1189–203.
- Fabiani, M., Donchin, E. (1995). Encoding processes and memory organization: a model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 224–40.
- Fabiani, M., Karis, D., Donchin, E. (1986). P300 and recall in an incidental memory paradigm. *Psychophysiology*, 23, 298–308.
- Fiske, S.T., Neuberg, S.L. (1988). A continuum model of impression formation: from category-based to individuating processes as a function of information, motivation, and attention. *Advances in Experimental Social Psychology*, 23, 1–108.
- Frith, U., Frith, C. (2001). The biological basis of social interaction. *Current Directions in Psychological Science*, 10, 151–5.
- Fyock, J., Stangor, C. (1994). The role of memory biases in stereotype maintenance. *British Journal of Social Psychology*, 33, 331–43.
- Gilbert, D.T., Pelham, B.W., Krull, D.S. (1988). On cognitive busyness: when person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, 54, 733–40.
- Harris, L.T., Todorov, A., Fiske, S.T. (2005). Attributions on the brain: neuro-imaging dispositional inferences, beyond theory of mind. *NeuroImage*, 28, 763–9.
- Hastie, R. (1980). Memory for behavioral information that confirms or contradicts a personality impression. In: Hastie, R., Ostrom, T., Ebbesen, E., Wyer, R., Hamilton, D., Carlston D., editors. *Person Memory: The Cognitive Basis of Social Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hastie, R., Kumar, P. (1979). Person memory: personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology*, 37, 25–38.
- Lupfer, M.B., Weeks, M., Dupuis, S. (2000). How pervasive is the negativity bias in judgments based on character appraisal? *Personality and Social Psychology Bulletin*, 26, 1353–60.
- McCarthy, G., Luby, M., Gore, J., Goldman-Rakic, P. (1997). Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI. *Journal of Neurophysiology*, 77, 1630–4.
- McKoon, G., Ratcliff, R. (1986). Interferences about predictable events. *Journal of Experimental Psychology: Learning, memory, and cognition*, 12, 82–91.
- Miller, E.K., Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.

- Mitchell, J.P., Banaji, M.R., Macrae, C.N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience*, 17, 1306–15.
- Mitchell, J.P., Cloutier, J., Banaji, M.R., Macrae, C.N. (2006). Medial prefrontal dissociations during processing of trait diagnostic and nondiagnostic person information. *SCAN*, 1, 49–55.
- Nichols, T.E., Holmes, A.P. (2002). Nonparametric permutations tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, 15, 1–25.
- Pascual-Marqui, R.D., Esslen, M., Kochi, K., Lehmann, D. (2002). Functional imaging with low resolution brain electromagnetic tomography (LORETA): a review. *Methods and Findings in Experimental & Clinical Pharmacology*, 24C, 91–5.
- Reeder, G.D. (1997). Dispositional inferences of ability: content and process. *Journal of Experimental Social Psychology*, 33, 171–89.
- Reeder, G.D., Brewer, M.B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, 86, 61–79.
- Reeder, G.D., Fulks, J.L. (1980). When actions speak louder than words: implicational schemata and the attribution of ability. *Journal of Experimental Social Psychology*, 16, 33–46.
- Reeder, G.D., Spores, J.M. (1983). The attribution of morality. *Journal of Personality and Social Psychology*, 44, 736–45.
- Satpute, A.B., Lieberman, M.D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, 1079, 86–97.
- Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, 16, 235–9.
- Saxe, R., Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in the “theory of mind”. *NeuroImage*, 19, 1835–42.
- Saxe, R., Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, 43, 1391–9.
- Saxe, R., Powell, L.J. (2006). It’s the thought that counts: specific brain regions for one component of theory of mind. *Psychological Science*, 17, 692–9.
- Skowronski, J.J., Carlston, D.E. (1987). Social judgment and social memory: the role of cue diagnosticity in negativity, positivity and extremity biases. *Journal of Personality and Social Psychology*, 52, 689–99.
- Skowronski, J.J., Carlston, D.E. (1989). Negativity and extremity biases in impression formation: a review of explanations. *Psychological Bulletin*, 105, 131–42.
- Skowronski, J.J., Carlston, D.E., Mae, L., Crawford, M.T. (1998). Spontaneous trait transference: communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology*, 74, 837–48.
- Slooman, S.A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22.
- Smith, E.R., DeCoster, J. (2000). Dual-Process models in social and cognitive psychology: conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4, 108–31.
- Srull, T.K., Wyer, R.S. (1989). Person memory and judgment. *Psychological Review*, 96, 58–83.
- Stangor, C., McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: a review of the social and social developmental literatures. *Psychological Bulletin*, 111, 42–61.
- Stern, R.M., Ray, W.R., Quigley, K.S. (2001). *Psychophysiological Recording*, 2nd edn. New York: Oxford University Press.
- Todorov, A., Uleman, J.S. (2002). Spontaneous trait inferences are bound to actors’ faces: evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, 83, 1051–64.
- Todorov, A., Gobbini, I.M., Evans, K.K., Haxby, J.V. (2006). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, 45, 163–173.
- Uleman, J.S. (1999). Spontaneous versus intentional inferences in impression formation. In: Chaiken, S., Trope, Y., editors. *Dual-Process Theories in Social Psychology*. New York: The Guilford Press, pp. 141–60.
- Uleman, J.S., Blader, S.L., Todorov, A. (2005). Implicit impressions. In: Hassin, R.R., Uleman, J.S., Bargh, J.A., editors. *The New Unconscious*. New York: Oxford University Press, pp. 362–92.
- Uleman, J.S., Newman, L.S., Moskowitz, G.B. (1996). People as flexible interpreters: evidence and issues from spontaneous trait inference. In: Zanna, M.P. editor. *Advances in Experimental Social Psychology*, Vol. 28, San Diego, CA: Academic Press, pp. 211–79.
- Van Overwalle, F., Labiouse, C. (2004). A recurrent connectionist model of person impression formation. *Personality and Social Psychology Review*, 8, 28–61.
- Van Overwalle, F., Drenth, T., Marsman, G. (1999). Spontaneous trait inferences: Are they linked to the actor or to the action? *Personality and Social Psychology Bulletin*, 25, 450–62.
- Wigboldus, D.H.J., Sherman, J.W., Franzese, H.L., van Knippenberg, A. (2004). Capacity and comprehension: spontaneous stereotyping under cognitive load. *Social Cognition*, 22, 292–309.
- Winter, L., Uleman, J.S. (1984). When are social judgments made? Evidence for the spontaneity of trait inferences. *Journal of Personality and Social Psychology*, 47, 237–52.
- Ybarra, O. (2002). Naïve causal understanding of valenced behaviors and its implications for social information processing. *Psychological Review*, 128, 421–41.