ERP Time Course and Brain Areas of Spontaneous and Intentional Social Inferences

Frank Van Overwalle & Kris Baetens
Vrije Universiteit Brussel, Belgium

To appear as chapter in Tomasz Maruszewski, Małgorzata Fajkowska and Michael W. Eysenck (Eds.) in Warsaw Lectures on Personality and Social Psychology. Volume 1: Personality from biological, cognitive and social perspectives. (Malgosia Fajkowska [weronika@psychpan.waw.pl])

Address for correspondence: Frank Van Overwalle, Department of Psychology, Vrije Universiteit Brussel, Pleinlaan 2, B - 1050 Brussel, Belgium; or by e-mail: Frank.VanOverwalle@vub.ac.be.

Running Head: Social Inferences and ERP

15 November, 2008
In our daily interactions with other people, it is beneficial to have some sense of their immediate intentions and long-term personality traits in order to achieve smooth communication and mutual understanding. Detecting underlying personality traits, and especially intentions of others seems even more crucial when we have to make instantaneous decisions about whether they mean us good or bad, so as to approach or avoid them. Our immediate well-being and survival may depend on such critical decisions. For instance, is someone making a move to hit me or to shake my hand? Is a person carrying a gun or a mobile phone? Prior research in the social cognition literature has revealed that we can make such social inferences very spontaneously and almost automatically, including judgments on goals (Hassin, Aarts & Ferguson, 2005) and traits (for a review, see Uleman, Blader & Todorov, 2005). However, this research tells little about the underlying neurological processes by which we make such inferences, the timing at which these processes occur and where in the brain they are computed. What are the neurological processes driving the detection of goals and traits of other people? How fast can we make goal and trait interpretations from behavioral information?

To answer these questions, this chapter briefly reviews prior neurological evidence on social processes, presents novel evidence on the timing of inferences like goals and traits, and shows how several brain areas are involved in sustaining these computations. In particular, to study the naturalness of these social processes, we focus on processing differences between spontaneous social inferences (SSI) versus intentional social inferences (ISI). Spontaneous inferences reflect that we make social judgments without intention or awareness, for instance, while being busy doing other things. In contrast, intentional inferences reflect that we have the explicit goal to deliberately create impressions of others, such as, for instance, while gossiping. This idea is prominent in dual-process models of person perception and social neuroscience, which propose that social information processing involves either spontaneous associative processes or controlled symbolic reasoning (Satpute & Lieberman, 2006; Smith & DeCoster, 2000; Keysers & Gazzola, 2007). Past behavioral research on spontaneous inferences in social psychology has focused mainly on traits of other persons. It has been demonstrated time and again that trait inferences can be made spontaneously, in the sense that if an actor's behavior is diagnostic of a trait, inferring that trait requires almost no intention or awareness, involves only little mental effort, and is difficult to suppress or modify (see Uleman, Blader & Todorov, 2005). Only very recently, researchers began to explore other types of
spontaneous inferences such as on the goals of someone’s actions (Hassin, Aarts & Ferguson, 2005; Van Duynslaeger, Coomans & Van Overwalle, 2008; Van Duynslaeger, Timmermans & Van Overwalle, 2008).

This earlier social cognition research, however, was limited to behavioral tasks, which does not allow for the exploration of the type and timing of the underlying neural circuits related to the identification of social inferences. At the neural level, we do not know whether goals and traits are identified at the same time and recruit the same brain areas or not, and to what extent spontaneous or intentional inferences modulate this neural process. In fact, we know little about the timing of spontaneous and intentional social inferences in general, and only somewhat more about their localization.

To provide more insight in these matters, this chapter provides a description of two recent studies using event-related brain potentials (ERP) derived from electroencephalogram (EEG) measurements which afford millisecond accuracy on the timing of brain processes. These brain waves also allow localizing the source of these ERP signals, so that the relevant brain regions involved in this process can be identified. Before embarking on this, we first provide an overview of the brain areas involved in social judgments like goals and traits on the basis of research using functional magnetic resonance imaging (fMRI), and then introduce the methodology and relevant social research on ERPs. We conclude this chapter with a discussion of the assumed process of social inference.

### Brain Areas recruited for Social Judgments

A recent meta-analysis of fMRI evidence by Van Overwalle (2008) suggests that two brain areas are predominantly involved in the understanding and attribution of mental states, such as goals and traits of others. As depicted in Figure 1A, inferences of goals (i.e., intentions and desires) and goal-directed behaviors by others mainly involve the **temporo-parietal junction (TPJ)**. For instance, when viewing animations of simple objects that move in a human-like fashion, observers have an immediate sense of intentionality (e.g., a triangle "chases" a square) that activates the TPJ (Martin & Weisberg, 2003; Ohnishi et al., 2004; Schultz et al., 2004). When seeking or reacting to an appropriate ending **for a story presented in a verbal or cartoon format**, the TPJ is more strongly activated for stories involving behavioral intentions than mere physical events (Blakemore et al., 2007; den Ouden et al., 2005; Saxe & Wexler, 2005; Völlm et al., 2006; Walter et al. 2004). When
reading stories, the TPJ is recruited more when these stories involve beliefs and reasoning by the actors than their physical appearances (Saxe & Powell, 2006) or when these stories involve moral dilemmas as opposed to impersonal decisions (Greene et al., 2001, 2004; Moll et al., 2001, 2002). In a recent study (Spiers & Maguire, 2006), participants were immersed in a virtual reality of taxi driving in central London from a first person perspective, and immediately afterwards they reported (without advanced warning) what they were thinking while viewing a video of their performance during scanning. It was found that the TPJ was engaged every time they thought about other agents' behaviors and intentions (e.g., customers, drivers and pedestrians). In sum, these tasks elicit the TPJ irrespective of the visual or verbal format of the material. Many of these tasks do not focus explicitly on the goal underlying the behaviors, and these inferences are thus relatively spontaneous.

In contrast, as depicted in Figure 1B, the medial prefrontal cortex (mPFC) is predominantly recruited for attributing enduring traits and attributes of others and the self (e.g., Mitchell, Banaji & Macrae, 2005; Todorov, Gobbini, Evans & Haxby, 2007). As can be seen, trait inferences of unfamiliar other people seem to activate the dorsal (superior) part of the mPFC, while trait inferences of familiar others (e.g., mother, siblings) and of the self mainly recruit the ventral (inferior) part of the mPFC. Note that these trait studies mostly involve intentional judgments. A recent study by Mitchell, Cloutier, Banaji and Macrae (2006) compared intentional with spontaneous (i.e., memory) instructions while participants were scanned using fMRI and found equally strong activation of the dorsal mPFC for trait-diagnostic sentences. However, these instructions were alternated between trials so that it is doubtful that the "spontaneous" trait inferences were made without any awareness and intention.

Taken together, all these recent fMRI findings as well as single cell-cell recordings of macaques (Keysers & Perret, 2004) have lead to the proposal that that the TPJ implements pre-reflective, intuitive and empathic representations about other people's intentions and beliefs. In contrast, the mPFC is involved in more deliberative reasoning (Keysers & Gazzola, 2007). In addition, the mPFC supports more integrative and abstract representations, like trait inferences (Van Overwalle, 2008). According to Van Overwalle (2008), this seems to suggest a processing sequence or neural pathway from automatic goal inference that informs later processing of traits. This pathway forms the hypothetical framework of this chapter and is illustrated in Figure 1C.

Briefly put, as depicted in the figure, visual input of human behaviors leads to the activation
of the superior temporal sulcus (STS) which is sensitive to human movement. From there, visual processing (as well as verbal processing; not shown) further continues to the TPJ, where the direction or end-point of the movement is identified. Although little is know about the underlying processes in the TPJ, most likely, the information on observed behavioral end-points is compared to one's own knowledge on behavioral orientations and associated goals, enabling the identification of ordinary goals of others. Evidence on this comes, in part, from single-cell recordings with macaques which reveal that some neurons react to the end-location of the movement (reflecting its goal) such as bringing food to the mouth (to eat) or to another location (to displace) rather than the movement itself (Keysers & Perrett, 2004). Similarly, recent evidence with humans (Decety & Lamm, 2007; Mitchell, 2007) suggests that the TPJ area identifies the orientation or directionality of movement and behavior across different content domains. Hence, this area is involved in the orientation of behaviors in space and beyond, that is, at a more abstract level reflecting the behavior's functional finality or goal. Once the intentions of others are inferred, this may lead to inferences of traits of the other persons. Because the mPFC can incorporate information over longer lapses of time, it is ideally suited to process stable trait attributes of self and others (see Van Overwalle, 2008), especially if the behavior from which the trait is inferred, is repeated over time.

In sum, based on past fMRI research, it appears that goal inferences are quickly inferred in the TPJ, while trait inferences require more processing in the mPFC. Note that many of the goal-directed behaviors that typically engage the TPJ, also elicited the mPFC to some substantial degree (Van Overwalle, 2008). This may be due to spontaneous trait inferences after goals are inferred. Because fMRI offers a time resolution in the range of a few seconds, it does not allow separating goal and trait inferences at different stages of the process. To provide more accurate evidence on the timing of social judgment processes, we now turn to research using ERP.

ERP Components of Social Stimuli

As illustrated in Figure 2, ERPs are averages from the brain's EEG waves, time-locked at the onset of a specific stimulus (denoted by 0 ms) and typically stretching several 100 milliseconds before and after that onset. The pre-stimulus time window serves as a baseline, to study the changes after the stimulus is presented. These ERPs waveforms therefore reflect electric activity of the brain while responding to specific stimuli, and their size or amplitude is expressed in µV (see Figure 2). To elicit ERP responses to particular stimuli, researchers typically use an "oddball" paradigm, in
which infrequent or inconsistent information is provided among otherwise frequent and consistent information. Such inconsistent stimuli often result in a deflection of the ERP waveforms, that is, an increase in their peak amplitudes. The timing of an ERP deflection reveals the processing stage at which the brain reacts to inconsistencies, and at minimum indicates that an inconsistency is detected and that its identification has begun. Different timings of ERP deflections are associated with different functions or manifestations of information processing activities, and research has revealed two main ERP types which index inconsistency detection (see also Figure 3). The first type is the P200 which reacts to early and automatic feature encoding and classification (Peters, Suchan, Zhang & Daum, 2005). The second type is the P300 which occurs later and responds to inconsistencies in task-related comprehension. These characteristics make the P200 and P300 ideally suited for exploring the neural correlates of early (automatic) and late (intentional) social inferences respectively.

**P200 and P300 as Indices of Inconsistency**

The P200 is an early positive waveform that peaks at about 200 ms after the critical stimulus. There is evidence that the P200 indexes early sensory stages of feature detection (Luck & Hillyard, 1994), encoding and classification (Dunn, Dunn, Languis & Andrews, 1998; Raney, 1993). In an oddball paradigm, the P200 reflects increased attention due to unexpected or improbable stimuli such as words (Peters et al., 2005). Research on lexical access indicates that low-frequency words lead to increased ERP amplitudes in comparison with high-frequency words as early as ~ 130 ms (Sereno, Rayner & Posner, 1998; Dambacher, Kliegl, Hofmann & Jacobs, 2006). More importantly, this research also revealed that a preceding sentence fragment may prime the appropriate meaning of an ambiguous word during a 130-200 ms window (Sereno, Brewer & O’Donnell, 2003). This suggests that a preceding sentence context may “direct early, lexical selection of the appropriate meaning” (Sereno et al., 2003, p. 328). These findings have lead to the view that the P200 and earlier ERP components index the selection or beginning of comprehension in response to preceding information.

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 is therefore also termed the *late positive potential*). Research using the oddball paradigm 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, Karis & Donchin, 1986; Fabiani & Donchin, 1995). These findings have led to the view that the P300 is an index of on-line updating of working memory after inconsistency detection.

**P200 and P300 during Social Impressions**

Social processes occur very rapidly. Several researchers found enhanced positive ERPs in response to emotional images, like angry, fearful or happy human faces or armed persons in comparison to neutral faces or unarmed persons, leading to a P200 or P300 (Carretié, Mercado, Tapia & Hinojosa, 2001; Correll, Urland & Ito, 2006; Eimer & Holmes, 2002; Eimer, Holmes & McGlone, 2003; Keil, Bradley, Hauk, Rockstroh, Elbert & Lang, 2002; Kubota & Ito, 2007). The P200 is sensitive to outgroup versus ingroup racial and gender differences, and this has been explained by the idea that outgroup members are often seen as more negative or threatening than ingroup members. Ito and colleagues found that Black or male outgroup faces evoke larger P200s than White or female ingroup faces (Correll, Urland & Ito, 2006; Ito & Urland, 2003, 2005; Kubota & Ito, 2007). The P200 race divergence occurs even when tasks direct attention away from race and the judgment is made spontaneously (Ito & Urland, 2005). Figure 3A illustrates a stronger P200 and P300 for Black faces in an (inconsistent) White context as opposed to Black context (note that downwards reflects a more positive amplitude; Ito & Urland, 2005). These findings suggest early attention allocation to more emotional or distinctive social stimuli and therefore reflect an automatic vigilance effect in which attention is quickly and spontaneously directed to stimuli with potentially negative implications.

More evidence on person inferences is available on the P300. Cacioppo and coworkers (Cacioppo, Crites, Berntson & Coles, 1993; Cacioppo, Crites, Gardner & Berntson, 1994) found that an 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 500 and 1000 ms. More pertinent to the present issue of deriving inferences from behavior, Bartholow and colleagues (2001, 2003) asked their participants to form intentional trait impressions about actors described in short behavioral 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). As shown in Figure 3B, the results showed significantly greater P300 activation in a 300-650 ms time window after presenting a critical word at the end of a behavioral description that revealed an inconsistent trait as opposed to a consistent trait. This result implies that intentional traits inferences are made at a late processing stage at ~ 400 ms post-stimulus; but note that this tells little about the timing of trait inferences during spontaneous conditions.

**An Inconsistent Behavioral Information Paradigm of Goal and Trait Inferences**

The prior overview suggests that goal inferences lead to an early and automatic activation of the TPJ, whereas trait inferences are slower and mainly recruit the mPFC. Especially during that latter processing stage, differences between spontaneous or intentional instructions may potentially show their greatest impact. In the remaining part of this chapter, we describe two recent studies comparing for the first time the neural timing and localization of spontaneous and intentional social inferences. One study explored goal inferences (Van der Cruyssen, Van Dyurslaeger, Cortoos & Van Overwalle, 2008) while the other investigated trait inferences (Van Duynslaeger, Van Overwalle & Verstraeten, 2007). In order to make the similarities and differences between goal and trait inferences more apparent, some of the results presented here were newly computed from these data and were not reported earlier.

**The Oddball Paradigm**

How can we explore the timing and localization of social processes? To do so, we borrowed the "oddball" paradigm as applied by Barholow et al. (2001, 2003) on trait-implying behavioral descriptions of other persons, and extended this paradigm to goal inferences and spontaneous processing instructions. Specifically, we provided short behavioral descriptions that were consistent with, inconsistent (opposite) with or irrelevant to the goal or trait implied during preceding behaviors. We instructed our participants either explicitly to infer the goal or trait of each target person (ISI) or to read the stimulus material carefully, without mentioning anything about a person's goals or traits (SSI). Table 1 illustrates the kind of sentences that were presented to the participants.

Social research has shown that negative inconsistencies in social inference give rise to more prominent responses (Cacioppo, Gardner & Berntson, 1999; Reeder & Brewer, 1979; Ybarra, 2002)
and stronger ERP deflections (Bartholow et al., 2003; Cacioppo, Gardner & Berntson, 1999). The reason is that negative violations of morality are more diagnostic. For instance, when an honest person is caught lying or stealing, this is much more diagnostic and revealing about the person than when a criminal pays a bill (e.g., a thief is not always stealing). Likewise, Van Duynslaeger et al. (2007) presented negative trait implications followed by positive trait violations, and did not found significant ERP differences. Therefore, we focus on positive goal or trait implications followed by the more diagnostic negative violations, as shown in Table 1. If the ERP waveforms show greater deflections following an inconsistency, this indicates that the participants identified the basic semantics of the inference and allocated early attention to it (for the P200), or dedicated more processing resources to identify and understand the implications of the inconsistency (for the P300).

**Manipulating Behavioral Consistency: The Oddball Paradigm in Action**

How was the oddball paradigm of Bartholow et al. (2001, 2003) applied in our research? Participants read several sentences that described the behavior of a fictitious target actor and from which a strong goal or trait could be inferred or not (see Table 1). Each set of sentences was introduced by showing the name of the actor on the computer screen for two seconds. For each target actor, a series of behavioral sentences was presented, each consisting of six words shown in the center of the computer screen. Every word was presented during 300 ms followed by a 350 ms blank (Osterhout, Bersick & McLaughlin, 1997). Critically, the last word of the last sentence determined the degree of consistency with the previously implied goal or trait: Consistent (C), inconsistent (I) and irrelevant (IR). These sentences had been pretested so that a majority of people spontaneously inferred the same implied goal or trait (> 70%), or the opposite inference (> 70% for inconsistent sentences), or no inference at all (< 30% for irrelevant sentences). This was further verified using scales in which the goal or trait was explicitly rated on a 11-point rating (0 = less applicable to 10 = more applicable) as well as its valence (0 = negative to 10 = positive).

Apart from these standard consistency conditions, we also added some extra inconsistency conditions. For goal inferences, we included descriptions containing observations of phenomena from nature that involve no intention or goal-directedness from the actor. Our question was whether these no-goal events would elicit stronger effects than goal-related events because they involve no intentionality at all. For trait inferences, we added inconsistent sentences that implied, instead of an opposite trait, a trait that involved a completely different personality content. In particular, we
provided behavioral descriptions implying high moral traits, followed by inconsistent sentences that implied the low competence of the actor. Our question here was whether such greater inconsistency, not only in valence but also in content, would be identified with the same ease and speed as behaviors implying only opposite (moral) traits?

**Spontaneous and Intentional Instructions and EEG Recording**

To manipulate spontaneous versus intentional processing goals, the participants were informed that they would read stories about several persons before they were presented with the stimulus material. Next, they received the instruction to "familiarize yourself with the material of the experiment" (SSI) or to "infer the goal that the person wants to reach / infer the personality trait of that person" (ISI; see also Todorov & Uleman, 2002). During reading, the EEG was recorded from 19 scalp sites according to the international 10-20 electrode system. In discussing the ERP results, we focus on the most meaningful results of 1 central midline (i.e., between hemispheres) site, termed Cz. For computing the localization of brain activity, however, all 19 sites are used.

**Did Spontaneous Inferences Occur? Memory Validation**

After the experiment, the participants were given two memory tasks typically used in behavioral research to measure and validate spontaneous social inferences. The first, a cued recall task (Winter & Uleman, 1984), requested participants to recall all information presented with the aid of a cue word that reflects the implied goal or trait. When observing or reading about a behavior, the implied goal or trait is assumed to be stored in memory together with the behavioral information from which it is inferred. As can be seen in the top panel of Table 2, the results revealed the predicted pattern of greater recall for consistent sentences after cuing with the implied goal or trait (Van der Cruyssen et al., 2008; Van Duynslaeger et al., 2007). As a consequence, this suggests that while reading the material, the implied goal and traits were inferred and memorized, under spontaneous as well as intentional instructions.

The second memory measure, a sentence completion task (Bartholow et al., 2001, 2003), requested to complete the last, critical word of the original sentences. Memory effects on this measure depend on the assumption that inconsistent information receives more cognitive processing and is memorized better—at the word–level—than consistent behavior (Stangor & McMillan, 1992). However, this measure might be less diagnostic for goals because goals are adaptive to new
circumstances, so that inconsistent behavior may suggest a change of proximal means in the pursuit of the same distal intention (e.g., buying an "inconsistent" gun for a party suggests a surprise party for Halloween or carnival) rather than a real discrepancy. As shown in the bottom panel of Table 2, the pattern of results was as expected. There were no memory differences between consistent and inconsistent words for goal-implying sentences (Van der Cruyssen et al., 2008), but there was stronger memory for inconsistent than consistent words from trait-implying sentences, although the difference was only significant for SSI (Van Duynslaeger et al., 2007).

In summary, the memory measures confirm that goals and traits were indeed inferred under both SSI and ISI instructions. There were no substantial differences in memory performance between SSI and ISI, indicating that SSI were made to the same degree as ISI, and are stored in memory to a similar degree. Having validated the social inferences processes under study, we can now turn to the results of the ERP measures.

**ERP Timing**

For goal inferences, a P200 deflection emerged that was substantially stronger for sentences that were inconsistent or irrelevant with the induced goal, or involved no goal at all, in comparison with goal-consistent sentences, and this effect was observed irrespective of instruction (Van der Cruyssen et al., 2008). Figure 4A depicts the ERP waveforms of the most important conditions at the Cz midline site. As can be seen, no-goal behaviors (dashed line) and irrelevant behaviors (light line) elicited a larger positive ERP than goal-consistent behaviors (dark line) starting at ~ 200 ms (see arrows). These findings suggest that after goal expectations were developed during the introductory sentences, goal discrepancies were detected at a very early and automatic processing stage. The difference with the irrelevant, and especially the no-goal condition was generally stronger than with the inconsistent condition, and continued until 1 second post-stimulus. This further confirms our suspicion that inconsistent goals imply a change of means or type of the same distal goal rather than real discrepancies, whereas irrelevant or absent goals imply a more drastic departure from goal-directed behavior.

Turning to trait inferences, a P300 deflection emerged that was substantially stronger when a moral actor performed an inconsistent immoral as opposed to a moral behavior, irrespective of SSI or ISI instructions (Van Duynslaeger et al., 2007). Figure 4B shows that when trait inferences (dark line) are violated by negative behaviors (light line), this elicits a larger positive ERP at the Cz scalp...
site starting at ~ 600 ms (see arrows) and continuing over a second. These results suggest that after trait expectations were induced during the introductory sentences, inconsistent trait implications were detected considerably later as indexed by the P300. A similar pattern of results was found for competence-inconsistent sentences, although these discrepancies led to an earlier P300 at ~ 400 ms. In another study by Van Duynslaeger et al. (2008) that investigated spontaneous trait inferences, a significant P300 was found for trait-inconsistencies, also at an earlier ~ 400 ms. This late onset of trait detection beginning no sooner than 400-600 ms suggests that, in contrast to goal inferences, longer and more processing is involved in trait inferences.

**Memory Validation**

To further validate that these ERP components are indicative of goal and trait identification, we computed correlations with cued recall and sentence completion. As can be seen in Table 3, there were significant correlations of at least one memory measure with early (P200) and late (P300) ERP waveforms, for goals as well as for traits. There are also differences between spontaneous and intentional instructions, as shown in stronger correlations of memory with intentional goals and with spontaneous traits. This may suggest that a minimal amount of focused attention or deliberation during goal inferences is necessary to develop a memory trace of them, because goal inferences are made very quickly and automatically. In contrast, reflective thought during deliberative trait inferences can be more extensive as they occur later, and may hamper memory because so much additional thoughts may differ between participants and so undermine these correlations (see also Fabiani, Karis & Donchin, 1986; Fabiani & Donchin, 1995). Taken together, the significant correlations with traditional behavioral memory tasks indicate that the ERP deflections are valid indicators of goal and trait inferences.

**ERP Localization**

The localization of electric activity in cortical brain areas was computed with the LORETA inverse solution method (Pascual-Marqui, Michel & Lehmann, 1994), which is a reliable method that estimates the EEG localization with fairly low errors "at worst in the order of 14 mm" (Pascual-Marqui, 1999, p. 85). This method has previously been applied to social interpretations of emotions (Esslen, Pascual-Marqui, Hell, Kochi & Lehmann, 2004) and although the spatial resolution of ERP waves is poorer than fMRI, a prime advantage is that their high time resolution allows for the
localization and imaging of brain activity in a millisecond range. Nevertheless, it should be realized that this solution is not unique, and should be confirmed by alternative methods such as fMRI. We focus our analyses on the consistent conditions and subtracted a baseline from these, in order to eliminate irrelevant electric activation due to mere sentence reading and comprehension. As irrelevant baseline for goals, we chose the no-goal condition that showed the most robust ERP differences with the consistent condition; for traits we chose the trait-irrelevant condition. We computed LORETA brain maps in a broad 100-700 ms time interval. To obtain robust results with less noise, these solutions were averaged over time intervals of 50 ms each (except for the critical 200-250 ms interval for goal inferences, where averages were taken for each 25 ms). When subsequent intervals did not show substantial changes, they were collapsed (i.e., averaged) together.

Figures 6 and 7 display LORETA brain maps marking the brain areas of maximum activation relative to other areas in the brain in dark grey. As can be seen, these "hot spots" fall at the two predicted brain areas (TPJ and mPFC), as well as at some other brain areas including the medial paracentral cortex (mPC) extending to the posterior mPFC (or dorsal anterior cingulate cortex), which is typically recruited during conflict monitoring (e.g., Stroop task; see Van Overwalle, 2008). Activation in this latter area may suggest that participants were aware of the inconsistencies embedded in the stimulus material and reacted to it, which is consistent with the finding that this activation appears under intentional instructions only. The figures display the brain maps for SSI (left panels) and ISI (middle panels). The activation of the TPJ was mainly at the right hemisphere and is therefore shown on that surface, whereas the activation of the mPFC and the posterior mPFC is shown on the medial surface. The time sequence is shown from top (100 ms) to bottom (700 ms), with on the top, the relevant fMRI locations of goal inferences in the TPJ, trait inferences in the mPFC and Stroop tasks in the pmPFC (adapted from the meta-analysis by Van Overwalle, 2008).

As can be seen for goal inferences depicted in Figure 5, in general, the TPJ is more involved during SSI while the (p)mPFC is more involved during ISI. This general pattern is interrupted by the onset of goal identification. During goal identification and shortly afterwards (~ 225 – 300 ms), the TPJ was most active under both SSI and ISI processing modes. After some more time and beginning at ~ 350 ms, the mPFC activation was strongest for both processing modes, perhaps reflecting deliberative thoughts related to the initial goal identification or other thoughts. These results were largely confirmed by statistical tests as shown in the right column of Figure 5. The only exception is
a short upsurge between 300-350 ms of more TPJ during ISI, which, at first sight, may seem surprising as the left (SSI) vs. middle (ISI) panels show rather an opposite difference. Note, however, that the activation under SSI and ISI is shown in terms of their local maxima (which may differ between brain maps), while the statistical tests calculate these differences directly.

As can be seen for trait inferences depicted in Figure 6, the TPJ (as well as posterior midline structures such as the cuneus) is generally more involved in SSI while the pmPFC is more involved in ISI, and t-tests confirm the greater activation of the pmPFC under ISI. This pattern is interrupted at the onset of trait identification (~ 600 ms), after which the TPJ is most active under SSI, while the mPFC is more active under ISI, as predicted. However, this difference fell short of significance, and the table at the bottom right explains why. Although the activation in the right TPJ is maximal for SSI and reduces (non-significantly) during ISI, the activation in the mPFC is quite similar under both conditions.

Considering the two inferences together (Figures 5 & 6), it is evident that there is a general tendency for the TPJ to be more active during SSI, while prefrontal midline structures (mPFC and pmPFC) tend to dominate ISI. However, the onset of social inferences disrupts this pattern. During and shortly after the onset of goal identification, the TPJ is most active and during trait identification and shortly afterwards, the mPFC is quite active (although the TPJ is more active under spontaneous trait inferences). This is consistent with prior brain imaging research implicating the TPJ in goal-directed judgments, and the mPFC in inferences of traits and stable attributes (see Van Overwalle, 2008). Being aware of inconsistencies and monitoring their occurrence may explain the additional activation of the pmPFC during ISI.

Memory Validation

To validate these LORETA localizations, we also computed correlations with the memory tasks. For goals, we did this for the critical 200-250 ms post-stimulus interval (of the P200) and for traits, we did this for the critical 600 ms post-stimulus time (of the P300). The correlations surpassing the significance levels shown in Figure 7 for goals (A-B) and traits (C-D) are depicted in dark grey. As can be seen, the results suggest that when people are spontaneously making social inferences (left column), better recall cued by the implied goal or trait is associated with more processing in the TPJ (and additionally in the mPFC for trait inferences). In contrast, during intentional instructions (right column), more activation in the mPFC is associated with a stronger
Social Inference

memory for inconsistencies (but only for goals and not for traits, which may be due to the fact that this region is always active for trait inferences). These correlations confirm that the TPJ is involved in spontaneous goal and trait inferences, while the mPFC is involved in deliberation about goal inconsistencies.

Implications and Conclusions

By using state-of-the-art neuroscientific ERP methodology, these results have advanced our understanding of the underlying neural processes involved in detecting and identifying goals and traits of others. We are now capable to answer when goals and traits of others are detected, which brain areas are implicated in their identification, and whether spontaneous versus intentional processing influences the underlying dynamics of this neural circuitry. Note that these results were obtained when reading short sentences (which may reflect what is going on during gossiping and so on), but that more evidence is needed to generalize to (visual) observations in real life situations.

Timing of Social Inferences

The results showed an early P200 component at ~200 ms indexing increased neural activation when an expected goal changed or was missing, and presumably reflects increased attention and processing of the behavioral information elicited by the discrepancy with the goal inferred from the preceding behaviors. A later P300 component at ~ 600 ms for was found for trait inferences and is consistent with earlier research indicating that this component reflects the brain's response to diagnostic and potentially threatening social interaction, when negative behaviors violate an actor's high morality implied during preceding behavioral descriptions (Cacioppo, Gardner & Berntson, 1999, Reeder & Brewer, 1979; Ybarra, 2002). The fact that the P300 was even faster at ~ 400 ms when also the content of the personality trait changed, rules out negative affect or valence as an alternative explanation for these findings because both inconsistent conditions were similar with respect to valence (although valence may potentially play an additional role). Instead, the P300 component presumably reflects a deliberative updating process by which an interpretation of the other person is adjusted and consolidated in long-term memory (Fabiani, Karis & Donchin, 1986; Fabiani & Donchin, 1995).

The early P200 component for goal inferences is a surprising novel finding, because such early inconsistency deflections were not expected on the basis of earlier trait inference research (e.g.,
Bartholow et al., 2001, 2003). This early timing strongly suggests that goals are inferred automatically. In comparison with earlier research on lexical access showing that a preceding sentence phrase may influence the semantic interpretation of ambiguous words as soon as 130 ms post-stimulus (Sereno, Brewer & O’Donnell, 2003), it is intriguing to see how fast goals are identified, requiring a mere extra 50 ms processing time. The fact that the P200 was strongest for goal-irrelevant or no-goal sentences which are relatively neutral to positive in affective content, again rules out negative affect or valence as an alternative explanation for these P200 effects (e.g., Correll, Urland & Ito, 2006; Eimer & Holmes, 2002; Eimer, Holmes & McGlone, 2003; Keil et al., 2002; Kubota & Ito, 2007). Instead, the fast identification of goals is consistent with the growing evidence that goal identification on the basis of behavioral information is part of an automatic system involving the TPJ (cf. Van Overwalle, 2008). Behavioral studies have also demonstrated that people can infer goals automatically without awareness or intent (Hassin, Aarts & Ferguson, 2005) and make goal inferences more rapidly than trait inferences (Van Duynslaeger, Coomans & Van Overwalle, 2008).

Localization of Social Inferences

The ERP timings discussed above show little evidence of timing differences between spontaneous and intentional processing modes. However, these processing differences become more evident when the activation in the different brain areas is considered. In general, spontaneous inferences tend to activate the TPJ more strongly, while intentional inferences lead to more mPFC (and pmPFC) activation. Perhaps, this differential activation reflects different attentional mechanisms or attentional boosts in different brain areas during preparatory or concluding phases when searching for some meaningful information (given spontaneous instructions) or specific details for deliberation (when making intentional judgments). Crucially, these general tendencies are interrupted during particular time windows, when goal or trait inferences are made and specific brain areas are maximally activated. During the 225-300 ms interval, after goal inferences have been identified, the most active brain area irrespective of instruction is the TPJ. This is consistent with the prediction that the TPJ dominates during early and automatic goal processing, and therefore reduces differences between spontaneous or intentional mind sets. During the 600-700 ms interval, when trait inferences are made, one of the most active areas is the mPFC, although the TPJ is maximally active during spontaneous trait processing.
Overall, these results are consistent with earlier fMRI research documenting that the TPJ is recruited during implicit goal identification when observing human-like goal-directed movement (Martin & Weisberg, 2003; Ohnishi et al., 2004; Schultz et al., 2004) or goal-directed behavior (Blakemore et al., 2007; den Ouden et al., 2005; Saxe & Powell, 2006; Saxe & Wexler, 2005; Völlm et al., 2006; Walter et al. 2004), whereas the mPFC is implicated in trait inferences (e.g., Mitchell, Banaji & Macrae, 2005; Mitchell, Cloutier, Banaji & Macrae, 2006; Todorov, Gobbini, Evans & Haxby, 2007) and the pmPFC is involved in conflict monitoring and controlled processing (Gilbert et al., 2006; Miller & Cohen, 2001).

**Implications**

Dual-process models in person perception typically suggest that spontaneous associative processes occur fast, and subserve and inform later intentional thoughts (e.g., Smith & DeCoster, 2000). However, we saw little evidence for such a sequential process of inference making, because the timings of spontaneous and intentional inferences were largely similar. Moreover, each inference recruited a specific brain area during its critical time window for identification, irrespective of a spontaneous or intentional processing mode.

Consequently, the present results seem more compatible with a more elaborated framework proposed by Van Overwalle (2008; see also Keysers & Gazzola, 2007), who argued that goal inferences are quickly and automatically inferred, and further guide trait inferences which require more processing mainly subserved by the mPFC, even under a spontaneous processing mode. Outside the critical time window in which goals and traits were inferred, we saw a general tendency for spontaneous instructions to activate the TPJ, whereas intentional instructions recruit the mPFC (or pmPFC). The selectivity of brain areas recruited for social inferences, but at different timings and areas, for goals and traits, is a unique contribution of recent neurophysiological research. Further studies may uncover the circumstances under which these timings and activations may vary, such as when goals and traits are combined, or when other social inferences such as causal attributions or beliefs by others are inferred. We still have a long way to go to fully unravel and understand the neural correlates of social judgments. Perhaps one future outcome may be to help uncovering the origins of pathologies of person inference, such as in autism and paranoia.
References


## Table 1:

Goal and Trait Implying Descriptions in the Introductory Sentences and in function of Consistency Conditions

<table>
<thead>
<tr>
<th>Sentences</th>
<th>Goal</th>
<th>Trait</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introductory</td>
<td>Washing a car</td>
<td>friendly</td>
<td>Spring</td>
</tr>
<tr>
<td></td>
<td>takes the long green hose</td>
<td></td>
<td>sees dew on the flowers</td>
</tr>
<tr>
<td></td>
<td>drives his family car outside</td>
<td>gives his brother a hug</td>
<td>smells a very fresh scent</td>
</tr>
<tr>
<td></td>
<td>searches for the powerful soap</td>
<td>gives his father a compliment</td>
<td>hears the young birds sing</td>
</tr>
<tr>
<td>Consistent</td>
<td>rubs softly with his sponge</td>
<td>gives his mother a rose</td>
<td>sees a completely blue sky</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>rubs softly with the mud</td>
<td>gives his mother a slap</td>
<td>sees large and big flakes</td>
</tr>
<tr>
<td>Irrelevant</td>
<td>rubs softly with his hands</td>
<td>gives his mother a spoon</td>
<td>sees a very small pimple</td>
</tr>
<tr>
<td>Competence</td>
<td>---</td>
<td>obtained for math an $F$</td>
<td>---</td>
</tr>
<tr>
<td>No-goal</td>
<td>sees a completely blue sky</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

*Note: The order of the introductory sentences was kept fixed for goal sentences. Consistent sentences describe behaviors that are consistent with the implied goal or trait from the introductory sentences, inconsistent sentences are opposite in valence, and irrelevant sentences describe neutral behaviors. Competence-inconsistent sentences imply a trait that is opposite in valence as well as different in personality content. No-goal sentences are part of observation stories depicted in the last column (i.e., observation consistent sentences). These sentences are from Van der Cruyssen et al. (2008) and Van Duynslaeger et al. (2007), and were translated from Dutch with slight adaptations in order to obtain six words in each sentence, with the critical word at the end.*
**Table 2**

*Proportion of Correct Memory as a function of Instruction and Consistency.*

<table>
<thead>
<tr>
<th></th>
<th>Spontaneous</th>
<th>Intentional</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Consistent</td>
<td>Inconsistent</td>
</tr>
<tr>
<td>Cued Recall Goals</td>
<td>14&lt;sub&gt;a&lt;/sub&gt;</td>
<td>04&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>Traits</td>
<td>07&lt;sub&gt;a&lt;/sub&gt;</td>
<td>00&lt;sub&gt;c&lt;/sub&gt;</td>
</tr>
<tr>
<td>Sentence Completion</td>
<td>29&lt;sub&gt;a&lt;/sub&gt;</td>
<td>23&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>Goals</td>
<td>05&lt;sub&gt;c&lt;/sub&gt;</td>
<td>08&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

*Note:* Means in a row of the Spontaneous or Intentional condition not sharing the same subscript differ significantly from each other according to a Fisher LSD test, *p* < .05. Correct memory was scored on the basis of verbatim accuracy of the sentences (without the actor’s name) in cued recall or of the last word in sentence completion, although synonyms were allowed. The means on goals were taken from Table 2 in Van der Cruyssen et al. (2008), and on traits were taken from Table 2 in Van Duynslaeger et al. (2007, p. 10).
### Table 3

*Pearson Correlations (across participants) between Memory Measures and positive ERP-deflections at Cz at different time segments as a function of Inference and Instruction*

<table>
<thead>
<tr>
<th>Inference</th>
<th>Interval</th>
<th>Cued Recall SSI</th>
<th>Cued Recall ISI</th>
<th>Sentence Completion SSI</th>
<th>Sentence Completion ISI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goals</td>
<td>50-300 ms</td>
<td>-0.14</td>
<td>0.55**</td>
<td>-0.19</td>
<td>0.48*</td>
</tr>
<tr>
<td></td>
<td>300-450 ms</td>
<td>0.13</td>
<td>0.41*</td>
<td>-0.08</td>
<td>0.48*</td>
</tr>
<tr>
<td>Traits</td>
<td>50-300 ms</td>
<td>0.21</td>
<td>0.23</td>
<td>0.46*</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>300-450 ms</td>
<td>0.25</td>
<td>0.22</td>
<td>0.11</td>
<td>0.22</td>
</tr>
</tbody>
</table>

*Note:* The % correct memory for goal-consistent and goal-inconsistent sentences was averaged before correlations (across participants) were computed with the P200 and P300 peaks in each interval. The correlations on goals were taken from Table 3 in Van der Cruyssen et al. (2008), and on traits were computed from the raw data of Van Duynslaeger et al. (2007).

* * p < .05; ** * p < .01
Figure Captions

**Figure 1.** Brain Locations of social judgments. The brain maps are oriented from the posterior to the anterior cortex and display either a medial view of the left hemisphere or a lateral view of the right hemisphere. Axes are set according to the Talairach atlas. [A] The TPJ involved in social inferences of intentionality; lateral view of the right hemisphere. [B] The mPFC involved in social inferences of traits (other and self). [C] The goal – trait processing stream: Visual input in the STS is propagated to the TPJ where the goal is identified, and then further propagated to the mPFC where trait inferences are made. Adapted with permission from Figure 1 of Van Overwalle (2008). Copyright 2008 by Wiley.

**Figure 2.** An illustration of ERP averaging. The EEG waveforms are time-locked at the onset of the stimulus (denoted by 0 ms) given a time window of -250 to 500 ms, and are then averaged into the waveform shown at the bottom.

**Figure 3.** Illustrations of ERP inconsistency findings (a positive amplitude is shown downward.). [A] A full line denotes a Black face in a Black context; a broken line denotes an (inconsistent) Black face in a White context. Adapted from Figure 1 in Ito & Urland (2005, p. 23). [B] A full line denotes an inferred trait that is consistent with preceding behaviors, and a broken line a trait that is inconsistent. Adapted with permission from Figure 1 in Bartholow et al. (2001, p. 201). Copyright 2001 by the American Psychological Society.

**Figure 4.** Effects of inconsistency on grand-averaged ERP waveforms showing early (P200) and late (P300) positive deflections at the central midline (Cz) scalp sites given spontaneous and intentional goal instructions. The timeline is given in ms. A positive amplitude is shown downward. [A] Goal inferences: The arrows indicate the onset of a P200 for goal-irrelevant and no-goal words, which are significantly different from goal-consistent words (p<.05). Adapted with permission from Figure 1 in Van der Cruyssen et al. (2008). Copyright 2008 by Psychology Press. [B] Trait inferences: The arrows indicate the onset of a P300 for trait-inconsistent words, which are significantly different from trait-consistent words (p<.05). Adapted with permission from Figure 1 in Van Duynslaeger et al. (2007). Copyright 2007 by Oxford Press.

**Figure 5.** LORETA source analysis of goal inferences. The first two columns depict the amplitudes of the goal-consistent condition minus the no-goal baseline condition under spontaneous
(left) and intentional (middle) instructions, averaged during consecutive post-stimulus time segments. The maps are scaled with respect to their maximum amplitude. The last column depicts maps of t-values, with black indicating greater activation in intentional than spontaneous maps $p<.025$. The brain maps are oriented from the posterior to the anterior cortex and display either a medial view of the left hemisphere or a lateral view of the right hemisphere. Computed from raw EEG data of Van der Cruyssen et al. (2008).

**Figure 6.** LORETA source analysis of trait inferences. The first two columns depict the amplitudes of the trait-consistent condition minus the trait-irrelevant baseline condition under spontaneous (left) and intentional (middle) instructions, averaged during consecutive post-stimulus time segments. The maps are scaled with respect to their maximum amplitude. The last column depicts maps of t-values, with black indicating greater activation in intentional than spontaneous maps, $p<.025$, unless noted otherwise. The t-map at the bottom right shows no significant differences, and instead a table shows the mean $\mu V$ for the right TPJ (rTPJ) and mPFC. The brain maps are oriented from the posterior to the anterior cortex and display either a medial view of the left hemisphere or a lateral view of the right hemisphere. Computed from raw EEG data of Van Duynslaeger et al. (2007).

**Figure 7:** LORETA maps of significant correlations (across participants) between cued-recall memory measures and mean source amplitude $\text{[A-B]}$ at 200-250 ms post stimulus for goals and $\text{[C-D]}$ at 600 ms post stimulus for traits. Dark grey denotes correlations surpassing the significance levels indicated at each map. The brain maps display a lateral view of the right hemisphere or medial view of the left hemisphere oriented from posterior to anterior. The inset displays a frontal view. $\text{[A-B]}$ are adapted with permission from Figure 3 of Van der Cruyssen et al. (2008). Copyright 2007 by Psychology Press. $\text{[C-D]}$ are from raw EEG data of Van Duynslaeger et al. (2007).
<table>
<thead>
<tr>
<th></th>
<th>Spontaneous</th>
<th>Intentional</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Cued Recall of Goal-Consistent Sentences</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TPJ p &lt; .05</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Cued Recall of Goal-Inconsistent Sentences</td>
<td>mPFC p &lt; .001</td>
</tr>
<tr>
<td>C</td>
<td>Cued Recall of Trait-Consistent Sentences</td>
<td>TPJ mPFC p &lt; .001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p &lt; .0025</td>
</tr>
<tr>
<td>D</td>
<td>Cued Recall of Trait-Inconsistent Sentences</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>p &lt; .0025</td>
</tr>
</tbody>
</table>