June 2023

Observed Error Monitoring as an Index of Theory of Mind

Kipras Varkala
University of South Florida

Observed Error Monitoring as an Index of Theory of Mind

by

Kipras Varkala

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Arts
Department of Psychology
College of Arts and Sciences
University of South Florida

Major Professor: Geoffrey F. Potts, Ph.D.
Peter Clayson, Ph.D.
Chad Dube, Ph.D.

Date of Approval:
April 28, 2023

Keywords: social cognition, event-related potential, perspective taking, observed error-related negativity

Table of Contents

List of Tables

List of Figures

Abstract

To thrive within our social environment, we must understand and learn from those around us. Theory of Mind (ToM) refers to our ability to put ourselves in other's shoes to comprehend what they are thinking. Observed error monitoring (OEM) refers to detecting others' mistakes in real time, which is ultimately thought of as a mechanism that supports our learning from those mistakes without personally having committed them. Current findings suggest that both ToM and OEM are dependent on one another when engaging in social learning. We first emulate the mental state of others to understand their goals motivations. We then monitor for errors in their behavior, determining if their actions align with their goals. In this way, we are able to learn vicariously through the mistakes of others. The relationship between ToM and OEM has not yet been directly studied in a healthy sample. The aim of the study was to validate OEM as an index of ToM performance. OEM was measured using two event-related potentials (ERPs), the observed error-related negativity (oERN) and the observed feedback-related negativity (oFRN), with a social go/no-go task. ToM was measured by the hinting task (HT) and the faux pas recognition test (FPRT). A final sample of thirty-six undergraduate participants from the University of South Florida was used for analysis. The HT as a predictor for the oERN trended towards significance and was found to be negatively correlated with the oERN after adding in combined household income as a control variable post-hoc. No other relationships were observed for the main predictions of the study. The results suggest that the oERN may be a viable

predictor of ToM performance while the oFRN may not be a sensitive enough to measure to detect this relationship.

Introduction (Preface)

Thriving socially requires us to understand and learn from one another. One of the ways that we learn from others is by watching them make errors. Being able to identify others' errors is called observed error monitoring (OEM) (Shane et al., 2008). To recognize that another person made an error, we have to first understand their goals and motivations. This ability to mentalize what another is thinking is called Theory of Mind (ToM) and can be classified within the wider scope of social cognition (SC), which refers to a collection of neurocognitive processes involving the identification, comprehension, and utilization of social cues within one's own actions and behavior (Adolphs, 2009; Ochsner, 2008; Penn et al., 1997).

ToM impairments span across several major mental disorders such as schizophrenia (SZ; Penn et al., 2008), autism spectrum disorder (ASD; Senju, 2003), and social anxiety disorder (SAD; Washburn et al., 2016). SC is a relatively new field of research that currently lacks the same level of validation and reliable empirical evidence as non-social neurocognitive constructs. When examining atypical populations characterized by social impairments, findings in SC often yield conflicting results (Silberstein & Harvey, 2018). In a study by de la Asuncion and colleagues (2015), patients with SZ showed impairment in monitoring their own behavior and not others, whereas in a study by Horat and colleagues (2018), patients with SZ showed impairment in monitoring others behavior and not their own. Different assessment techniques along with slightly different event-related potentials (ERPs) measured might explain why these findings are inconsistent. These conflicting findings demonstrate the potential utility of developing an operationalized description of SC constructs and what their neural indices are within the Research Domain Criteria project (RDoC) framework (Gur & Gur, 2016).

Theory of Mind (ToM)

Making inferences about observed behavior is referred to as ToM (Gehring et al., 2012). ToM is the capacity for individuals to identify others' mental states such as intentions, beliefs, emotions, knowledge, and expectations for when those mental states differ from their own (Leslie, 1987). A classic assessment of ToM is the Sally-Anne test, which has the participant determine where Sally will look for a marble after Anne moved it while Sally was not looking, developed for assessing children with autism spectrum disorder (ASD; Baron-Cohen et al., 1985). An individual with ToM capacities will conclude that Sally will look for the marble in the wrong place, demonstrating awareness that Sally is missing the necessary information to update her expectations. In broader terms, ToM helps us predict what someone else is going to do.

ToM is one of the four proposed social cognition domains which also include social perception, emotion processing, and attributional style/bias (Pinkham et al., 2016). Social perception can be defined as the ability to decode interpersonal cues such as, tone of voice, facial expressions, and body language (Bänziger et al., 2011). Emotion processing is the aptitude for identifying, comprehending, producing, and regulating emotions. Attribution style/bias is the degree to which one attributes events, negative or positive, to the self, others, and situations (Yamada et al., 2019). Pinkham and colleagues (2016) ran a panel of tasks used to assess ToM and the other three SC domains, with findings demonstrating fair-good convergent and discriminant validity (Pinkham et al., 2014), suggesting that these SC skills function independently from one another.

ToM is sometimes erroneously described as empathy. Stietz and colleagues (2019) clarify that ToM is cognitive processing of others' emotions (i.e., mentalizing rather than sharing another's affective state) and empathy is comprised of an entirely different neural network than ToM (see also Dvash & Shamay-Tsoory, 2014). ToM was shown to differentiate in its activation depending on if the inference made was simply cognitive (e.g., "Dave made a mistake.") or required affective processing as well (e.g., "I think Dave is sad.") (Schlaffke et al., 2015). Cognitive and affective ToM are distinguished by having the participant, not only recognize another's mental state, but also rate the emotional load observed in a false-belief story, a type of ToM task that will be discussed in the *ToM Psychometrics* section. In a review by Dvash and Shamay-Tsoory (2014), functional neuroimaging along with lesion-based findings showed activation for cognitive ToM in the dorsal anterior cingulate cortex (dACC), dorsal medial prefrontal cortex (dmPFC), dorsal lateral prefrontal cortex (dlPFC), temporal parietal junction, and the superior temporal sulcus. In a similar review by Abu-Akel and Shamay-Tsoory (2011), affective ToM showed activation in the orbitofrontal and ventromedial cortices, ventral striatum, amygdala, and ventral anterior cingulate cortex. These results suggest that cognitive and affective ToM do not overlap at a neural or a behavioral level.

Observed Error Monitoring (OEM)

Social learning, sometimes referred to as observational learning, occurs through the observation of others' actions and makes it possible to acquire knowledge through others experiences, rather than our own. (Bandura et al., 1963; Pollock & Lee, 1992) One such way we can do this is by watching others make mistakes (OEM; Bellebaum & Colosio, 2014). When watching for others' mistakes, ToM is the mechanism involved in modelling others' beliefs by envisioning their mental state to understand what their action goals are. When there is a violation of the modelled belief (e.g., watching someone spill coffee when trying to drink it), the brain has to track that an error has occurred to learn from the social experience (Joiner et al., 2018).

The anterior cingulate cortex (ACC) has been the putative source of error-related activity and has been implicated in the learning process, with the goal of adjusting behavior for minimizing future errors (Apps et al., 2016; Kolling et al., 2016; Magno et al., 2006, Shane et al., 2008). The ACC has been shown to reflect or "mirror" the same activity for observed errors as well (Davies & Stone, 1995, Koban & Pourtois, 2014; Mann-Wrobel, 2009; Decety et al., 2002). This class of neuron, responsible for mirroring activity, is referred to as a mirror neuron (MN) and appears to be involved in responding to the performance of others (e.g., reward cue, Apps et al., 2016; Frith & Frith, 2001). However, both Spaulding (2013) and Heyes & Catmur (2021) propose that MNs act in a contributory capacity rather than a causal one, putting into question whether the ACC is associated with "intention reading". To explain this, Shane and colleagues (2008) propose that the ACC is part of a hierarchical organization that receives, monitors, and codes input from other brain regions (e.g., ToM regions previously discussed) for the eventual outcome observed, rather than where intention reading actually takes place.

ToM and OEM

ToM develops by the 3rd or 4th year of age and is critical for social learning (Leslie et al., 2004). The inability to develop ToM may contribute to the social impairment demonstrated in children with ASD (Baron-Cohen et al., 1985, Baron-Cohen, 2000). In the Sally-Anne example, those who are unable to put themselves in Sally's shoes will say that the marble is where Anne has hidden it rather than where Sally would actually look. In order to predict what Sally is going to do, the participant has to take on Sally's beliefs and desires about the world in order to make an accurate prediction about her actions (Leslie et al., 2004). A response toward an action prediction based off on Sally's perspective precipitates observed error monitoring; "Sally made an error when looking for the marble."

The literature implicates observational learning as essential for propagating and synchronizing social and cultural norms (Bandura & Walters, 1977; Tomasello, 1999). Young girls in Chiapas, Mexico observe their mother's work the loom for years without being allowed to touch one themselves (Maynard & Greenfield 2006). When finally presented with the opportunity, they are expected to perform as if they have done it before. The girls perform with great accuracy, albeit, at a slower rate. This ability to engage in imitative thought and action representation requires the observer to envision themselves as if they were the ones performing the action, and, in the process, determine which observed actions are appropriate.

In a mouse study, headed by Jeon and colleagues (2009), a deletion of calcium channels in the ACC resulted in impaired observational fear learning (e.g., watching another mouse get repeated foot shocks). Animal studies should be observed cautiously as they may not be translational. In humans, the ACC has also been shown to be involved in active observational learning by using a probabilistic selection task (Kobza et al., 2011).

In functional imaging (fMRI and PET) and ERP (EEG) studies, ToM (Calarge et al., 2003; Lissek et al., 2008) and OEM (Apps et al., 2016; Horat et al., 2017) have shown activation in the ACC. In the patient populations that demonstrate ToM impairment, (Baron-Cohen et al., 1999), empirical and theoretical evidence converges toward functional and structural abnormalities in the ACC with ASD patients who have shown ToM impairment (see Santesso et al., 2011). Evidence in regard to ToM performance correlating with ACC activity is not homogenous as seen in an fMRI study by Brüne and colleagues (2008). They found that when patients with ongoing positive symptoms of SZ were engaged in a ToM task (picture sequencing task; Langdon & Colheart, 1999), they produced less activation in the right ACC, along with the insula, thalamus, and striatum, when compared to healthy controls (HC). Yet, ToM scores in patients with SZ did not differ from those of healthy controls. However, Langdon and colleagues (2001) did find a difference between patients and HC associated specifically with negative symptoms of SZ as opposed to positive symptoms. A few regions also saw greater activation than that of controls (supplementary motor area, dlPFC) suggesting that these regions acted in a compensatory manner to match healthy performance.

In summary, both OEM and ToM appear to be affected by impaired ACC performance and appear to be functionally dependent on one another during the process of social learning. As such, OEM could potentially index ToM performance if the study's current findings are supportive of an interdependent relationship existing between the two constructs. ERPs are one way that we can assess OEM performance at the neurocognitive level.

In another fMRI study, patients with SZ demonstrated significantly worse performance for ToM ability, as measured by the Reading the Mind in the Eyes task (RMET; Baron-Cohen et al., 1997), when compared to their healthy matched counterparts (Hirao et al., 2008). Gray matter volume reductions were observed in the ACC, insula, right superior temporal gyrus (STG), ventromedial prefrontal cortex (vmPFC), left ventrolateral prefrontal cortex (vlPFC), and the dmPFC. Specifically, the vlPFC was associated with poor performance on the ToM task. The RMET has been argued to be a measure of emotion recognition whereas the picture sequencing task, a measure of ToM ability (Fernandes et al., 2018; Oakley et al., 2016). The differences in methodology and participant selection strategies most likely account for the conflicting results.

OEM ERP Components

EEG, fMRI, and PET have all been used in identifying the brain structures that underlie OEM. The choice of utilizing ERPs allows for one of the only easily accessible and direct measures of neural activity occurring in real-time with millisecond temporal resolution (Bartholow, 2010). EEG studies are practical as they are less costly and time consuming compared to other imaging methods. The existing evidence supports the current conceptualization of performance/error monitoring investigation within social settings to warrant its application in the current study (Koban & Pourtois, 2014). Ultimately, ERPs are capable of isolating the influence of numerous neural components in relation to behavioral responses usually observed in cognitive-social paradigms (Bartholow, 2010).

Observed known error monitoring: observed Error Related Negativity (oERN)

The error-related negativity (ERN) was initially discovered in a speeded reaction time task (Ghering et al., 1993l; Renault et al., 1980). The ERN is a negative ERP deflection with a fronto-central scalp distribution peaking at about 100ms following a response error (i.e., “I just made a mistake”) (Dehaene, Posner, & Tucker, 1994). Visual, auditory, and somatosensory modalities have been shown to be indexed by the ERN suggesting that it is the “generic” error detection system (Miltner, Braun, & Coles, 1997). Over the course of over 25 years since its discovery, the ERN has been a robust index of an error response (Gehring et al., 2018).

With the ACC as its putative source, the ERN has been identified to index loss aversion, with punishing (e.g., loud white noise) and no-reward errors eliciting a greater negativity, as opposed to failing to provide a correct response in order to gain money (Ghering et al., 2012; Potts, 2011). This means that the ERN cannot be a simple mechanism for error detection. A level of evaluative processing, within the presence of contextually dependent motivation, must be represented by the ERN as well. The ERN has been considered as a performance monitoring system that recognizes errors with the purpose of adjusting behavior to minimize future errors (behavioral adaptation) and learning from them (Gehring et al., 1993; Scheffers & Coles, 2000, Tamnes et al., 2013).

Originally detected by Miltner and colleagues (2004), the oERN is an ERN-like component that is evoked after observing someone else's errors, as opposed to one's own. The oERN is generated in a social version of an ERN-eliciting task with a near identical source as the ERN. This component occurs anywhere from ~30 and ~150ms (early oERN; e.g., go/no-go) to ~250 and ~300ms (late oERN; e.g., Flanker) depending on the type of task and has a smaller peak amplitude when compared to the ERN (de Bruijn & von Rhein, 2012; Miltner et al., 2004; van Schie et al., 2004). Five studies have assessed the ERN in an observation context and found supporting results on the existence of this component (Bates et al., 2005; Miltner et al., 2004; Koelewijn et al., 2008; van Schie et al., 2004; Yu & Zhou, 2006). These studies of healthy participant samples, demonstrated that the neural elements that generate the ERN are also involved when an individual observes errors being made by another participant.

When participants perceived themselves as similar to the person they were observing, as generated by fake test results provided by the researcher, oERN amplitudes were greater in size when compared to perceptions of no-similarity (Carp et al., 2009; Quandt, 2007). Koban and colleagues (2010) discovered that by generating a competitive (competing reward) and cooperative (joint reward) environment for participants, early oERN amplitudes were present only in the cooperative condition while a late oERN was present during the competition condition. In contrast, de Bruijn and von Rhein (2012) demonstrated that the oERN was comparable between competitive and cooperative contexts. Methodological differences (go/no-go vs flanker tasks) might explain why findings are inconsistent. The complexity of the Eriksen flanker task (Eriksen & Eriksen, 1974) might require greater recruitment of attention and, as a result, might attenuate ERP amplitudes due to some participants not paying attention. The oERN could also be independent of reward and respond only to errors. In turn, a cooperative and competitive environment for participants will help avoid misinterpretation of the results. Comparing ERN amplitudes to oERN amplitudes within-subjects will also aid in validating the presence of the oERN.

Sources of the oERN. ERP source localization suggests the oERN generator is the ACC (Koban et al. 2010). The dorsal ACC (dACC) has been functionally linked to more cognitive processing, whereas the rostral region (rACC) to emotional functionality (Bush et al., 2000). The latter is more closely related to affective processing of errors with strong associations to mood, social context, personality, and performance (Pizzagalli et al., 2006). Poor social skill performance has been associated with rACC hypoactivity (Santesso et al., 2011). An fMRI study using healthy subjects, showed rostral and ventral ACC activity during error observation (Shane et al., 2009). The dACC may be the primary oERN generator with contribution of the rACC. The oERN could also overlap with ERP components associated with visual processing. The N1 component, a negativity generated between ~150 to ~200 ms after a stimulus is perceived (Mangun & Hillyard, 1991), might influence the oERN as it occurs within a similar latency and is widely distributed across the scalp.

A SZ study by Mann-Wrobel (2009) found that oERN amplitudes did not differ between patients with schizophrenia and HC, nor was there a relationship observed between the oERN and ToM scores. This particular study might have suffered from fatigue effect, a questionable effort-reward balance (many trials with low-payout), and an insufficient sample size, as outlined by the author. Fatigue is also proposed to affect findings related to ACC activity if there is an effort-reward ratio imbalance (Lorist et al., 2005).

Observed feedback error monitoring: observed Feedback-Related Negativity (oFRN)

Feedback error monitoring is when knowledge that an error was made or not is not inherently known and is provided through feedback instead (such as a red “X” appearing on the screen informing the participant that an error has occurred). This ERP component is called the feedback error-related negativity (FRN or fERN) and is evoked in the fronto-central region of the scalp, ~200 to ~300 ms after relevant negative feedback (i.e., “I just found out I made a mistake”) (Koban et al., 2012; Miltner et al., 1997). The FRN is very similar to the ERN in terms of its source (ACC) and process (error monitoring) (Ghering et al., 2012). However, a two-source model of an additional frontal (OFC) or medial prefrontal cortex (mPFC) subcomponent, along with the ACC, was found to be a complimentary generator of the FRN (Potts, 2011; Müller et al., 2005; van Veen et al., 2004). This was not the case in a more recent study (Horat et al., 2017). Similar to the ERN, the FRN is a reflection of a multi-modal “generic” error detection system (Miltner, Braun, & Coles, 1997). The FRN has been shown to be reflective of a performance monitoring system through gambling task designs, where participants received feedback about whether they won or lost money (Gehring & Willoughby, 2002). Originally thought to only be indexing worse than expected outcomes, the FRN was discovered to index unexpected outcomes, regardless of their valence (Soder & Potts, 2018). In other words, performance monitoring is achieved through monitoring the salience of errors rather than their valence.

Just as there is an ERN mirroring component during observation of others' errors, in the same fashion, an FRN mirroring component named the oFRN has been observed as well (Bellebaum et al., 2010; Yu & Zhou, 2006). Occurring between ~200 to ~ 350ms, it is invoked after observing another person receive visual or auditory performance feedback (Koban & Pourtois, 2014). The oFRN is smaller in amplitude and has a more posterior topography (Mann-Wrobel, 2009; Yu & Zhou, 2006). The oFRN presence was shown to be dependent on the individual observed being real as opposed to a computer-generated player (Fukushima & Hiraki, 2009). However, in a PET study and fMRI study, the ACC saw activity during both real and computer-generated players, with the former generating greater activity (PET: Gallagher, Jack, Roepstorff, & Frith, 2002; fMRI: McCabe, Houser, Ryan, Smith, & Trouard, 2001).

oFRN amplitudes were present in both a cooperative and competitive environment, but responses were greater for cooperators when compared to competitors (Koban et al., 2012). This observed feedback processing might be less involved with evaluative processing when compared to the oERN. Yet, observing friends, as opposed to strangers, led to greater oFRN activation suggesting an effect of friendship on social motivation (Kang et al., 2010). It might be that the oFRN is not as sensitive in registering differences in low-motivating social interactions as the oERN does with strangers in a cooperative environment or in situations where a sense of similarity is perceived between the individual and the observed performer. Marco-Pallarés and colleagues (2010) also included a "neutral" condition where the observer was asked to only pay attention to the player. This additional condition was proposed to help determine whether the feedback error was related to one's own performance or that of the player. Their findings suggested that only when the observer does not have a competing stake in the game does their oFRN reflect that of the player's. It is worth replicating this design element for both the oFRN and the oERN to further the findings of this relationship.

Sources of the oFRN. Very little source localization has been performed on the oFRN, but the putative source appears to be the ACC (Koban et al., 2012), as seen in the oERN as well. If the oFRN is reflective of two sources, like the FRN, it might help us understand some of the discrepancies that arise between it and the oERN. For example, Frith and Frith (2003) suggest that the mPFC is responsible for decoupling physical state representations from mental ones.

The oFRN might not be as responsive as the oERN to social motivation and might impact observational learning in a social setting. As a result, oFRN measurement will be an exploratory aim that will serve to add to the very limited oFRN literature. This study proposes that the oFRN will have a weaker relationship with ToM performance when compared to oERN amplitudes.

Empathy and OEM

Although empathy has been differentiated from cognitive and affective ToM, it may still have an impact on OEM. Brazil and colleagues (2011) found that in patients with psychopathy, a mental disorder that exhibits antisocial behavior, oERN amplitudes were markedly attenuated when compared to those of controls. One possible explanation is that patients with psychopathy have demonstrated no impairments for ToM (Dolan & Fullam, 2004; Richell et al., 2003), but have shown impairment in empathy when compared to healthy controls (Lee & Lee, 2016; Mullins-Nelson et al., 2006).

Fukushima and Hiraki (2007) managed to find that oMFN (a.k.a. oFRN) amplitudes were correlated with empathy trait scores, whereas, perspective taking was not correlated. An fMRI study discovered the ACCg significantly covaried with emotion contagion, an essential ability to being able to empathize with someone (Lockwood et al., 2015).

The findings are not entirely consistent as Shamay-Tsoory and colleagues (2010) have found that patients with psychopathy only differed from controls in their affective ToM. In contrast to Fukushima and Hiraki's (2007) findings, oFRN amplitudes were unrelated to empathy in one study with an ASD population (Bellebaum et al., 2014) and oERN amplitudes in another with a healthy sample (Quandt, 2007). In the former study, cognitive empathy was also found to be associated with observational learning.

Thoma & Bellebaum (2012) suggest that cognitive empathy (a.k.a. affective ToM) is represented by dorsomedial and ventromedial prefrontal cortices while affective empathy is activated in the ACC, anterior insula, inferior parietal lobe, and inferior frontal gyrus. A quantitative meta-analysis proposes that this network is not distinct, but shared, particularly through the ACC, with cognitive empathy more consistently activated in the left dorsal anterior midcingulate cortex and the anterior insula for affective empathy (Fan et al., 2011). Psychometric and neurological results suggest further exploration into the relationship between empathy, affective ToM, and OEM. Bellebaum and colleagues (2020) come to the conclusion that empathy may modulate performance monitoring after finding that higher trait empathy was associated with a more pronounced oMFN. Quandt (2007) highlights the limitation that the studies that examine this relationship, usually fail to assess how empathetic one feels towards the person they are observing. Inclusion of a state empathy rating scale within this study, along with psychometrics assessing cognitive and affective empathy, may help explain which findings discussed are worth supporting.

Task Rationale

ERP Tasks

oERN & ERN. Commonly used tasks to bring about the oERN are socialized versions of ERN-eliciting tasks. The go/no-go task (Bates et al., 2005) has participants respond to correct trials and inhibit a response toward incorrect trials. The Simon task (de la Asuncion et al., 2015) requires an individual to correctly respond to a stimulus in congruent (click left when “left” appears on the *left* side of the screen) and incongruent (click left when “left” appears on the *right* side of the screen) trials. The flanker task, similar in concept to the Simon task, has participants accurately respond to a stimulus appearing on the center of the screen during the simultaneous appearance of congruent (SSSSS) and incongruent (HHSHH) task-irrelevant “flanking” stimuli (Eriksen & Eriksen, 1974). These tasks are “socialized” by having the participant observe another performer participate in the task (a confederate in this study) (Kobza et al., 2011). Attention is kept by having the participant count the performer’s errors and a cooperative environment will be induced by telling the observers that they will be monetarily awarded based on a combined accurate account of the confederate’s performance and the confederates account of the participants performance (van Schie et al., 2004). A competitive environment will be generated by letting the players know that only the one with the better performance will win more money (Koban et al., 2012) If the competitive condition does not show differential activation from cooperation condition, then the results would suggest that only the absolute value is coded during observed error monitoring and, as de Bruijn and colleagues (2009) found, the striatum may be more sensitive to rewards during observation than the ACC-generated oERN.

Only real-life designs might be able to elicit the oFRN (Fukushima & Hiraki, 2009). Although not directly studied, might also be the case for the oERN. The complexity of the task may impact oERN amplitude; the more complex flanker task, which requires mismatch detection, elicited a smaller oERN than the simpler go/no-go task (Koban & Pourtous, 2014). Social context of similarity most likely moderates the strength of the oERN (Carp et al., 2009; Quandt, 2007). A real-life social go/no-go task within a neutral, cooperative, and competitive environment should be the most suitable type of task for this study.

oFRN & FRN. Paradigms used to evoke this potential have been either performance-based (good/bad) or economic in nature (e.g., gain or loss), as shown in a study by Holroyd and colleagues (2003). The observation condition setup for either type of task is the same as for the oERN. To generate the oFRN using monetary feedback, gambling task designs have been used (Fukushima & Hiraki, 2009; Yu & Zhou, 2006). The latter two studies had a similar design where the player was presented two cards labeled "5" and "25" side by side. After the player chooses one of the cards, feedback in the form of a monetary gain or loss was provided. In performance-based designs, the social version of the flanker task, where a confederate intentionally commits errors to generate negative feedback trials, has been used (Clawson et al., 2014; Mann-Wrobel, 2009). A social version of the go/no-go task with a feedback component has also been employed (Koban et al., 2012). Since it has been discussed that task complexity can influence oERN amplitude strength, the same might hold for the oFRN. The social go/no-go feedback task has been chosen for this study. The only difference between this version, and the one that will be used for the oERN, is that the go/no-go stimulus will be a colored arrow that is preceded by a black arrow that acts as an orientation primer. In order to get positive feedback, the player has to respond correctly to an arrow that matches the orientation of the preceding arrow and the "go" color in the stimulus. Any other combination or too slow of a reaction time will result in negative feedback. Difficulty of the game is modulated through online performance tracking of the go/no-go stimulus presentation. The same type of attentional control check and cooperation technique, as the one used for the oERN social go/no-go task, will be used. Many of the design considerations for the oERN also apply to the oFRN as well, namely, realism and social context. Considering how these ERPs are sensitive to their social environment and the design of the task, a real-life social go/no-go feedback task, within a neutral, cooperative, and competitive environment, will be implemented in this study.

Empathy Psychometrics

The Interpersonal Reactivity Index (IRI: Davis, 1983) has seen the most widespread use when measuring for dispositional empathy and its component parts. This questionnaire is made up of four subscales that assess cognitive empathy ("fantasy" scale + "perspective taking" scale) and affective empathy ("empathetic concern" scale + "personal distress" scale). The fantasy scale captures the degree to which one envisions themselves in an imaginative scenario. The perspective taking scale assesses their ToM ability. The empathetic concern scale determines how compassionate one is to the distress of another. The personal distress scale measures how much distress does one experience when watching someone else suffering. The downside to employing the IRI is that the prima facie evidence does not differentiate from cognitive empathy (affective ToM) and cognitive ToM. This 'perspective taking' item for example, "I sometimes find it difficult to see things from the 'other guy's' point of view," is an assessment of cognitive ToM. Whereas this other 'perspective taking' item says, "Before criticizing somebody, I try to imagine how I would feel if I were in their place," is a measure of affective ToM. The Empathy Quotient (EQ: Baron-Cohen & Wheelwright, 2004) could be argued to be suffering from the same issues in regard to measuring for cognitive empathy. It is similar in design to the IRI and was designed in response to the shortcomings of the IRI and other empathy-measuring questionnaires (Allison et al., 2011). Particularly, to be sensitive enough to capture differences within healthy populations and between genders, with females usually scoring higher. Considering that a healthy population is being assessed in this study and the presence of evidence in support of its reliability and validity in measuring for empathy, the EQ has been chosen for this study to assess affective empathy only (Lawrence et al., 2004).

Affective ToM/Cognitive Empathy Psychometrics

Reading the Mind in the Eyes task (Eyes/MIE/RMET: Baron-Cohen et al., 1997), which requires participants to accurately pair an emotion with an image of a pair of eyes, has been commonly used as a standard measure for assessing affective ToM. It has also been shown to not be associated with cognitive ToM (Ahmed & Miller, 2011). It has been argued to exhibit weak psychometric properties, as Healey and colleagues (2015) suggest, and that it actually assesses emotion recognition instead of ToM (Pinkham et al., 2016). An event-related fMRI study by Mier and colleagues (2010) developed and validated a task specific to assessing affective ToM and emotion recognition separately by having participants make decisions based on the face's intention and emotion separately. The two components were positively correlated and had overlapping activation suggesting little differentiation. Within a natural environment, making inferences about someone's emotional state is closely in line with recognizing what emotional behaviors they are exhibiting (Premack & Woodruff, 1978). In light of its widespread use and for scores to vary within a healthy population, the RMET will be employed as a measure of affective ToM in the current study (Gallant & Good, 2019; Yildirim et al., 2011).

Cognitive ToM Psychometrics

Pinkham and colleagues (2014, 2016) evaluated a variety of SC-domain-measuring tasks for test-retest reliability, validity, internal consistency, simplicity, and sensitivity to group differences. In regard to ToM, they concluded that the hinting task (HT; Corcoran et al., 1995), which has participants comprehend indirect speech and make inferences of the mental states of others, demonstrated strong psychometric properties that warrant a recommendation for clinical application. The HT is a type of false-beliefs task (FB), considered as the "gold standard" for measuring ToM, involves having the subject understand that another individual can have false beliefs that are contrary to one's own "correct" knowledge (Dennett, 1978). The HT has also been shown to be an exclusive measure of cognitive ToM (Lindgren et al., 2018). Similar to the HT, the faux pas recognition test (FPRT; Gregory et al., 2002; Stone et al., 1998), has the participant identify, from a variety of scenarios, whether a character has made a faux pas (FP; a blunder) or not. A version of a FB, the picture sequencing task (PST-FB), originally formulated by Baron-Cohen and colleagues (1986), was improved upon by Langdon and Colleagues (2006) to not depend on the performer being capable of adequate executive functioning or intelligence (IQ). This makes it applicable to not just normal, but also atypical populations (e.g., SZ, schizotypal personality disorder) that may demonstrate impairments in the latter facets, but not the former. It appears that the HT and the modified PST-FB are good contenders for assessing ToM. However, Mallawaarachchi and colleagues (2018) found low convergent validity ($r = .38$, $p < .01$) between the two measures, suggesting that one might be a more accurate measure of ToM than the other. Guastella and colleagues (2015) propose that the HT, along with the FPRT, stand out against the PST-FB, in that it incorporates the social nuances that transpire during normal social communication.

Among the tasks evaluated for this study, the HT and the FPRT appear to be the most favorable tasks for measuring ToM. The HT is used to assess ToM impairment in atypical populations with standard versions of the task usually exhibiting a ceiling effect in controls (Marjoram et al., 2005). Therefore, a more difficult version of the task, derived from Marjoram and colleagues (2005) will be employed in this study to avoid ceiling effects. The FPRT has also been found to have a varied distribution of scores among healthy controls (Shamay-Tsoory, Tomer, Berger, & Aharon-Peretz, 2003; Zhu et al., 2007). The HT proves to be a good overall measure, whereas, the FPRT is multidimensional and will aid verifying whether ToM is single or multi-factored through an exploratory factor analysis.

Relationships between ToM and Empathy psychometrics are predicted to have negative linear correlations with the observation-ERPs as similar distributions have been observed in an ERP (Fukushima & Hiraki, 2009) and an fMRI study (Lockwood et al., 2015).

The studies that have been discussed have varied in their methodologies, psychometric properties, administration techniques, age, and small sample/effect sizes (see Albrecht & Bellebaum, 2021; Healey et al., 2015; Hinterbuchinger et al., 2018; Mann-Wrobel, 2009). Age has also been shown to be negatively correlated with performance monitoring (Mathewson et al., 2005) and is a common control variable in OEM studies. These dissociations of measurement from one study to another make for difficult assessment and interpretation of current findings. Some of these limitations will be addressed, such as by including multiple ToM assessments, generating a cooperative and competitive environment, using age and sex as control variables, and an a priori power analysis to determine sufficient sample size. While there are decades of supporting research on how these two constructs might be related to each other through social learning, validating OEM as an index of ToM performance in a healthy sample still remains to be attempted.

Hypotheses[1]

1) Theory of mind performance is negatively correlated with the observed known error monitoring.
    a. oERN amplitudes, as measured by the social go/no-go task, will be negatively correlated with HT scores.
    b. oERN amplitudes, as measured by the social go/no-go task, will be negatively correlated with FPRT scores.
2) Theory of mind performance is negatively correlated with observed feedback error monitoring.
    a. oFRN amplitudes, as measured by the social go/no-go feedback task, will be negatively correlated with HT scores.
    b. oFRN amplitudes, as measured by the social go/no-go feedback task, will be negatively correlated with FPRT scores.
3) Theory of mind performance will have no correlation with self-error monitoring for known and feedback-based errors.
    a. ERN amplitudes will not be correlated with HT and FPRT scores
    b. FRN amplitudes will not be correlated with HT and FPRT scores.
4) Participants will have comparable observed known error monitoring performance in the neutral and cooperative conditions and inverse observed known error monitoring performance in the competitive condition.

[1] Since these ERPs are negative, and more negativity is representative of greater activation, a negative correlation actually means a positive association with the variable they are being correlated with.

a. oERN amplitudes will not be statistically different during the neutral and cooperative conditions.
b. oERN amplitudes in the competitive condition will be significantly more negative than the neutral and cooperative conditions.

5) Participants will have comparable observed feedback error monitoring performance in the neutral and cooperative conditions and inverse observed known error monitoring performance in the competitive condition.
    a. oFRN amplitudes will not be statistically different during the neutral and cooperative conditions.
    b. oFRN amplitudes in the competitive condition will be significantly more negative than the neutral and cooperative conditions.
6) Observed known error monitoring is more representative of theory of mind performance when compared to observed feedback error monitoring.
    a. oERN amplitudes will have a stronger negative correlation with HT scores compared to oFRN amplitudes.
    b. oERN amplitudes will have a stronger negative correlation with FPRT scores compared to oFRN amplitudes.

Methods

Participants

An a priori power analysis using G*Power 3.1.9.7 was used to determine the preferred sample size for the within-subjects study. If the desired medium effect size parameters are to be met (f^2=0.35, α=0.05, β=0.95). At least 40 participants are to be recruited from the University of South Florida psychology department through student research participation (SONA). Participants were excluded if they had a history of mental illness and if they have consumed alcohol or other psychoactive substances (regardless of whether prescribed or illicit) within the past 24 hours. 40 participants were recruited through SONA. 4 participants did not have enough clean trials to use for analysis. 36 participants and subsets of this sample size were used in the final analysis. Sample size breakdown by type of ERP used for analysis can be seen in Table 1 and by psychometrics in Table 2.

Measures

Demographics

A short digital form was administered asking for the following demographic information: gender, age, handedness, ethnicity, race, socioeconomic status, family mental illness history, any current mental disorders, recent consumption of any drugs or alcohol, and substance abuse. Demographic statistics can be seen in Table 3.

Empathy Quotient (EQ)

The questionnaire is made up of 60 questions. 40 items pertain to empathy while 20 items are filler questions. Participants are prompted to choose from one of four responses, strongly disagree, slightly disagree, slightly agree, and strongly agree. Items are scored one point for mild responses and two points for strong responses. Using results from a factor analysis by Lawrence and colleagues (2004), 9 items that exclusively loaded onto the factor 'cognitive empathy' (19, 25, 26, 41, 44, 52, 54, 55, 58) will be excluded from analysis. The other two factors, 'emotional reactivity' (6, 21, 22, 27, 29, 32, 36, 42, 43, 48, 50, 59) and 'social skills' (1, 4, 8, 12, 14, 35, 57), comprised of 19 total questions (range: 0 – 38 points), will be retained for analysis.

Reading the Mind in the Eyes task (RMET)

A revised version for adults will be used (Baron-Cohen et al., 2001). The test is comprised of 36 photographs of eye pairs and requires the participant to identify, "Which word best describes what the person in the picture is thinking and feeling?" out of a set of four labeled emotions. As a control question, prior to each emotion identification, participants will also be asked to identify to which sex the eyes belong. Responses to emotion identification questions where the corresponding control question is incorrectly answered will be omitted from total scoring. Behavioral performance will be measured in terms of accuracy (percentage of items answered correctly).

State Empathy

Derived from Bellebaum and colleagues (2013), participants will be asked to rate their current empathetic attitude toward the other player at the end of each observation block. The questionnaire is made up of six items (as seen below) in which they have a five-point scale to use for their responses (not at all, a little, a decent amount, a lot, very much). The items are further divided into congruent and incongruent and are scored as two separate mean scores accordingly.

1. How happy were you for the other participant when they were gaining money? (congruent)
2. How angry were you about the other participant's gains? (incongruent)
3. Did you pity the other participant when they were losing money? (congruent)
4. Were you angry with the other participant when they were losing money? (incongruent)
5. Were you angry together with the other participant when they were losing money? (congruent)
6. Were you delighted when the other participant was losing money? (incongruent)

Hinting Task (HT)

The task comprises 10 passages of a conversation between two characters. At the end of each scenario, one of the characters concludes with an obvious hint. The participant is then asked to interpret the meaning of what the character is alluding to. If the participant is struggling to come up with the right answer, a following hint is provided with a more specific follow-up question to potentially receive partial points. An example is provided below. The task and scoring procedure can also be found in the appendix.

Example:

Alan is watching television, and his wife Jill sits down to join him. She says: "I see you're watching the football. Isn't there anything else on at the moment?"

Question: What does Jill really mean when she says this?

Hint: Jill then says to Alan: “I thought there was a good play on the other channel”.

Question: What does Jill want Alan to do?

Faux Pas Recognition Test (FPRT)

Adapted from Gregory et al. (2002) and Stone and colleagues (1998), participants are read 10 stories that involve a social faux pas and 10 control stories that do not contain a FP incident within them and help rule out incomprehension of a similar FP story. In addition to having the stories read to them, participants are also provided the stories so they can read along. This is to reduce any working-memory influences. Regardless of which story is read, the participants are asked a series of questions to determine if they have understood and recognized that a FP has occurred. A maximum of 40 points can be earned during this task (1 point for each FP related question and 1 point for each control story identified). In addition to an overall score, sub scores can be calculated for FP detection, understanding inappropriateness, intentions, beliefs, and empathy. The FPRT can be found in the supplementary material and an example of a passage can be seen below.

Example:

Helen’s husband was throwing a surprise party for her birthday. He invited Sarah, a friend of Helen’s, and said, “Don’t tell anyone, especially Helen.” The day before the party, Helen was over at Sarah’s and Sarah spilled some coffee on a new dress that was hanging over her chair. “Oh!” said Sarah, “I was going to wear this to your party!” “What party?” said Helen. “Come on,” said Sarah, “Let’s go see if we can get the stain out.”

Did anyone say something they shouldn't have said or something awkward?

If **yes**, ask:

Who said something they shouldn’t have said or something awkward?

Why shouldn’t he/she have said it or why was it awkward?

Why do you think he/she said it?

Did Sarah remember that the party was a surprise party?

How do you think Helen felt?

Control question: In the story, who was the surprise party for?
What got spilled on the dress?

Social go/no-go Task (ERN/oERN)

Based off the design by Bates and colleagues (2005) and Koban and colleagues (2012), participants are seated next to a confederate posing as another student and both are provided instructions on the task. The confederate's hand/keyboard clicks are obstructed from the participant's view while the confederate's responses are time-locked to a yellow border appearing around the edges of the screen. A visualization of the setup can be seen in Figure 1.

In a randomized sequence, the letters X and K will be presented in a black color (80% X, 20% K) in the center of the screen against a white background. Both will be informed that a response is required anytime the letter X appears and no response when the letter K appears; regardless of color. The stimulus will appear for 250 ms with the interstimulus interval randomly varying between 1.5, 1.75, 2, 2.25, and 2.5 s. The message "too slow" will appear if reaction time is slower than 450 ms in black letters for 1 s. A total of twelve blocks (48 go: 12 no-go trials per block) with the pair taking turns in playing two blocks as the observer and two blocks as the player with the order being counterbalanced for who observes first and whether the condition is cooperative or competitive. The first condition, however, will always be the "neutral" condition as to avoid the influence of the incentive paradigm on the participant. During the neutral condition, the pair is told to count every error observed. During the cooperation condition, the pair is told that half of the other player's points will be awarded in the final calculation of who wins the overall point for the round. During the competition condition, the pair is told that only the player with the best performance of the round will be awarded a 10-dollar Amazon gift card. The player with the highest number of round points at the end of the study will receive the gift card. Unknown to the participant, they will get a 10-dollar Amazon gift card regardless of their performance. The sequence of events can be seen in Figure 2.

Social go/no-go Feedback Task (FRN/oFRN)

Adapted from Koban and colleagues (2012), the only difference between this version, and the one that will be used for the oERN, is that the go/no-go stimulus will be a blue or red colored arrow (350 ms - online) that is preceded by a black arrow (random time of 1-2 s) that acts as an orientation primer. In order to get positive feedback; ‘fast hit’, the player has to respond correctly and fast enough to an arrow that matches the orientation of the preceding arrow and the “go” color (blue) in the stimulus. Any other combination or too slow of a reaction time results in negative feedback; a “miss or slow hit”. Prediction difficulty in the game is modulated through online performance tracking (mean of correct hit RTs and the base stimulus presentation time of 350 ms) and subsequent adjustment of go/no-go stimulus presentation. Correct hits are ones that are made within the online stimulus presentation window and Correct no-hits (during no-go trials) that are not made within the same time window. A feature that the participants are not made aware of. Feedback is provided 1 s after response, with either a green dot for when a correct keypress is made within the allotted time (fast hit) during go-trails and when no keypress is made during no-go trials. A red dot will appear after a correct keypress is made outside of the allotted time (slow hit), when a keypress is made during a no-go trial, and when a no response is detected during go-trials. The same design for number of trials, blocks, ISI’s, attentional control checks, and condition techniques will be used as the one for the social go/no-go task. The sequence of events can be seen in Figure 3.

Procedure

Participants will fill out the informed consent and complete the demographic questionnaire on a computer. Participants will then be administered all of the psychometric tests and all of the ERP tasks in counterbalanced order using a Latin square design. The only exclusion being the state empathy assessment that will be administered after each observation block within the ERP tasks. For the psychometric tasks, the researcher will administer the HT and FPRT in the form of a structured interview while the EQ, RMET, and state empathy tests will be taken on the same computer used for the ERP tasks. For the ERP tasks, the EEG net is applied and the participant is introduced to the confederate. Both the participant and confederate will receive the same instructions for each task and will be told who is observing or performing first. For both ERP tasks, the neutral condition will be introduced first and the performers will have a shortened practice block consisting of 20 go trials and 5 no-go trials prior to every real task. Upon completion of all tasks, the participant will receive the pay-out, regardless of their performance, and be debriefed about the nature and use of deception in the study.

Data Processing

Electroencephalographic (EEG) Acquisition & ERP Extraction Procedure

EEG data will be collected using 128- electrode geodesic sensor nets (Magstim EGI) through Net Station 5.4 acquisition software at a sampling rate of 250 Hz with .1-100 Hz analog filtering, referenced to the vertex. For offline filtering, a 0.10 Hz highpass and a 20 Hz lowpass filter will be applied. Epochs will be segmented to 1000-ms from 200ms before to 800ms after onset of either response or feedback. Using the Net Station artifact detection tool, artifacts, such as eye-blinks and movement, will be filtered out automatically and through manual inspection.

Bad channels will then be replaced and the epochs averaged and re-referenced to the average reference. Averaged ERPs will be baseline corrected over the 200ms pre-feedback or pre-response period. To generate the ERPs, initial piloting will take place in order to determine a region of interest (ROI) and time window that is in line with the current setup and previous literature for both the ERN/oERN and the FRN/oFRN. The tasks will be presented on a 19-in Dell monitor with a screen resolution of 640 x 480. Both tasks will be designed and presented using E-Prime version 3.0 (PST Inc., Pittsburg). A mouse will be used to collect player responses. To ensure the oERN is present, the mean amplitude difference between timely (0-450 ms after stimulus presentation) correct responses and incorrect responses (keypress during 'K' trial 0-450 ms after stimulus presentation) will be used along with a within-subjects comparison of the ERN. For confirming the presence of the oFRN, the mean amplitude difference between go-trials, where feedback is attributed to a slow hit, will be used to compare against positive feedback trials using a paired-samples t-test along with a within-subjects comparison of the FRN.

Statistical Analyses

Hypothesis 1.

Null Model for Comparison to Hypothesis 1 (Null Hypothesis where predictor mean = 0)

$$\text{oERN } \mu\text{V} \sim 1 + \text{Age} + \text{Sex}$$

The test of Hypothesis 1a. will use a linear regression model with HT Score as the predictor variable and Age and Sex as constants. A negative R coefficient is predicted.

HT score as the predictor variable for oERN amplitude

$$\text{oERN } \mu\text{V} \sim 1 + \text{HT Score} + \text{Age} + \text{Sex}$$

The test of Hypothesis 1b. will use a linear regression model with FPRT Score as the predictor variable and Age and Sex as constants. A negative R coefficient is predicted.

FPRT task score as the predictor variable for oERN amplitude

$$\text{oERN } \mu\text{V} \sim 1 + \text{FPRT Score} + \text{Age} + \text{Sex}$$

Hypothesis 2.

Null Model for Comparison to Hypothesis 2 (Null Hypothesis where predictor mean = 0)

$$\text{oFRN } \mu\text{V} \sim 1 + \text{Age} + \text{Sex}$$

The test of Hypothesis 2a. will use a linear regression model with HT Score as the predictor variable and Age and Sex as constants. A negative R coefficient is predicted.

HT score as the predictor variable for oFRN amplitude

$$\text{oFRN } \mu\text{V} \sim 1 + \text{HT Score} + \text{Age} + \text{Sex}$$

The test of Hypothesis 2b. will use a linear regression model with FPRT Score as the predictor variable and Age and Sex as constants. A negative R coefficient is predicted.

FPRT score as the predictor variable for oFRN amplitude

$$\text{oFRN } \mu\text{V} \sim 1 + \text{FPRT Score} + \text{Age} + \text{Sex}$$

Hypothesis 3.

The test of Hypothesis 3a. will use a linear regression model with HT Score as the predictor variable and Age and Sex as constants and FPRT score as the predictor variable and Age and Sex as constants. No correlation is predicted.

HT score and FPRT score (separately) as the predictor variable for ERN amplitude

$$\text{ERN } \mu\text{V} \sim 1 + \text{HT Score} + \text{Age} + \text{Sex}$$

$$\text{ERN } \mu\text{V} \sim 1 + \text{FPRT Score} + \text{Age} + \text{Sex}$$

The test of Hypothesis 3b. will use a linear regression model with HT Score as the predictor variable and Age and Sex as constants and FPRT score as the predictor variable and Age and Sex as constants. No correlation is predicted.

HT score and FPRT score (separately) as the predictor variable for FRN amplitude

$$\text{FRN } \mu\text{V} \sim 1 + \text{HT Score} + \text{Age} + \text{Sex}$$

$$\text{FRN } \mu\text{V} \sim 1 + \text{FPRT Score} + \text{Age} + \text{Sex}$$

Hypothesis 4 & 5.

The test of Hypothesis 4 (a & b) and 5 (a & b) will use repeated measures ANOVA to determine whether ERP amplitudes are different across conditions.

Hypothesis 6.

The test of Hypothesis 6 (a & b) will use a multivariate multivariable regression model to account for covariance attributed by predictors toward both dependent variables and be able to compare model fit statistics. Pillai test statistic will be used to determine whether or not to include the predictors prior to comparing model fit with Akaike's Information Criterion.

a. (oERN μV, oFRN μV) ~ 1 + HT Score + Age + Sex

b. (oERN μV, oFRN μV) ~ 1 + FPRT Score + Age + Sex

Table 1. ERP Amplitudes by Composite values and Conditions

ERP (n)	Condition	Min	Max	Mean (SD)
oERN (25)	Composite	-7.17	4.45	-0.90 (2.87)
	Neutral	-4.63	11.55	0.34 (3.61)
	Cooperative	-9.76	4.92	-0.84 (3.95)
	Competitive	-10.52	3.08	-2.20 (3.95)
oFRN (36)	Composite	-5.18	1.02	-1.79 (1.31)
	Neutral	-8.61	1.96	-1.88 (1.85)
	Cooperative	-7.53	4.68	-1.63 (1.94)
	Competitive	-8.95	0.98	-1.79 (1.31)
ERN (20)	Composite	-7.31	5.67	-2.26 (3.15)
	Neutral	-15.98	2.60	-2.85 (4.21)
	Cooperative	-8.78	6.26	-1.77 (3.51)
	Competitive	-9.4	8.16	-2.16 (4.08)
FRN (34)	Composite	-8.14	0.46	-2.64 (1.76)
	Neutral	-8.83	0.59	-2.66 (2.02)
	Cooperative	-8.22	1.15	-2.48 (2.00)
	Competitive	-9.81	1.16	-2.78 (2.59)

Table 2. Psychometrics Summary Statistics by ERPs

Data Set (n)	Psychometric	Min	Max	Mean (SD)
oERN (25)	HT	11	20	17.32 (2.11)
	FPRT	70	97.5	88.81 (7.82)
	RMET	15	32	25.96 (4.85)
	EQ	14	30	21.96 (3.78)
	State Empathy -			
	Congruent	1	3.62	2.12 (0.70)
	Incongruent	1	2.5	1.44 (0.34)
oFRN (36)	HT	11	20	17.53 (2.00)
	FPRT	50.62	98.75	87.27 (10.38)
	RMET	13	32	25.53 (4.90)
	EQ	12	30	21.39 (4.34)
	State Empathy -			
	Congruent	1	3.62	2.03 (0.67)
	Incongruent	1	2.5	1.43 (0.33)
ERN (20)	HT	11	20	17.47 (2.34)
	FPRT	50.62	97.5	86.32 (11.87)
	RMET	13	32	24.78 (5.13)
	EQ	16	30	21.63 (3.57)
	State Empathy -			
	Congruent	1.13	3.40	2.12 (0.63)
	Incongruent	1	1.83	1.36 (0.21)
FRN (34)	HT	11	20	17.76 (1.96)
	FPRT	50.62	100	87.52 (10.84)
	RMET	15	32	25.70 (4.59)
	EQ	12	30	21.64 (4.24)
	State Empathy -			
	Congruent	1	3.62	2.07 (0.64)
	Incongruent	1	2.5	1.40 (0.32)
ALL (13)	HT	11	20	17.38 (2.72)
	FPRT	79.41	97.5	91.37 (5.58)
	RMET	18	32	25.46 (4.35)
	EQ	17	30	22.15 (3.57)
	State Empathy -			
	Congruent	1.13	3.4	2.23 (0.60)
	Incongruent	1	1.83	1.33 (0.24)

Table 3. Final Sample Demographic Information (n = 36)

Category	Group	%
Sex	Male	33.3
	Female	66.6
Race	African American/Black	2.7
	American Indian	0
	Caucasian/White	52.7
	Asian/Pacific Islander	11.1
	Hispanic	25
	Other	8.3
Age	18	47.2
	19	25
	20	22.2
	22	2.7
	23	2.7
Combined Household Income	$0 - $19,999	36.1
	$20,000 - $39,999	11.1
	$40,000 - $59,000	8.3
	$60,000 - $79,000	11.1
	$80,000 - $99,999	2.7
	over $100,000	30.5

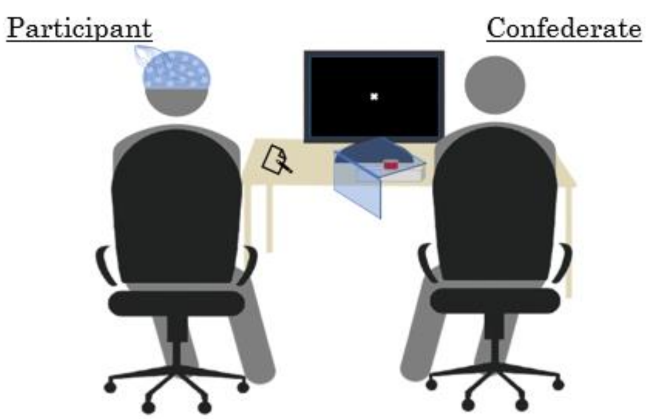

Figure 1. Observation Condition Setup

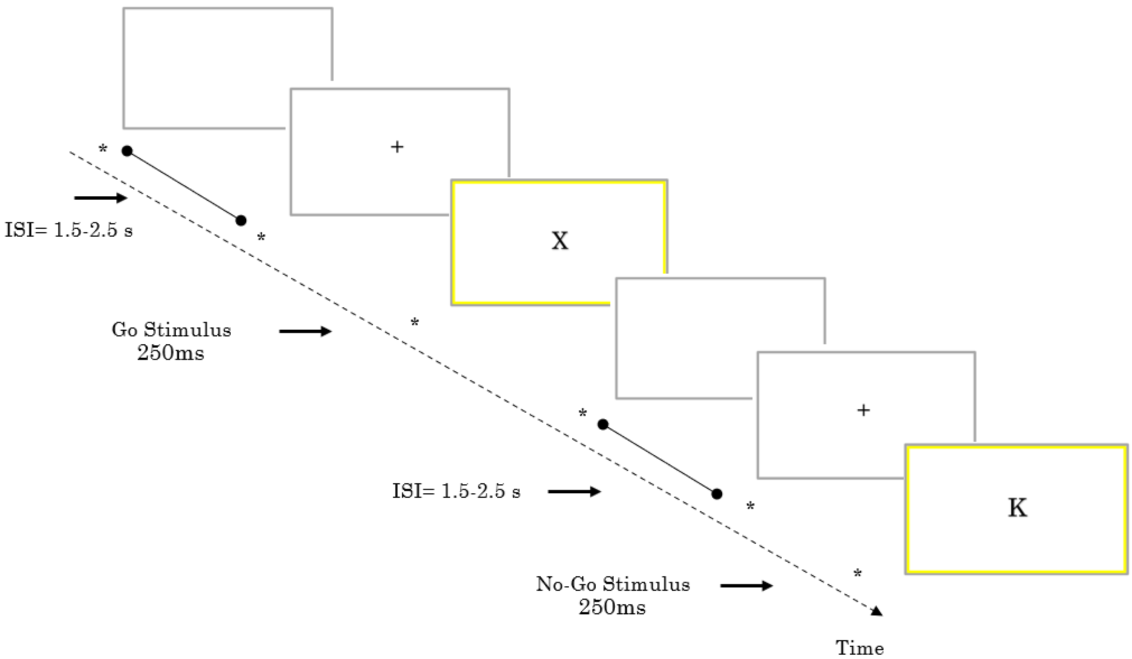

Figure 2. Social go/no-go Task (oERN)

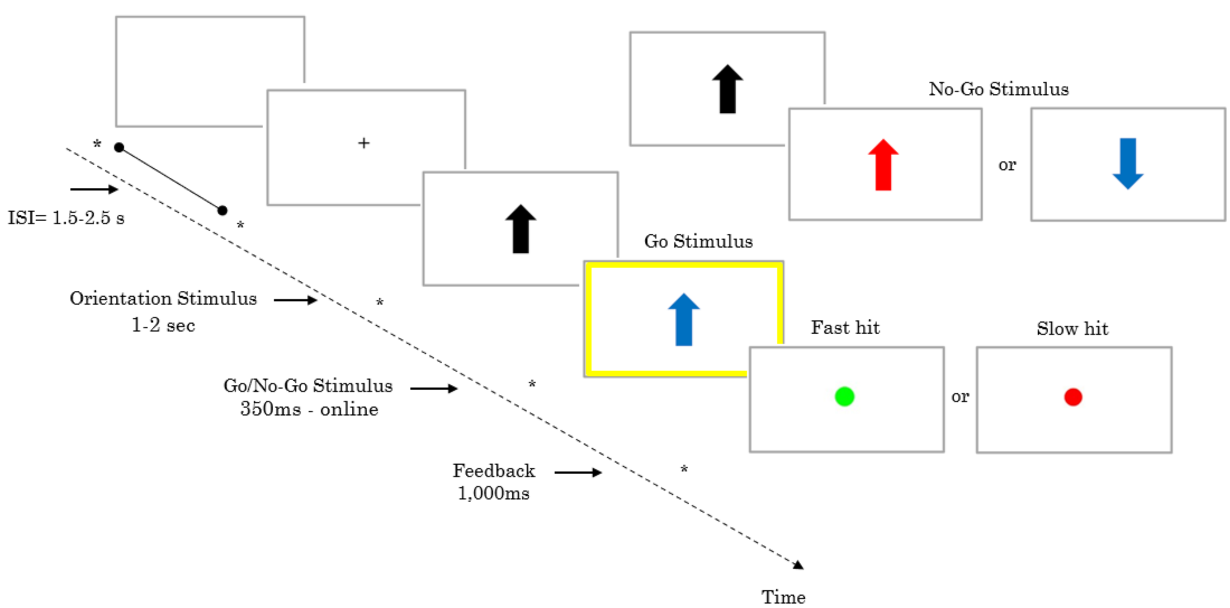

Figure 3. Social go/no-go Feedback Task (oFRN)

Results

Data were tested for model assumptions and analyzed using R 4.2.3 and IBM SPSS 28.

In-line with theoretical knowledge and visual inspection, the following electrode region of interest was used for all ERP measures: 4, 10, 11, 16, 18, 19. For analysis of the ERN/oERN and behavioral data, a 2 (Response: Correct, Error) by 3 (Condition: Neutral, Cooperative, Competitive) repeated measures ANOVA was used to assess for mean differences in amplitude. For analysis of the FRN/oFRN data, a 2 (Feedback: Positive, Negative) by 3 (Condition: Neutral, Cooperative, Competitive) repeated measures ANOVA was used to assess for mean differences in amplitude. Listwise deletion was used to handle missing data. Bonferroni post hoc tests were used to conduct follow-up pairwise comparisons.

Behavioral Data

An exploratory assessment of reaction times (milliseconds) for the Go/No-Go task used to elicit the ERN/oERN, there was a main effect of Response type, F (1,34) = 17.28, $p < .001$, $\eta p^2 = .33$. Erroneous responses (M=369.78, SE=3.30) were faster than correct responses (M=378.62, SE=3.33). There was also a main effect of condition, F (2,68) = 3.63, $p < .001$, $\eta p^2 =$ .09. Follow-up tests showed that participants responded faster in the competitive condition (M=372.90, SE=3.20) and the cooperative condition (M=371, SE=3.18) when compared to the neutral condition (M=378.71, SE=4.25). No interaction was observed. The behavioral data for reaction times can be seen in Table 2.

ERPs

oERN (150-250ms). There was a main effect of Condition, F (2,48) = 4.63, $p < .05$, $\eta p^2 = .16$. The oERN in the competitive condition (M=-1.41, SE=0.48) was more negative when compared to the neutral condition (M=-0.09, SE=0.41), but not to the cooperative condition (M=-0.47, SE=0.37). A significant interaction was observed for Response x Condition, F (2,48) = 4.38, $p < .05$, $\eta p^2 = .15$. Follow-up tests showed the competitive condition was significantly more negative for erroneous responses (M=-2.20, SE=0.79) when compared to the competitive condition for correct responses (M=-0.62, SE=0.29) and the neutral condition for erroneous responses (M=0.34, SE=0.72). The grand average ERP is displayed in Figure 4.

oFRN (200-300ms). A significant interaction was observed for Feedback x Condition, F (2,70) = 5.5, $p < .05$, $\eta p^2 = .13$. Follow-up tests showed the oFRN in the neutral condition was significantly more negative for positive feedback (M=-2.7, SE=0.45) when compared the neutral condition for negative feedback (M=-1.87, SE=0.35) and the cooperative condition for positive feedback (M=-1.35, SE=0.31). The cooperative condition was also significantly more negative for negative feedback (M=-1.88, SE=0.31) when compared to the cooperative condition for positive feedback (M=-1.35, SE=0.31). The grand average ERP is displayed in Figure 5.

ERN (50-150ms). There were no significant main effects or interactions observed. The grand average ERP is displayed in Figure 6.

FRN (200-300ms). There were no significant main effects or interactions observed. The grand average ERP is displayed in Figure 7.

A summary table of all ERPs by condition can be seen in Table 3.

Hypothesis 1

The null model, HT predictor model, and FPRT predictor for oERN amplitude were all found not have significant results. It is worth noting that the HT predictor model almost reached significance, β =-0.58, ${}_{adj}R^2 = .05$, $t(21) = -2.06$, $p = .051$, 95% CI [-1.16, 0]. Considering that there was a monetary incentive in the experiment design, it was suspected that household income could influence the level of attention paid to participation. An exploratory analysis with combined household income as an additional predictor in the HT predictor model revealed a significant relationship, β =-0.68, ${}_{adj}R^2 = .15$, $t(17) = -2.41$, $p = .02$, 95% CI [-1.27, -0.08]. For every 1-point increase in HT score, oERN amplitude was reduced by -0.68 µV. No changes were observed when applying the model adjustment for the null or FPRT predictor model.

The Akaike Information Criterion (AIC) model selection method was used to distinguish which of the models tested fit the data best. Within the original Hypothesis 1 models, the HT predictor model was the most parsimonious and carried 61% of the cumulative model weight. When adding in combined household income as a control variable to all of the models, the HT predictor model was still the most parsimonious, but carried a greater cumulative model weight of 74%. It is worth noting that the FPRT predictor model fit worse than the null model in both cases. When comparing the two leading models against one another. The original model was more parsimonious, holding 8.82 fewer units of AIC, suggesting evidence in favor of retaining the simpler model. However, the adjusted R squared differences between the models suggests that the more complex model accounts for 10% more variation observed in oERN amplitudes.

Hypothesis 2

The null model, HT predictor model, and FPRT predictor for oFRN amplitude were all found not have significant results. Exploratory analyses adding in combined household income into the models yielded non-significant results. Using AIC model selection, the null models fit best in both the proposed and exploratory analyses. Adjusted R squared differences also suggest retaining the null models.

Hypothesis 3a

The null model, HT predictor model, and FPRT predictor for ERN amplitude were all found not have significant results. It is worth noting that the models did not meet the assumption of linearity due to observed sex differences. All attempts at accounting for this difference through various transformations and non-linear regression techniques were unsuccessful. Instead, the models were separated by gender for further analysis. Exploratory analyses adding in combined household income into the models yielded non-significant results. The AIC model selection method was used to distinguish which of the models tested fit the data best. The null models fit best in both the proposed and exploratory analyses. Adjusted R squared differences also suggest retaining the null models.

Hypothesis 3b

The null model, HT predictor model, and FPRT predictor for FRN amplitude were all found not have significant results. Exploratory analyses adding in combined household income into the models yielded non-significant results. The AIC model selection method was used to distinguish which of the models tested fit the data best. The null models fit best in both the proposed and exploratory analyses. Adjusted R squared differences also suggest retaining the null models.

Hypothesis 4

A one way (Condition: Neutral, Cooperative, Competitive) repeated measures ANOVA was used to assess for mean differences in oERN amplitude for erroneous responses. There was a main effect for Condition, F (2,48) = 5.1, $p < .05$, $\eta p^2 = .17$. Simple effects analysis revealed that the oERN in the competitive condition was significantly more negative for erroneous responses (M=-2.20, SE=0.79) when compared to the neutral condition for erroneous responses (M=0.34, SE=0.72).

Hypothesis 5

A one way (Condition: Neutral, Cooperative, Competitive) repeated measures ANOVA was used to assess for mean differences in oFRN amplitude for erroneous responses. There were no significant main effects or interactions observed.

Hypothesis 6

A multivariate multivariable regression was used to assess if there were any significant differences in the null, HT, and FPRT predictor models across oERN and oFRN amplitudes. No significant results were observed.

Psychometrics

An exploratory linear regression analysis was used to determine whether substituting any of the other psychometrics as predictors for Hypothesis 1 and 2 in oERN and oFRN amplitude. All models were found to have non-significant results. Further exploratory analyses revealed that

HT scores and FPRT scores were found to bear no correlation with each other. EQ and RMET scores were found to have a significantly positive correlation with each other, β =0.43, ${}_{adj}R^2$ = .21, $t(34) = 3.25$, $p < .05$, 95% CI [3.37, 17.37]. A summary table of all psychometrics categorized by the ERP used for each analysis can be seen in Table 4.

Table 4. Reaction Times in the Go/No-Go task by Condition

Condition	Type	Mean (SE)
Neutral	Correct	382.78 (4.07)
	Error	374.63 (4.76)
Cooperative	Correct	375.91 (3.30)
	Error	366.08 (3.64)
Competitive	Correct	377.17 (3.47)
	Error	368.63 (3.69)

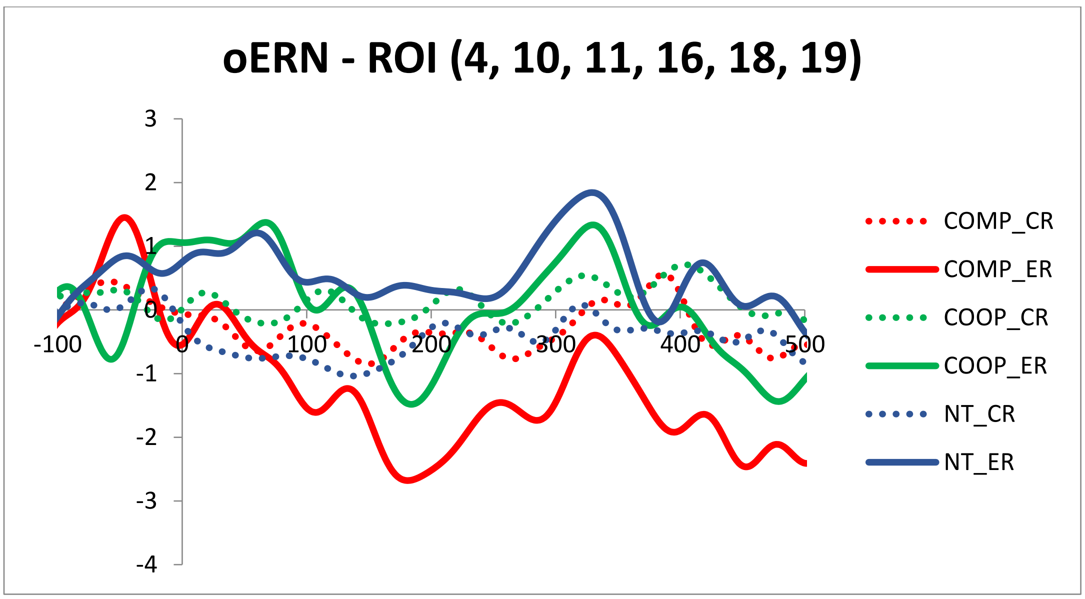

Figure 4. oERN Grand Average Overplot of all Response Types and Conditions.

*NT = Neutral condition, COOP = Cooperative condition, COMP = Competitive condition, CR= Correct response, ER = Erroneous response. ROI = Region of Interest. Shaded area represents the ERP extraction window.

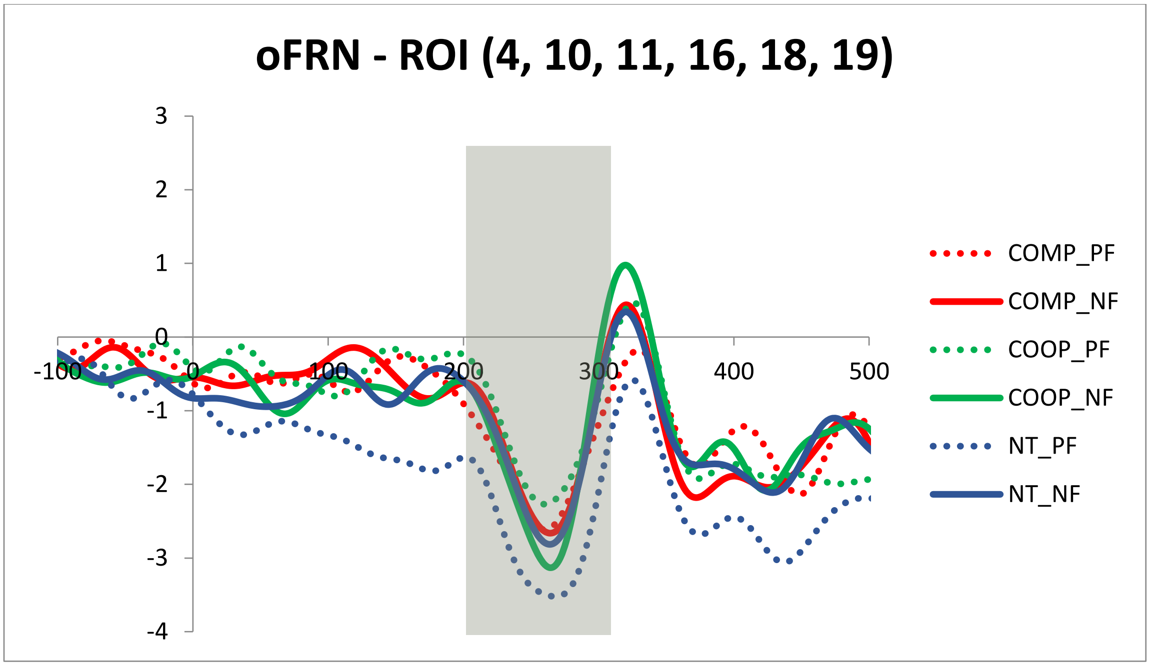

Figure 5. oFRN Grand Average Overplot of all Response Types and Conditions.

** NT = Neutral condition, COOP = Cooperative condition, COMP = Competitive condition, PF= Positive feedback, NF = Negative feedback. ROI = Region of Interest. Shaded area represents the ERP extraction window.*

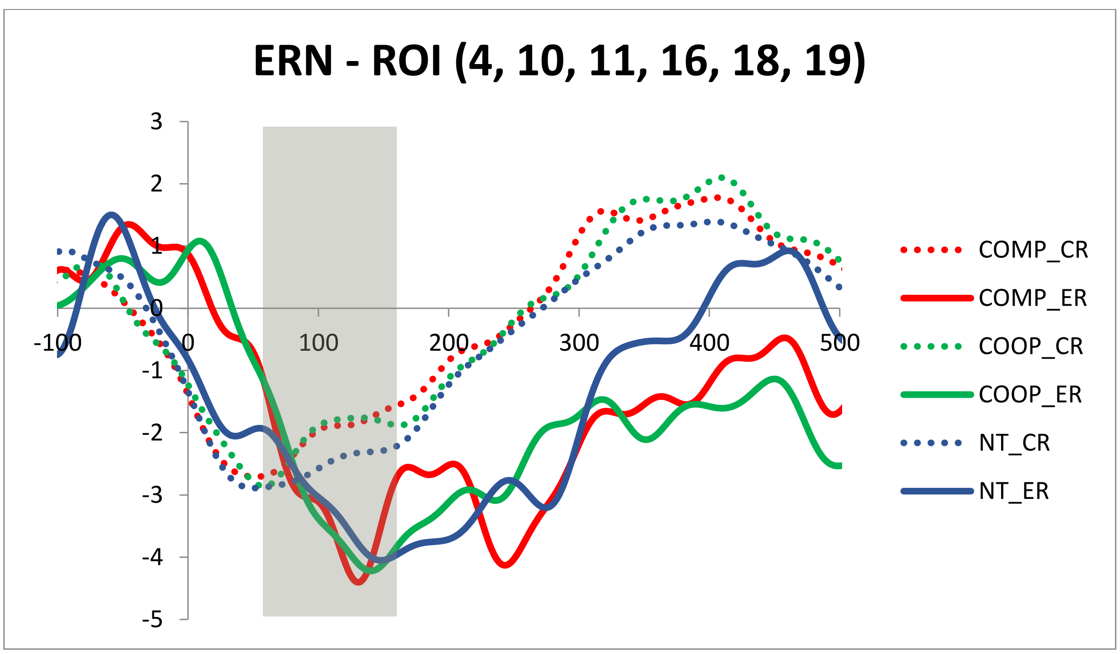

Figure 6. ERN Grand Average Overplot of all Response Types and Conditions**.**

**NT = Neutral condition, COOP = Cooperative condition, COMP = Competitive condition, CR= Correct response, ER = Erroneous response. ROI=Region of Interest. Shaded area represents the ERP extraction window.*

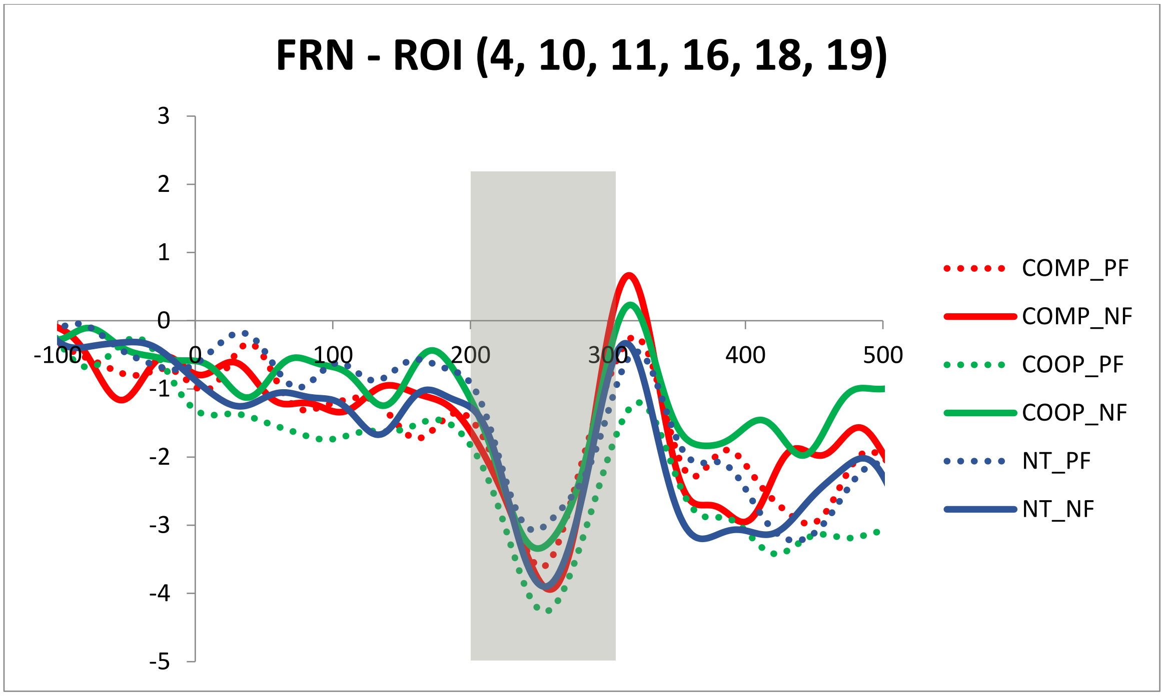

Figure 7. FRN Grand Average Overplot of all Response Types and Conditions.

**NT = Neutral condition, COOP = Cooperative condition, COMP = Competitive condition, PF= Positive feedback, NF = Negative feedback. ROI = Region of Interest. Shaded area represents the ERP extraction window.*

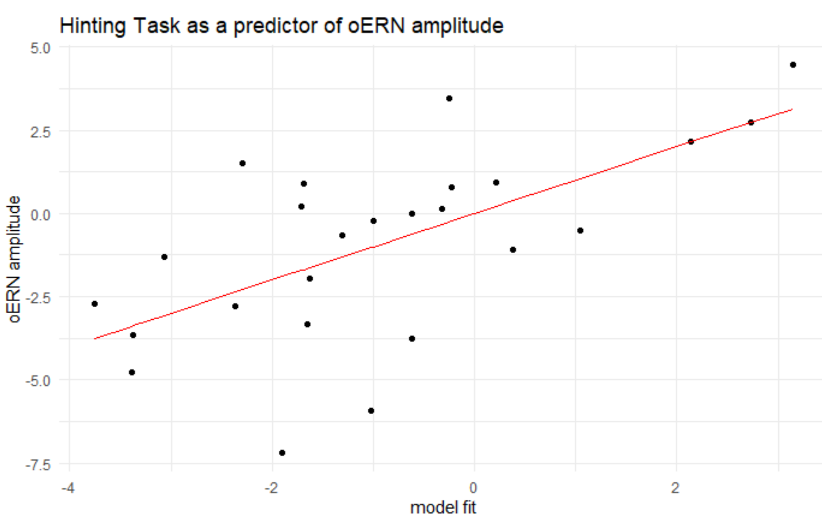

Figure 8. Hinting Task as a predictor of oERN amplitude while controlling for Age, Sex, and Combined Household Income.

Discussion

The purpose of this study was to determine if our ability to cognitively understand another's mental state correlates with the region of activity in the brain that is believed to index observed error monitoring. This study took a novel approach to understand the proposed relationship by using various measures to assess for ToM and empathy, two different measures for OEM, and manipulation of the social context. The results of the study indicate that observed known error monitoring may be an index of ToM performance. These findings can offer an alternative approach to measuring for social impairments in healthy populations and potentially in clinical populations too. It is not overly clear why some of the predicted results were found to be non-significant and some potential confounds were discussed. Regardelss, some model effects and manipulation differences were still observed and are worthy of further discussion.

HT and FPRT as Predictors of OEM Performance

Hypothesis 1 results were somewhat in line with predictions. The leading hypothesis was that HT scores will be negatively correlated with oERN amplitudes. The proposed relationship was found to be non-significant, but the results were still promising with a strong negative correlation ($\beta = -0.58$) and a *p* value of 0.51. Such findings warranted an assessment of any confounding variables that may be introducing error into the prediction's results. Perception toward money was noted as a potential confounding variable during data collection due to some participants explicitly mentioning that they were not interested in pursuing the monetary reward that was used to elicit the competitive and cooperative conditions. Others may have shared the same feeling, but chose to not share their thoughts out of consideration for the experimenters. Considering that several studies have used a monetary incentive to generate motivation in paying attention to the player being observed, it is likely that the lack of interest for offered reward may also reflect a lack of interest toward paying attention in general (de Bruijn & von Rhein, 2012; Koban et al., 2010). The FRN and the oFRN have both been found to be attenuated toward losing money when compared to gaining it (Yu & Zhou, 2006), suggesting that their amplitudes are modulated by perception of reward. Considering that the FRN has been shown to be a type of error monitoring (Ghering et al., 2012), OEM measures, such as the oERN in this case, may reflect a similar relationship with perception of reward. An exploratory analysis of the oERN by condition supports this idea with the competitive condition oERN being significantly more negative than the neutral condition oERN. This relationship between the oERN and perception of reward may be further moderated by one’s relationship with money, as measured by combined household income in this study. A recent publication suggests that those in higher household income brackets have higher loss aversion in a monetary based task design (Gächter et al., 2022). ERPs similar to the oERN, the FRN, has also been found to be positively correlated with loss aversion (Kokmotou et al., 2017). Therefore, the relationship one has with money coming into a study may moderate how they react to potential rewards and losses they are faced with at both a behavioral and neural level. As a result of this investigation, a post-hoc analysis with combined household income as an added covariate to the proposed relationship was run. The relationship between the HT and oERN amplitudes was found to have a strong significant negative correlation and a medium effect size (β =-0.68, ${}_{adj}R^2$ = .15) and can be seen in Figure 8. These findings not only support the main prediction of the study that ToM can be indexed by OEM, but that the generation of the oERN may be dependent on successfully establishing an interpersonal context between the observer and player. In this case, using a monetary incentive to create a competitive and cooperative social environment.

Hypothesis 2 failed to produce significant results in both proposed and exploratory analyses. Additionally, model comparisons suggested that the null models were worth retaining in this case. This tells us that ToM measures only serve to worsen the predicted changes in the oFRN. Ultimately, these findings did not support the prediction that oFRN amplitudes are correlated with HT or FPRT scores. The introduction of the oFRN as a dependent measure in the study was done so in an exploratory manner considering that the background literature had mixed outcomes on the presence (Kang et al., 2010; Koban et al., 2012) and absence (Marco-Pallarés et al., 2010) of the oFRN when the observer had a competing stake in the task. Instead, the results ended up being supportive of the assumption made that the oFRN may not be as sensitive in registering differences in low-motivating social interactions as the oERN was predicted to be (Koban et al., 2012). An additional analysis of whether the oERN, when compared to the oFRN, would be more representative of the relationship between the HT and the FPRT, failed to produce significant results in both proposed and exploratory analyses.

Hypothesis 3a results were in line with predictions, but required splitting up the models by gender in order to conform to the linearity assumption in linear regression. Model results were non-significant in both proposed and exploratory analyses. Additionally, model comparisons suggested that the null models are worth retaining in this case. Hypothesis 3b had similar findings, however, the models did not need to be split by gender in order to be assessed. These findings are in support of the prediction that ERN and FRN amplitudes are not correlated with HT or FPRT scores and adding in either type of ToM measure as a predictor only serves to worsen model fit. A simple explanation is ToM engagement is not required when the participant is performing the task rather than observing someone play. Several studies have shown that the ERN (Romney et al., 2015, Voegler et al., 2018) and the FRN (Voegler et al., 2019) are more negative under observation as opposed to when the participant is alone. However, no differences were observed in the ERN under observation between competitive and cooperative conditions (Koban et al., 2010). These findings suggest that, although the participant's ERN/FRN may be enhanced under observation, it does not reflect self-other processing as indicated by the absence of differences in the ERPs between changes in interpersonal context induced through the competitive and cooperative conditions.

Social Context Manipulation

Go/No-Go task reaction times were found to be faster in the cooperative and competitive conditions when compared to the neutral condition. One explanation is that there is evidence of successful manipulation of reward incentive. Known as the "reward effect", it has been found to exist only when the participant is able to preprogram their response in advance (Mir et al., 2011). The reward effect is supported in the context of this study, because participants were able to anticipate which button they had to press in order to respond to the stimulus. Another explanation could be that there was a presence of the practice effect. The neutral condition always came before the introduction of the monetary incentive conditions in order to control for any motivational changes that could be elicited with the chance of winning a reward. This may have given everyone a chance to practice the game and hone in on their peak performance prior to playing in the non-neutral conditions. The most likely conclusion is that both explanations contribute to the observed findings.

Hypothesis 4 results were partially supported. oERN amplitudes were significantly more negative in the competitive condition when compared to the neutral condition, but not with the cooperative condition. Differences between neutral and cooperative conditions were not observed. As mentioned earlier, the oERN may only code the absolute value of the condition outcome; “their mistake is going to affect my chance of winning the reward” vs “their mistake is not going to affect my chance of winning the reward”. Instead, a study using the same cooperation/competition condition design found that it was within the striatum that rewards were differentiated by condition (de Brujin et al., 2009) What was unexpected was that oERN amplitudes did not differ in the cooperative condition from the competitive condition like they did with the neutral condition. Since one’s gain was also a half gain for the other in cooperation rounds, the relative value of points gained when observing errors was reduced to half of the points that could be gained in competitive rounds. The reduction in the significance of the error observed for the participant may be great enough that it does not differ from when there is no incentive to keep track of error trials.

Hypothesis 5 failed to produce significant results in both proposed and exploratory analyses. No differences were observed in oFRN amplitude between conditions. In conjunction with Hypothesis 2 findings, the assumption that the oFRN is not sensitive to low-motivating social contexts is supported at the condition level as well.

Exploratory Findings

oERN amplitudes being compared by condition and response type were only found be more negative for erroneous responses made in the competitive condition when compared to correct responses in the competitive condition and erroneous responses in the neutral condition. Perhaps the discrepancy between erroneous and correct responses shown in the competitive condition is due to the highest stakes for effortful performance generated by this condition when compared to the other, less motivating conditions. The latter relationship between the competitive and neutral condition for erroneous responses could also be explained by the differences in motivation to pay attention. Overall, mean trends suggest that the conditions and their response types are predicted to differ with a more sufficient sample size.

oFRN amplitudes being compared by condition and feedback type were found to be more negative for positive feedback in the neutral condition when compared to negative feedback in the neutral condition and positive feedback in the cooperative condition. These findings are somewhat surprising considering that negative feedback has been found to generate a larger oFRN when compared to the oFRN for positive feedback (Bellebaum et al., 2010; Koban et al., 2012). The Go/No-Go task used to elicit the oFRN was noted to be fairly difficult to get positive feedback on by most throughout the study and it is possible that positive feedback was more informative about how to perform the task accurately prior to receiving financial motivation to pay attention to negative feedback.

Exploratory analyses of whether any of the psychometrics correlated with each other, including HT and FPRT scores, yielded a significant correlation between EQ and RMET scores. These findings are in support of the argument that the RMET is closer to a measure emotion recognition and appears to capture the latent construct of empathy, as measured by the EQ, more closely than ToM (Pinkham et al., 2016). This correlation has also been observed in another study as well (Voracek & Dressler, 2006).

Limitations

There were several limitations observed in this study. The total sample sized used across analyses was less than what was proposed to observe the desired effects (N=36) and more so considering that individual ERP analyses used smaller subsets of the sample. A major contributor to low power throughout the study was the use of listwise deletion for missing data due to linear regression analysis assumptions requiring a complete-case data set. The reason for missing cases was that there were no clean trials left to analyze for certain conditions after the EEG data was processed. The generalizability of this study's findings is narrow with the sample majority being female (66.6%) and homogenous in age group (18-23). Administration of the neutral condition prior to the incentivized conditions may have introduced practice and order effects.

Considering that the ERN had the one of the smallest sample sizes (N=20) to draw on for analysis due to an insufficient amount of clean error trials, no differences between response types and conditions were observed. Segmentation was also determined to start shortly after the onset of the negativity and to end 100ms later. This segmentation technique could have also contributed to the lack of effects throughout the rest of the ERP analyses as well. Changing segmentation windows, by making them narrower in this case, allows for comparison of the wave forms at their peak activity and where differences are most apparent in amplitude both visually and statistically. Further data collection and a different approach to segmentation may help clarify which relationships are present or not.

The way the cooperative condition was generated may be problematic if the observer saw the situation as a weaker form of competition due to reduced gains in points when tallying for errors. This interpretation would contribute to explaining why there were no differences observed between cooperative and competitive conditions for the oERN. No research exists to date on the impact of reward magnitude on the oERN and the presence of differences between cooperative and competitive conditions are mixed (de Bruijn & von Rhein, 2012; Koban et al., 2010). This study's findings could provide preliminary evidence for the relationship between reward magnitude and the oERN.

Conclusion and Future Directions

The major strength of this study was addressing the relationship between OEM and ToM. Although the design itself was novel, and in large part, exploratory, robust tasks and measures were used with theory-driven predictions. As a result, several hypotheses were supported despite the lack of power and wide range used for segmentation. Most importantly, the primary hypothesis of this study almost reached significance for the HT negatively predicting changes in oERN amplitudes. After observing some participants demonstrate a lack of interest in the monetary aspect of this study's design, one's perception toward money may have influenced their overall engagement toward the interpersonal context generated by having the observer and player compete or cooperate for the reward. Upon including combined household income as a control for attention and motivation post-hoc, the HT had a large significant negative linear correlation with oERN amplitudes (β =-0.68) with a medium effect size ($_{adj}R^2$ = .15). The oFRN was not shown to be predicted by the ToM measures used and follow-up analyses were supportive of the suspicion that it may not be as sensitive to social contexts.

Future studies attempting to continue this line of research should make several improvements to this study's design. A larger sample size should be gathered so that each ERP being assessed has at least 40 participants worth of clean data. A more effective approach would be to use Structural Equation Modeling, which can utilize and incomplete data, to test predictions. Furthermore, broadening the demographics of the sampled population will aid in the generalizability of findings. It would be informative if the current cooperation condition design was included alongside a potentially purer version of the cooperation condition where participants are told if they score a combined accuracy of 95% for the round, they both get the round point and are required to have a minimum number of round points by the end of the study to qualify for winning the reward. Lastly, it would be prudent to address the main hypothesis within clinical samples that demonstrate poor ToM.

References

Abu-Akel, A., & Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia*, *49*(11), 2971-2984.

Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annual review of psychology*, *60*, 693-716.

Ahmed, F. S., & Miller, L. S. (2011). Executive function mechanisms of theory of mind. *Journal of autism and developmental disorders*, *41*(5), 667-678.

Albrecht, C., & Bellebaum, C. (2021). Disentangling effects of expectancy, accuracy, and empathy on the processing of observed actions. *Psychophysiology*, *58*(9), e13883.

Allison, C., Baron-Cohen, S., Wheelwright, S. J., Stone, M. H., & Muncer, S. J. (2011). Psychometric analysis of the Empathy Quotient (EQ). *Personality and Individual Differences*, *51*(7), 829-835.

Apps, M. A., Rushworth, M. F., & Chang, S. W. (2016). The anterior cingulate gyrus and social cognition: tracking the motivation of others. *Neuron*, *90*(4), 692-707.

Bandura, A., & Walters, R. H. (1977). *Social learning theory* (Vol. 1). Prentice Hall: Englewood cliffs.

Bandura, A., Ross, D., & Ross, S. A. (1963). Vicarious reinforcement and imitative learning. *The Journal of abnormal and social psychology*, *67*(6), 601.

Bänziger, T., Scherer, K. R., Hall, J. A., & Rosenthal, R. (2011). Introducing the MiniPONS: a short multichannel version of the profile of nonverbal sensitivity (PONS). *Journal of Nonverbal Behavior*, *35*(3), 189-204.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of autism and developmental disorders*, *34*(2), 163-175.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger Syndrome. *Journal of Child Psychology and Psychiatry, 38*(7), 813-22.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*(1), 37-46.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. *British Journal of developmental psychology*, *4*(2), 113-125.

Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., & Williams, S. C. (1999). Social intelligence in the normal and autistic brain: an fMRI study. *European journal of neuroscience*, *11*(6), 1891-1898.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241-251.

Bartholow, B. D. (2010). On the role of conflict and control in social cognition: Event-related brain potential investigations. *Psychophysiology*, *47*(2), 201-212.

Bates, A. T., Patel, T. P., & Liddle, P. F. (2005). External behavior monitoring mirrors internal behavior monitoring: Error-related negativity for observed errors. *Journal of Psychophysiology*, *19*(4), 281-288.

Bellebaum, C., & Colosio, M. (2014). From feedback-to response-based performance monitoring in active and observation learning. *Journal of cognitive neuroscience*, *26*(9), 2111-2127.

Bellebaum, C., Brodmann, K., & Thoma, P. (2014). Active and observational reward learning in adults with autism spectrum disorder: relationship with empathy in an atypical sample. *Cognitive Neuropsychiatry*, *19*(3), 205-225.

Bellebaum, C., Kobza, S., Thiele, S., & Daum, I. (2010). It was not MY fault: event-related brain potentials in active and observation learning from feedback. *Cerebral Cortex*, *20*(12), 2874-2883.

Brüne, M. (2005). Emotion recognition,'theory of mind,'and social behavior in schizophrenia. *Psychiatry research*, *133*(2-3), 135-147.

Brüne, M., Lissek, S., Fuchs, N., Witthaus, H., Peters, S., Nicolas, V., ... & Tegenthoff, M. (2008). An fMRI study of theory of mind in schizophrenic patients with "passivity" symptoms. *Neuropsychologia*, *46*(7), 1992-2001.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in cognitive sciences*, *4*(6), 215-222.

Calarge, C., Andreasen, N. C., & O'Leary, D. S. (2003). Visualizing how one brain understands another: a PET study of theory of mind. *American Journal of Psychiatry*, *160*(11), 1954-1964.

Carp, J., Halenar, M. J., Quandt, L. C., Sklar, A., & Compton, R. J. (2009). Perceived similarity and neural mirroring: evidence from vicarious error processing. *Social neuroscience*, *4*(1), 85-96.

Clawson, A., Clayson, P. E., Worsham, W., Johnston, O., South, M., & Larson, M. J. (2014). How about watching others? Observation of error-related feedback by others in autism spectrum disorders. *International Journal of Psychophysiology*, *92*(1), 26-34.

Corcoran, R., Mercer, G., & Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: investigating "theory of mind" in people with schizophrenia. *Schizophrenia research*, *17*(1), 5-13.

Davies, M., & Stone, T. (1995). Mental simulation: evaluations and applications-reading in mind and language.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113.

de Bruijn, E. R., de Lange, F. P., von Cramon, D. Y., & Ullsperger, M. (2009). When errors are rewarding. *Journal of Neuroscience*, *29*(39), 12183-12186.

de Bruijn, E. R., & von Rhein, D. T. (2012). Is your error my concern? An event-related potential study on own and observed error detection in cooperation and competition. *Frontiers in neuroscience*, *6*, 8.

de la Asuncion, J., Docx, L., Morrens, M., Sabbe, B., & de Bruijn, E. R. (2015). Neurophysiological evidence for diminished monitoring of own, but intact monitoring of other's errors in schizophrenia. *Psychiatry research*, *230*(2), 220-226.

Decety, J., Chaminade, T., Grèzes, J., & Meltzoff, A. N. (2002). A PET exploration of the neural mechanisms involved in reciprocal imitation. *Neuroimage*, *15*(1), 265-272.

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological science*, *5*(5), 303-305.

Dennett, D. C. (1978). Beliefs about beliefs [p&w, sr&b]. *Behavioral and Brain sciences*, *1*(4), 568-570.

Dvash, J., & Shamay-Tsoory, S. G. (2014). Theory of mind and empathy as multidimensional constructs: Neurological foundations. *Topics in Language Disorders*, *34*(4), 282-295.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics*, *16*(1), 143-149.

Fernandes, J. M., Cajão, R., Lopes, R., Jerónimo, R., & Barahona-Corrêa, J. B. (2018). Social cognition in schizophrenia and autism spectrum disorders: a systematic review and meta-analysis of direct comparisons. *Frontiers in psychiatry*, *9*, 504.

Frank, M. J., Woroch, B. S., & Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. *Neuron*, *47*(4), 495-501.

Frith, U., & Frith, C. (2001). The biological basis of social interaction. *Current directions in psychological science*, *10*(5), 151-155.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 459-473.

Fukushima, H., & Hiraki, K. (2009). Whose loss is it? Human electrophysiological correlates of non-self reward processing. *Social neuroscience*, *4*(3), 261-275.

Gächter, S., Johnson, E. J., & Herrmann, A. (2022). Individual-level loss aversion in riskless and risky choices. *Theory and Decision*, *92*(3-4), 599-624.

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *Neuroimage*, *16*(3), 814-821.

Gallant, C., & Good, D. (2020). Examining the "reading the mind in the eyes test" as an assessment of subtle differences in affective theory of mind after concussion. *The Clinical Neuropsychologist*, *34*(2), 296-317.

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, *295*(5563), 2279-2282.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (2018). The Error-Related Negativity. *Perspectives on Psychological Science*, *13*(2), 200–204.

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological science*, *4*(6), 385-390.

Gehring, W. J., Liu, Y., Orr, J. M., & Carp, J. (2012). The error-related negativity (ERN/Ne). In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 231–291). Oxford University Press.

Gregory, C. , Lough, S., Stone, V.E., Erzinclioglu, S., Martin, L., Baron-Cohen, S. & Hodges, J. (2002). Theory of mind in frontotemporal dementia and Alzheimer's disease: Theoretical and practical implications. Brain, 125, 752-64.

Guastella, A. J., Ward, P. B., Hickie, I. B., Shahrestani, S., Hodge, M. A. R., Scott, E. M., & Langdon, R. (2015). A single dose of oxytocin nasal spray improves higher-order social cognition in schizophrenia. *Schizophrenia research*, *168*(3), 628-633.

Gur, R. C., & Gur, R. E. (2016). Social cognition as an RDoC domain. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *171*(1), 132-141.

Healey, K. M., Combs, D. R., Gibson, C. M., Keefe, R. S., Roberts, D. L., & Penn, D. L. (2015). Observable social cognition–a rating scale: an interview-based assessment for schizophrenia. *Cognitive neuropsychiatry*, *20*(3), 198-221.

Heyes, C., & Catmur, C. (2021). What happened to mirror neurons?. *Perspectives On Psychological Science*.

Hinterbuchinger, B., Kaltenboeck, A., Baumgartner, J. S., Mossaheb, N., & Friedrich, F. (2018). Do patients with different psychiatric disorders show altered social decision-making? A systematic review of ultimatum game experiments in clinical populations. *Cognitive neuropsychiatry*, *23*(3), 117-141.

Hirao, K., Miyata, J., Fujiwara, H., Yamada, M., Namiki, C., Shimizu, M., ... & Murai, T. (2008). Theory of mind and frontal lobe pathology in schizophrenia: a voxel-based morphometry study. *Schizophrenia Research*, *105*(1-3), 165-174.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *Neuroreport*, *14*(18), 2481-2484.

Horat, S. K., Favre, G., Prévot, A., Ventura, J., Herrmann, F. R., Gothuey, I., ... & Missonnier, P. (2018). Impaired social cognition in schizophrenia during the Ultimatum Game: an EEG study. *Schizophrenia research*, *192*, 308-316.

Horat, S. K., Prévot, A., Richiardi, J., Herrmann, F. R., Favre, G., Merlo, M. C., & Missonnier, P. (2017). Differences in social decision-making between proposers and responders during the ultimatum game: an eeg study. *Frontiers in integrative neuroscience*, *11*, 13.

Itagaki, S., & Katayama, J. I. (2008). Self-relevant criteria determine the evaluation of outcomes induced by others. *Neuroreport*, *19*(3), 383-387.

Jeon, D., Kim, S., Chetana, M., Jo, D., Ruley, H. E., Lin, S. Y., ... & Shin, H. S. (2010). Observational fear learning involves affective pain system and Ca v 1.2 Ca 2+ channels in ACC. *Nature neuroscience*, *13*(4), 482-488.

Kang, S. K., Hirsh, J. B., & Chasteen, A. L. (2010). Your mistakes are mine: self-other overlap predicts neural response to observed errors. *Journal of Experimental Social Psychology*, *46*(1), 229-232.

Koban, L., & Pourtois, G. (2014). Brain systems underlying the affective and social monitoring of actions: an integrative review. *Neuroscience & Biobehavioral Reviews*, *46*, 71-84.

Koban, L., Pourtois, G., Bediou, B., & Vuilleumier, P. (2012). Effects of social context and predictive relevance on action outcome monitoring. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(3), 460-478.

Koban, L., Pourtois, G., Vocat, R., & Vuilleumier, P. (2010). When your errors make me lose or win: event-related potentials to observed errors of cooperators and competitors. *Social Neuroscience*, *5*(4), 360-374.

Kobza, S., Thoma, P., Daum, I., & Bellebaum, C. (2011). The feedback-related negativity is modulated by feedback probability in observational learning. *Behavioural brain research*, *225*(2), 396-404.

Koelewijn, T., van Schie, H. T., Bekkering, H., Oostenveld, R., & Jensen, O. (2008). Motor-cortical beta oscillations are modulated by correctness of observed action. *Neuroimage*, *40*(2), 767-775.

Kokmotou, K., Cook, S., Xie, Y., Wright, H., Soto, V., Fallon, N., ... & Stancak, A. (2017). Effects of loss aversion on neural responses to loss outcomes: An event-related potential study. *Biological Psychology*, *126*, 30-40.

Kolling, N., Wittmann, M. K., Behrens, T. E., Boorman, E. D., Mars, R. B., & Rushworth, M. F. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nature neuroscience*, *19*(10), 1280-1285.

Langdon, R., & Coltheart, M. (1999). Mentalising, schizotypy, and schizophrenia. *Cognition*, *71*(1), 43-71.

Langdon, R., Coltheart, M., & Ward, P. (2006). Empathetic perspective-taking is impaired in schizophrenia: evidence from a study of emotion attribution and theory of mind. *Cognitive Neuropsychiatry*, *11*(2), 133-155.

Langdon, R., Coltheart, M., Ward, P. B., & Catts, S. V. (2001). Mentalising, executive planning and disengagement in schizophrenia. *Cognitive Neuropsychiatry*, *6*(2), 81-108.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: reliability and validity of the Empathy Quotient. *Psychological medicine*, *34*(5), 911-920.

Leslie, A. M. (1987). Pretense and representation: The origins of" theory of mind.". *Psychological review*, *94*(4), 412.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in cognitive sciences*, *8*(12), 528-533.

Lindgren, M., Torniainen-Holm, M., Heiskanen, I., Voutilainen, G., Pulkkinen, U., Mehtälä, T., ... & Therman, S. (2018). Theory of mind in a first-episode psychosis population using the Hinting Task. *Psychiatry research*, *263*, 185-192.

Lissek, S., Peters, S., Fuchs, N., Witthaus, H., Nicolas, V., Tegenthoff, M., ... & Brüne, M. (2008). Cooperation and deception recruit different subsets of the theory-of-mind network. *PloS one*, *3*(4), e2023.

Lockwood, P. L., Apps, M. A., Roiser, J. P., & Viding, E. (2015). Encoding of vicarious reward prediction in anterior cingulate cortex and relationship with trait empathy. *Journal of neuroscience*, *35*(40), 13720-13727.

Lorist, M. M., Boksem, M. A., & Ridderinkhof, K. R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cognitive Brain Research*, *24*(2), 199-205.

Ma, Q., Jin, J., Meng, L., & Shen, Q. (2014). The dark side of monetary incentive: how does extrinsic reward crowd out intrinsic motivation. *Neuroreport*, *25*(3), 194-198.

Magno, E., Foxe, J. J., Molholm, S., Robertson, I. H., & Garavan, H. (2006). The anterior cingulate and error avoidance. *Journal of Neuroscience*, *26*(18), 4769-4773.

Mallawaarachchi, S. R., Cotton, S. M., Anderson, J., Killackey, E., & Allott, K. A. (2019). Exploring the use of the hinting taskin first-episode psychosis. *Cognitive neuropsychiatry*, *24*(1), 65-79.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology: Human perception and performance*, *17*(4), 1057.

Mann-Wrobel, M. C. (2009). *Error observation in schizophrenia*. University of Maryland, College Park.

Marco-Pallarés, J., Krämer, U. M., Strehl, S., Schröder, A., & Münte, T. F. (2010). When decisions of others matter to me: an electrophysiological analysis. *BMC neuroscience*, *11*(1), 1-8.

Marjoram, D., Gardner, C., Burns, J., Miller, P., Lawrie, S., & Johnstone, E. (2005). Symptomatology and social inference: a theory of mind study of schizophrenia and psychotic affective disorder. *Cognitive Neuropsychiatry*, *10*(5), 347-359.

Mathewson, K. J., Dywan, J., & Segalowitz, S. J. (2005). Brain bases of error-related ERPs as influenced by age and task. *Biological psychology*, *70*(2), 88-104.

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the national academy of sciences*, *98*(20), 11832-11835.

Mier, D., Lis, S., Neuthe, K., Sauer, C., Esslinger, C., Gallhofer, B., & Kirsch, P. (2010). The involvement of emotion recognition in affective theory of mind. *Psychophysiology*, *47*(6), 1028-1039.

Miltner, W. H., Brauer, J., Hecht, H., Trippe, R., & Coles, M. G. (2004). Parallel brain activity for self-generated and observed errors.

Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a "generic" neural system for error detection. *Journal of cognitive neuroscience*, *9*(6), 788-798.

Mir, P., Trender-Gerhard, I., Edwards, M. J., Schneider, S. A., Bhatia, K. P., & Jahanshahi, M. (2011). Motivation and movement: the effect of monetary incentive on performance speed. *Experimental brain research*, *209*, 551-559.

Morrison, K. E., Pinkham, A. E., Kelsven, S., Ludwig, K., Penn, D. L., & Sasson, N. J. (2019). Psychometric evaluation of social cognitive measures for adults with autism. *Autism Research*, *12*(5), 766-778.

Müller, S. V., Möller, J., Rodriguez-Fornells, A., & Münte, T. F. (2005). Brain potentials related to self-generated and external information used for performance monitoring. *Clinical Neurophysiology*, *116*(1), 63-74.

Negrão, J., Akiba, H. T., Lederman, V. R. G., & Dias, Á. M. (2016). Faux Pas Test in schizophrenic patients. *Jornal Brasileiro de Psiquiatria*, *65*, 17-21.

Ochsner, K. N. (2008). The social-emotional processing stream: five core constructs and their translational potential for schizophrenia and beyond. *Biological psychiatry*, *64*(1), 48-61.

Penn, D. L., Corrigan, P. W., Bentall, R. P., Racenstein, J., & Newman, L. (1997). Social cognition in schizophrenia. *Psychological bulletin*, *121*(1), 114.

Penn, D. L., Corrigan, P. W., Martin, J., Ihnen, G., Racenstein, J. M., Nelson, D., ... & Hope, D. A. (1999). Social cognition and social skills in schizophrenia: the role of self-monitoring. *The Journal of nervous and mental disease*, *187*(3), 188-190.

Pinkham, A. E., Penn, D. L., Green, M. F., & Harvey, P. D. (2016). Social cognition psychometric evaluation: Results of the initial psychometric study. *Schizophrenia bulletin*, *42*(2), 494-504.

Pinkham, A. E., Penn, D. L., Green, M. F., Buck, B., Healey, K., & Harvey, P. D. (2014). The social cognition psychometric evaluation study: results of the expert survey and RAND panel. *Schizophrenia bulletin*, *40*(4), 813-823.

Pizzagalli, D. A., Peccoralo, L. A., Davidson, R. J., & Cohen, J. D. (2006). Resting anterior cingulate activity and abnormal responses to errors in subjects with elevated depressive symptoms: A 128-channel EEG study. *Human brain mapping*, *27*(3), 185-201.

Pollock, B. J., & Lee, T. D. (1992). Effects of the model's skill level on observational motor learning. *Research quarterly for exercise and sport*, *63*(1), 25-29.

Potts, G. F. (2011). Impact of reward and punishment motivation on behavior monitoring as indexed by the error-related negativity. *International Journal of Psychophysiology*, *81*(3), 324-331.

Potts, G. F., Martin, L. E., Kamp, S. M., & Donchin, E. (2011). Neural response to action and reward prediction errors: Comparing the error-related negativity to behavioral errors and the feedback-related negativity to reward prediction violations. *Psychophysiology*, *48*(2), 218-228.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, *1*(4), 515-526.

Quandt, L. (2007). *I Know How You Feel: The Effect of Similarity and Empathy on Neural Mirroring* (Doctoral dissertation).

Renault, B., Ragot, R., & Lesèvre, N. (1980). Correct and incorrect responses in a choice reaction time task and the endogenous components of the evoked potential. In *Progress in brain research* (Vol. 54, pp. 647-654). Elsevier.

Romney, C. E., Larson, M., Sandberg, J., Steffen, P. R., & Baldwin, S. (2015). Marital Satisfaction, Error-observation, and the Brain: Harmful or Beneficial Effects of Spouse Observation?.

Santesso, D. L., Drmic, I. E., Jetha, M. K., Bryson, S. E., Goldberg, J. O., Hall, G. B., ... & Schmidt, L. A. (2011). An event-related source localization study of response monitoring and social impairments in autism spectrum disorder. *Psychophysiology*, *48*(2), 241-251.

Scheffers, M. K., & Coles, M. G. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 141–151.

Schlaffke, L., Lissek, S., Lenz, M., Juckel, G., Schultz, T., Tegenthoff, M., ... & Brüne, M. (2015). Shared and nonshared neural networks of cognitive and affective theory-of-mind: A neuroimaging study using cartoon picture stories. *Human brain mapping*, *36*(1), 29-39.

Shamay-Tsoory, S. G., Tomer, R., Berger, B. D., & Aharon-Peretz, J. (2003). Characterization of empathy deficits following prefrontal brain damage: the role of the right ventromedial prefrontal cortex. *Journal of cognitive neuroscience*, *15*(3), 324-337.

Shane, M. S., Stevens, M. C., Harenski, C. L., & Kiehl, K. A. (2009). Double dissociation between perspective-taking and empathic-concern as predictors of hemodynamic response to another's mistakes. *Social cognitive and affective neuroscience*, *4*(2), 111-118.

Shane, M. S., Stevens, M., Harenski, C. L., & Kiehl, K. A. (2008). Neural correlates of the processing of another’s mistakes: a possible underpinning for social and observational learning. *Neuroimage*, *42*(1), 450-459.

Silberstein, J. M., Pinkham, A. E., Penn, D. L., & Harvey, P. D. (2018). Self-assessment of social cognitive ability in schizophrenia: Association with social cognitive test performance, informant assessments of social cognitive ability, and everyday outcomes. *Schizophrenia research*, *199*, 75-82.

Soder, H. E., & Potts, G. F. (2018). Medial frontal cortex response to unexpected motivationally salient outcomes. *International Journal of Psychophysiology*, *132*, 268-276.

Spaulding, S. (2013). Mirror neurons and social cognition. *Mind & Language*, *28*(2), 233-257.

Stietz, J., Jauk, E., Krach, S., & Kanske, P. (2019). Dissociating empathy from perspective-taking: Evidence from intra-and inter-individual differences research. *Frontiers in Psychiatry*, *10*, 126.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of cognitive neuroscience*, *10*(5), 640-656.

Tamnes, C. K., Walhovd, K. B., Torstveit, M., Sells, V. T., & Fjell, A. M. (2013). Performance monitoring in children and adolescents: a review of developmental changes in the error-related negativity and brain maturation. *Developmental cognitive neuroscience*, *6*, 1-13.

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard university press.

van Schie, H. T., Mars, R. B., Coles, M. G., & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature neuroscience*, *7*(5), 549-554.

Voegler, Rolf, Jutta Peterburs, Hannah Lemke, Sebastian Ocklenburg, Roman Liepelt, and Thomas Straube. "Electrophysiological correlates of performance monitoring under social observation in patients with social anxiety disorder and healthy controls." *Biological psychology* 132 (2018): 71-80.

Voegler, Rolf, Jutta Peterburs, Christian Bellebaum, and Thomas Straube. "Modulation of feedback processing by social context in social anxiety disorder (SAD)–an event-related potentials (ERPs) study." *Scientific reports* 9, no. 1 (2019): 4795.

Voracek, M., & Dressler, S. G. (2006). Lack of correlation between digit ratio (2D: 4D) and Baron-Cohen's "Reading the Mind in the Eyes" test, empathy, systemising, and autism-spectrum quotients in a general population sample. *Personality and Individual Differences*, *41*(8), 1481-1491.

Washburn, D., Wilson, G., Roes, M., Rnic, K., & Harkness, K. L. (2016). Theory of mind in social anxiety disorder, depression, and comorbid conditions. *Journal of anxiety disorders*, *37*, 71-77.

Yamada, Y., Inagawa, T., Sueyoshi, K., Sugawara, N., Ueda, N., Omachi, Y., ... & Sumiyoshi, T. (2019). Social cognition deficits as a target of early intervention for psychoses: a systematic review. *Frontiers in psychiatry*, *10*, 333.

Yildirim, E. A., Kasar, M., Güdük, M., Ateş, E., Küçükparlak, I., & ÖZALMETE, E. O. (2011). Investigation of the reliability of the" reading the mind in the eyes test" in a Turkish population. *Turkish Journal of Psychiatry*, *22*(3).

Yu, R., & Zhou, X. (2006). Brain responses to outcomes of one's own and other's performance in a gambling task. *Neuroreport*, *17*(16), 1747-1751.

ZHU, C. Y., Lee, T. M., LI, X. S., JING, S. C., WANG, Y. G., & Wang, K. (2007). Impairments of social cues recognition and social functioning in Chinese people with schizophrenia. *Psychiatry and clinical neurosciences*, *61*(2), 149-158.

Appendix A: Hinting Task

Hinting Task

Lisa is about to leave the house when her father's car pulls up in the driveway. When he enters she says to him: ``I really need to go shopping, but it's so far away and the rain is terrible".

Question: What does Lisa really mean when she says this?

Hint: Lisa goes on to say: ``It's only five minutes in the car".

Question: What does Lisa want her father to do?

Alan is watching television, and his wife Jill sits down to join him. She says: ``I see you're watching the football. Isn't there anything else on at the moment?"

Question: What does Jill really mean when she says this?

Hint: Jill then says to Alan: ``I thought there was a good play on the other channel".

Question: What does Jill want Alan to do?

Sarah is spending the morning with her next-door neighbor Caroline, having coffee. They are talking about Sarah's forthcoming holiday abroad, when Sarah says to Caroline: ``I'm worried that all my plants will be dead by the time I get back".

Question: What does Caroline really mean when she says this?

Hint: Caroline then says to Sarah: ``I have a spare key for the front door".

Question: What does Caroline want Sarah to do?

Jack and his father are talking about the recent form of the local football team, which they both support. Jack says: ``You know United are playing at home to their big rivals this weekend. I'm sure it will be very exciting."

Question: What does Jack really mean when he says this to his father?

Hint: Jack goes on to say: ``I have never been to watch a football match".

Question: What does Jack want his father to do?

Jim and his brother Richard are getting ready for work in the morning. Jim goes to the bathroom and finds that Richard is about to use the shower, and says to him: ``I've got an early start today and I'm running late".

Question: What does Jim really mean when he says this to Richard?

Hint: Jim goes on to say to Richard: ``It won't take me long to get ready".

Question: What does Jim want Richard to do?

Harry and Chris work together in the same office. One day Harry says to Chris: ``I would really like an extra-long lunch break today, as I have to go to the bank. Will you be going out for lunch today?"

Question: What does Harry really mean when he says this?

Hint: Harry then says to Chris: ``Do you think our boss would mind if only one of us were here?"

Question: What does Harry want Chris to do?

On a weekday evening, Martin goes to see his friend Lucy at home. He is trying to persuade her to go out for a meal, but she says: ``I'm really busy writing a report tonight. I don't even have time to chat".

Question: What does Lucy really mean when she says this?

Hint: Lucy then says: ``I really have to be getting on with my work, is there someone else you could ask".

Question: What does Lucy want Martin to do?

Two children, Emma and Katie are playing, when Emily breaks an old statue belonging to Katie's mother. Emma says to Katie: ``If your Mum finds out it was me that broke it, I won't be allowed to come here anymore".

Question: What does Emma really mean when she says this?

Hint: Emma then says to Katie: ``She wouldn't punish you though".

Question: What does Emma want Katie to do?

Tony and his girlfriend Alison are giving a dinner party at their new flat. They are going through the list of guests when Alison exclaims: ``Oh! It says here you've invited your ex-girlfriend. Is that right?"

Question: What does Alison really mean when she says this?

Hint: Alison goes on to say to Tony: ``I don't get on with her very well".

Question: What does Alison want Tony to do?

Simon is enjoying an evening out at the pub with his friend Gareth. Gareth is about to buy some more drinks when Simon says: ``I have a very busy day tomorrow, and I need to be at my best".

Question: What does Simon really mean when he says this?

Hint: Simon then says to Gareth: ``We have already had quite a lot to drink".

Question: What does Simon want do?

Appropriate responses are credited with 2 points (maximum) and the next scene is presented. If the participant fails to respond accurately (e.g., paraphrasing what was asked and with no obvious use of inferential skills), a rather obvious hint is read to the participant and the second question of what the characters wants to do is asked. This time, if the response is appropriate, a maximum of 1 point can be credited, but if the response is inappropriate, a score of 0 is given and the next scene is presented. Modified task created by Marjoram et al., 2005*

Appendix B: IRB Approval Letter

April 12, 2022

Dear Kipras Varkala:

On 4/11/2022, the IRB reviewed and approved the following protocol:

Application Type:	Initial Study
IRB ID:	STUDY003999
Review Type:	Expedited 4, 7
Title:	Observed Error Monitoring as an Index of Theory of Mind.
Approved Protocol and Consent(s):	• OEMT_Protocol_04.08.22.docx; • 3999 AdultSigned Version 4 4.11.22.pdf; Approved study documents can be found under the 'Documents' tab in the main study workspace. Use the stamped consent found under the 'Last Finalized' column under the 'Documents' tab.

Within 30 days of the anniversary date of study approval, confirm your research is ongoing by clicking Confirm Ongoing Research in BullsIRB, or if your research is complete, submit a study closure request in BullsIRB by clicking Create Modification/CR.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

Jennifer

Walker
IRB Research Compliance Administrator