CAUSAL ROLES OF THE PREFRONTAL CORTEX AND TEMPORO-PARIETAL JUNCTION IN SOCIAL DECISION MAKING

Patricia Christian

Graduate School of
Systemic Neurosciences

LMU Munich

Dissertation at the
Graduate School of Systemic Neurosciences
Ludwig-Maximilians-Universität München

Munich, 31st March 2023

Supervisor: Dr. Alexander Soutschek
Department of General and Experimental Psychology
Faculty of Psychology and Educational Sciences
Ludwig–Maximilians–Universität München

Second Supervisor: Prof. Dr. Paul Sauseng
Research Unit Biological Psychology
Faculty of Psychology and Educational Sciences
Ludwig–Maximilians–Universität München

Third Supervisor: Prof. Paul Taylor
Department of General and Experimental Psychology
Faculty of Psychology and Educational Sciences
Ludwig–Maximilians–Universität München

Fourth Supervisor: Prof. Dr. Simone Schütz-Bosbach
Department of General and Experimental Psychology
Faculty of Psychology and Educational Sciences
Ludwig–Maximilians–Universität München

First Reviewer: Dr. Alexander Soutschek
Second Reviewer: Prof. Dr. Paul Sauseng

Date of Submission: 31st March 2023

Date of PhD Defense: 27th June 2023

Summary

How much we as humans are able to interact with others by following social norms is one of our crucial abilities to build up relationships, a good reputation and become a part of society. This might be one of the main reasons why humans show strong fairness preferences to avoid inequity between themselves and others and punish norm violators when someone deviates from social expectancies. Therefore, the ability to properly guide our own decision making in the social context is crucial for social functioning, which can be affected in psychiatric disorders. Thus, it is highly relevant to understand the precise psychological mechanisms which drive social decision making as well as the underlying brain regions which are causally involved to promote our ability to guide our own decision making in social interactions. In this thesis, I investigated which brain regions are crucial for implementation of one's own choices in the social context as well as adaption of our own social strategical behaviour in response to the other's actions. Furthermore, I examined how our social cognitive abilities such as inference of the others perspective can explain our tendency for fairness preferences in social interactions.

The first research project addressed the crucial role of the right dorso-lateral pefrontal cortex (rDLPFC) in norm enforcement. It has been controverseley debated whether the rDLPFC promotes norm-guided behaviour or implements selfish choices in the social context to maximise one's own monetary payoff. By calculating a meta-analysis of previous studies assessing the rDLPFC's function with transcranial magnetic stimulation (TMS) on social decision making we were able to demonstrate that the rDLPFC's role in social decision making crucially depends on the social context.

The second research project investigated the role of the right temporo-parietal junction (rTPJ) and the right lateral prefrontal cortex (rLPFC) in pro-social fairness. It still remained unknown whether these brain regions were associated with either advantageous or

disadvantageous inequity aversion. By using noninvasive transcranial alternating current stimulation (tACS) we provided direct evidence that rTPJ and rLPFC show dissociable roles for moderating aversion to advantageous and disadvantageous inequity. Further, our results demonstrated that the rTPJ's role for perspective taking strengthen the aversion to unequal splits.

For the third research project we used electroencephalography (EEG) and transcranial magnetic stimulation (TMS) to assess whether the dmPFC is functionally relevant to update our beliefs in response to violated social expectancies in cooperatitive - competitive contexts. While the dmPFC had been linked with mental model representations to build up beliefs about the others strategy, it's precise role in social strategical behaviour still remained unknown. Our results reveal that the dmPFC shows an early neurobiological response when the other choose a stronger competitive strategy then expected. Moreover, our results provide direct evidence that the dmPFC is crucially relevant for updating our beliefs and adapt our behavioural responses when the other unexpectedly defected.

Taken together, this thesis provides causal evidence of the neurocognitive mechanisms which drive social decision making in social interactions. The present findings expand our understanding of the precise neuro-psychological determinants which implement our ability to act and flexibly adapt our own strategy in the social contex to promote goal-directed behaviour. Based on our findings we can gain a better understanding how these brain regions might be affected in psychiatric disorders, which might result in clinical implications for alternative therapeutic interventions such as brain stimulation. Further, the findings give us great insight how our fairness preferences and social cognitive abilities guide our decision to either act more or less pro-social depending on the social context.

Table of Contents

1. General Introduction

1.1. Fundamentals of Social Decision Making

In social interactions humans are required to guide their own decision making based on the trade off between social preferences and one's own self-interests (Fehr & Fischbacher, 2003). Our ability to make decisions in everyday social interactions, which will affect ourselves and our social environment, might be one of the most fundamental abilities to build up relationships and become a part of human society. In complex social interactions such as social decision making we recuit distinct psychological mechanisms such as higher-order social cognition to monitor the others behaviour with the purpose to guide our own choices (Frith & Singer, 2008). Finding the optimal choice in the social context is often challenging because it involves a conflict how much we are willing to override our own selfish interests to follow social norms (Fehr & Fischbacher, 2003). Moreover, our choices are often driven not only by our own pro-social tendencies but depend on the others fairness preferences as well why we need to represent our own goals and strategies in social interaction while at the same time flexibly adapt to the others fairness behaviour (Rilling & Sanfey, 2011).

The psychological and neuro-cognitive mechanisms underlying normative behaviour in social decision making have been commonly studied with economic games (Camerer, 2003). Economic games are widely used to assess social preferences for fairness norms, in which participants are instructed to decide how to divide an amount of money between themselves and others in varying social contexts (Fehr & Camerer, 2007; Sanfey, 2007). Even though there is evidence of external validity of experimental findings on social decision-making it has been controversally debated whether these findings inside of the laboratory are truly relevant outside of the laboratory (Franzen & Pointner, 2012; Laury & Taylor, 2008; Levitt & List, 2007). More recent findings combining physiological recordings and social decision making provide further

evidence that physiological responses in the laboratory corresponds to those outside of the laboratory associated with fairness preferences (Fooken, 2017).

Even though game theory models predict humans to behave rational and selfish by maximising their own payoff, a broad research literature demonstrates that humans show strong fairness preferences by restricting one's own selfish motives in economic games (Camerer, 2003; Fehr & Camerer, 2007; Fehr & Fischbacher, 2003; Fehr & Fischbacher, 2004; Rilling & Sanfey, 2011). Previous findings suggest that conformity with social norms might be deeply rooted in humans: humans and non-human primates show stronger fairness preferences and pro-social behaviour among others, as f.e. norm enforcement in response to violation of social expectancies (Brosnan, 2013; Burkart, Brügger, & van Schaik, 2018). Further, from a developmental perspective previous findings reveal that children already show fairness preferences at an early stage of their development, as f.e. preferences for an equal split between themselves and other children (Guroglu, van den Bos, & Crone, 2014; McAuliffe, Blake, Steinbeis, & Warneken, 2017; McAuliffe, Jordan, & Warneken, 2015). Nevertheless, how much humans are willing to override their own selfish interest depends on the type of social interaction and the social context, as f.e. whether we are confronted to make choices for the punishment of out-group or ingroup members (Bernhard, Fischbacher, & Fehr, 2006; Rahal, Fiedler, & De Dreu, 2020). Thus, our tendency to act more or less pro-social depends on contextual social factors.

1.2. Psychological mechanisms and higher-oder social cognition underlying social decision making

Social decision making is a complex human behaviour which is driven by our own fairness preferences and selfish interests as well as our social expectancies towards others. That's why

it is crucial to understand the precise psychological mechanisms underlying social decision making as well as the socio-cognitive abilities which enable us to act flexibly in social environments.

It has been controversially debated which psychological mechanisms drive our willingness to behave fairly towards others, even when threat of punishment is absent. Proactive fairness behaviour can be understood as the willingness to fairly distribute resources when norm violations in response to unfair behaviour cannot be punished (Hallsson, Siebner, & Hulme, 2018). Previous findings on proactive fairness demonstrate that humans show preferences to be treated equally by avoiding unfair splits between themselves and others (Fehr & Schmidt, 1999). Interestingly, humans not only try to avoid being worse off than others (disadvantageous inequity) but prefer to receive an equal split, even when they are better off than others (advantageous inequity) (Charness & Rabin, 2002; Fehr & Schmidt, 1999). Previous evidence reveals that humans show a stronger aversion to disadvantageous in contrast to advantageous inequity (Fehr & Schmidt, 1999; Loewenstein, Bazerman, & Thompson, 1989). Nevertheless, humans prefer an equal split even when they need to sacrifice their own monetary benefits to establish an equal treatment in accordance with social norms (advantageous inequity). It has been proposed that aversion to advantageous inequity might reflect fairness concerns to promote long term goals such as sustained cooperation and building up one's own reputation (Dawes, Fowler, Johnson, McElreath, & Smirnov, 2007). Further, it has been hypothesized that the psychological mechanisms underlying this tendency to reject advantageous inequity might rely on perspective taking promoting pro-social choices (Imuta, Henry, Slaughter, Selcuk, & Ruffman, 2016; Underwood & Moore, 1982). Indeed, it has been proposed that the ability for understanding the others intentions and goals as well as to differentiate between our own and the others mental state shows a strong impact on one's own fairness preferences and social behaviour (Frith & Frith, 2012; Van Overwalle & Baetens, 2009). Investigations of the

psychological mechanisms underlying disadvantageous inequity show that the aversion to receive less than others is associated with strong negative affect such as anger or envy which may need to be downregulated to overcome disadvantageous inequity (McAuliffe et al., 2017). Thus, previous evidence suggest that advantageous and disadvantageous inequity might be implemented by dissociable psychological mechanisms which could be reflected by dissociable neuro-cognitive mechanisms underlying different brain areas.

Moreover, our social behaviour is driven by norm enforcement to maintain fairness behaviour in society: humans tend to punish norm deviant behaviour and show stronger normative behaviour themselves when experiencing the threat of punishment (Fehr & Fischbacher, 2004; Strang et al., 2015). Even when the option to punish deviant behaviour is costly or result in negative consequences for themselves, humans are willing to enforce social norms (Baumgartner, Knoch, Hotz, Eisenegger, & Fehr, 2011). Similarly, non-human primates show the tendency to punish deviant behaviour even when this is costly, providing evidence for a strong biological predisposition for punishment when others violate one's own social expectancies (Leimgruber, Rosati, & Santos, 2016). Interestingly, humans are willing to punish norm violators even when they are not directly affected by the negative consequences, i.e. Third Party Punishment (Fehr & Fischbacher, 2004). The tendency to punish norm violators might be rooted in the long term benefits of reinstatement of justice and maintained cooperation in society as well as direct or indirect reciprocity towards the punisher (Boyd, Gintis, Bowles, & Richerson, 2003). Indeed, punishment shows long-term benefits such as increasing one's own status or reputation, be more likely to be choosen as a partner for social exchange and to be perceived as more trustworthy (Barclay, 2006; Jordan, Hoffman, Bloom, & Rand, 2016). Theoretical accounts propose that the motivation to punish deviant behaviour could be driven by negative affect such as anger or envy towards norm violators (Gilam, Abend, Shani, Ben-Zion, & Hendler, 2019; Harth & Regner, 2017; Reuben & van Winden, 2008). In fact, previous

findings combining physiological data (f.e. heart rate) with behavioural performance demonstrate that participants experience real physiological stress and emotional arousal in response to unfair proposals which is associated with higher rejection rates (Dulleck, Schaffner, & Torgler, 2014; Dunn, Evans, Makarova, White, & Clark, 2012). Nevertheless, humans vary in their response behaviour to either accept or reject unfair proposals depending on the social contextual factors (Bechler, Green, & Myerson, 2015). Further, it is still an open debate which precise neuro-cognitive mechanism drive decision making in response to unfair monetary allocations. Previous accounts propose that based on our cognitive control abilities humans are able to override selfish tendencies to maximise one's own payoff thus promoting higher punishment in response to unfair proposals (Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006). More recent accounts suggest that the implementation of our decision to either reject or accept unfair proposals might be based on strategical thinking trading off external signals such as social norms with internal goals and intentions to guide our decision making in response to perceived unfairness (Buckholtz et al., 2015). Hence, our higher-order cognitive abilities might promote more flexible behavioural responses to proposed monetary allocations rather than simply rejecting or accepting these proposed offers per se.

Further, there is considerable evidence that our tendency to establish social cooperation with others is influenced by enforcing social norms through sanctions why humans show stronger tendencies for social cooperation when facing the option to be punished (Fehr & Gächter, 2000). Social cooperation is one of the most complex forms of human social behaviour when individuals provide resources to others while expecting to receive something equivalent in the long term (Axelrod & Hamilton, 1981; Rilling et al., 2002). One of the main preconditions which promote the willingness to cooperate is a prolonged social interaction with the same social partner to build up an alliance from which both might benefit in the long term (Axelrod & Hamilton, 1981). Further, cooperative strategies are implemented more strongly when our

own willingness to cooperate is in fact reciprocated by the other player in repetitive social interactions (King-Casas et al., 2005; Rand & Nowak, 2013). Previous findings demonstrate that the evolution of direct and indirect forms of reciprocity support our tendency to be cooperative which encourages the persistence to follow social norms (Nowak & Sigmund, 2005; Santos, Santos, & Pacheco, 2018). Hence, humans have a stronger preference to select more cooperative individuals as long-term social interaction partners why being cooperative is crucially relevant to build up a good reputation and social network (De Cremer & Barker, 2003; Feinberg, Willer, & Schultz, 2014). For building up social cooperation both social partners depend on the others choices which requires us to infer the others mental state and make predictions about the others future actions to guide decision making in the social context (Brown & Brüne, 2012; Hampton, Bossaerts, & O'Doherty, 2008). Thus we build up mental model representations about the others intentions and behavioural strategy to adapt our own choices in accordance with the others current strategy (Stallen & Sanfey, 2013). Hence, one of the major challenges in sustained cooperation is understanding the others' fairness preferences to discriminate between those who are willing to return the favor or those who deviate from our own fairness norms (Emonds, Declerck, Boone, Vandervliet, & Parizel, 2012). In fact, humans show a strong preference for cooperation when they belief that the other is willing to choose cooperative over defective strategies (Fehr & Fischbacher, 2003; Rilling et al., 2002). Further, the others tendency to choose cooperative over competitive strategies depend on our behavioural strategy as well. Thus, it is essential for us to represent our own and the others fairness preferences to guide our own decision making (Hill et al., 2017). That's why our own willingness to choose a cooperative strategy might be promoted by second - order beliefs about the others social expectancies (Chang, Smith, Dufwenberg, & Sanfey, 2011). Conclusively, in complex social interactions, in which there is a trade off between cooperation and competition, humans are required to build up and update beliefs about the others intentions and behavioural strategies while at the same time representing the others fairness expectations to guide their

own decision making. However, the precise neuro-cognitive mechanisms underlying belief updating still remain unknown.

1.3. Neuro-cognitive mechanisms of Social Decision Making

Based on previous findings we can assume that socio-cognitive abilities such as taking the others perspective or prediction of the others behavioural strategy enable us to implement goal-directed decision making in the social context (Frith & Singer, 2008). While we already have a basic understanding which psychological mechanisms drive decision making in the social context it is still debated which underlying neuro-cognitive mechanisms reflect these pro-social tendencies and social strategical choices. Depending on the type of social interactions we recruit distinct higher-order cognitive functions to implement social decision making which suggest that dissociable brain regions might be involved to implement decision making depending on the contextual factors of our social environment.

Recent neuroimaging findings suggest that social decision making is a manifold human behaviour which is associated with a broad neural network, recruiting the prefrontal cortex, temporo-parietal cortex, anterior insula and the amygdala (Gangopadhyay, Chawla, Dal Monte, & Chang, 2021; Lee, 2008; Luo, 2018; Sanfey, 2007). Previous evidence reveals that prefrontal cortex and temporo-parietal cortex are recruited in social decision making when we are being confronted with real human agents evoking social conflicts between our own and the others interests (Knoch et al., 2006; Rilling & Sanfey, 2011; Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004). Interestingly, these brain regions are associated with a domain general role of basic neuro-psychological processes across different types of social behaviour (Fehr & Camerer, 2007; Suzuki & O'Doherty, 2020). Hence, these neural circuits might be commonly recruited for social cognition and social decision making depending on the contextual factors (Feng et al., 2021; Hare, Camerer, Knoepfle, & Rangel, 2010; Ruff & Fehr, 2014).

It has been proposed that the right temporo-parietal junction (rTPJ) is one of the brain regions which play a key role in promoting pro-social choices and implementing fairness behaviour towards others (Obeso, Moisa, Ruff, & Dreher, 2018; Rilling, King-Casas, & Sanfey, 2008). More specifically, the rTPJ is involved when we promote proactive fairness towards others (Morishima, Schunk, Bruhin, Ruff, & Fehr, 2012) and has been linked with social preferences for promoting inequity aversion (Hutcherson, Bushong, & Rangel, 2015). Furthermore, previous findings demonstrate that the rTPJ is recruited when we demand socio-cognitive abilities to guide our behaviour in the social context, such as mentalizing, perspective taking and self-other distinction (Frith & Frith, 2006, 2012; Gallagher & Frith, 2003; Saxe & Kanwisher, 2003; Van Overwalle & Baetens, 2009). Based on previous evidence that the rTPJ is recruited across social contexts, it has been hypothesized that the neuro-computational role of the rTPJ for promoting pro-social behaviour might be explained by it's key role in promoting social cognition (Decety & Lamm, 2007; Soutschek, Ruff, Strombach, Kalenscher, & Tobler, 2016; Strombach et al., 2015). In line with this theoretical account past research demonstrates that the same brain regions which are involved in mentalizing are recruited for decision making in the social context, when we make more generous choices (Cutler & Campbell-Meiklejohn, 2019; Young, Cushman, Hauser, & Saxe, 2007). Indeed, past research suggest that social preferences for inequity aversion might be implemented by the rTPJ's role for perspective taking (McAuliffe et al., 2017). Nevertheless, it still remains unknown whether the rTPJ is causally involved to promote either advantageous inequity or inequity aversion per se and whether the preference to avoid unequal splits in the proactive fairness context can be explained by the rTPJ's role for perspective taking.

Previous neuroimaging findings contributed to a greater understanding of the neural mechanisms underlying punishment in social interactions implying that the rDLPFC is one of the key areas involved in norm-based decision making (Buckholtz et al., 2008; Sanfey, Rilling,

Aronson, Nystrom, & Cohen, 2003; Spitzer, Fischbacher, Herrnberger, Grön, & Fehr, 2007; Strobel et al., 2011). Direct evidence from brain stimulation studies suggest that the rDLPFC enhances rejection rates of unfair proposals ascribing this brain region a crucial role for punishment of norm violators (Baumgartner et al., 2011; Knoch et al., 2006). Moreover, previous findings demonstrate that disruption of the rDLPFC reduced punishment rates without impairing the participants' ability to judge these offers as unfair which suggest that the rDLPFC is mainly involved to implement decision making irrespective of the perceived unfairness of these offers (Buckholtz et al., 2015; Knoch et al., 2006). Nevertheless, contrary findings show that perturbation of the rDLPFC enhances norm-guided tendencies which suggest that the rDLPFC promotes selfish choices to increase one's own monetary payoff rather than implement norm enforcement (Brune et al., 2012; Christov-Moore, Sugiyama, Grigaityte, & Iacoboni, 2017; Maier et al., 2018). Thus, it has been controversely debated whether the rDLPFC implements fairness behaviour across social contexts or whether it is involved to either promote pro-social or selfish behaviour. Moreover, it has been debated which precise neuro-cognitive mechanisms underly the rDLPFC's involvement in norm-based decision making. While previous accounts suggest that the rDLPFC inhibits prepotent selfish tendencies based on cognitive control mechanisms to promote norm enforcement (higher punishment rates) (Knoch et al., 2006), more recent accounts suggest that the rDLPFC is recruited to integrate internal intentions and motives as well as external signals (f.e. social norm) to flexibly implement decision making promoting goal maintenance (Buckholtz & Marois, 2012; Buckholtz et al., 2015). This is in line with the domain general role of the rDLPFC for higher order cognition to promote goal directed decision making by representing internal goals and how to achieve them (Friedman & Miyake, 2017; Miller & Cohen, 2001; Passingham & Sakai, 2004). Hence, it is still debated which precise function can be ascribed to the rDLPFC in social decision making.

The medial prefrontal cortex (mPFC) has been shown to be strongly linked with pro-social behaviour, when making donations for charity or rejecting unfair proposals (Baumgartner et al., 2011; Hare et al., 2010; Tricomi, Rangel, Camerer, & O'Doherty, 2010). Further, the mPFC is recruited in competitive social interactions when the other unexpectedly choose defective over cooperative strategies (Hertz et al., 2017). This is in line with previous research which suggests that the mPFC is sensitive to unexpected outcomes when simulating the other's actions (Dungan, Stepanovic, & Young, 2016; Lee & Seo, 2016) and to build up representations of one's own and the others mental states (Nicolle et al., 2012; Zhu, Mathewson, & Hsu, 2012). Thus, the mPFC might be functionally relevant to implement and update representations about the other's current thoughts and goals to guide our own decision making in complex social interaction such as social cooperation. Indeed, past research propose that the mPFC encodes social value signals which increase our ability to flexibly adapt our own choices in response to changes in our social environment (Yoshida, Saito, Iriki, & Isoda, 2011) and is linked with updating of mental model representations (Haroush & Williams, 2015; Nicolle et al., 2012). Nevertheless, it is still unclear whether the mPFC is causally relevant to promote prediction changes about the others behavioural strategy in response to unexpected outcomes. Thus, it's crucially relevant to gain a better understanding of the neuro-cognitive mechanisms and neurobiological basis in response to unexpected defection in social cooperation.

1.4. Advances of brain stimulation methods for causal inference in neuroscientific research

Previous findings from neuroimaging research propose that the prefrontal cortex and temporo-parietal cortex play key roles in implementing decision making in the social context (Lee & Seo, 2016), however, their precise neuro-computational roles are still debated. Based on limited correlational inference of neuroimaging research it still remains unknown which brain regions are causally involved to promote pro-social tendencies in social interactions.

Thus, it is crucial to gain a greater understanding of the precise function of these brain regions in social decision making and how they implement our choice behaviour depending on the social context.

Neuroimaging methods offer great insights into the neural mechanisms underlying social decision making by identifing a broad network of brain regions linked with normative behaviour including the prefrontal cortex and temporo-parietal cortex (Gabay, Radua, Kempton, & Mehta, 2014; Yang, Zheng, Yang, Li, & Liu, 2019). However, understanding the neuronal signature of decision making in the social context with the assessment of neuroimaging methods is limited to correlational inference. Whenever brain regions show an increased activation pattern simultaneously to behavioural task performance this has often been interpreted as an involvement of these brain regions to implement social decision-making. Nevertheless, based on the fact that we can observe changes in brain activity patterns while at the same time performing a task does not necessarily mean that this brain area is crucially relevant for task performance. Thus, based on methodological constrains of neuroimaging methods we are not able to make any causal inferences about brain regions associated with social decision making (Marini, Banaji, & Pascual-Leone, 2018; Polania, Nitsche, & Ruff, 2018).

Therefore the application of neuroscientific methods such as non-invasive brain stimulation (NIBS), as f.e. transcranial magnetic stimulation (TMS), transcranial direct-current stimulation (tDCS) or transcranial alternating current stimulation (tACS), should be exploited more strongly to establish causal links between the recruitment of specific brain regions and choice behaviour (Polania et al., 2018). NIBS methods can be applied to either enhance or interfere with ongoing neural processes in a specified target area depending on the stimulation protocol (Bolognini & Ro, 2010; Polania et al., 2018; Veniero, Strüber, Thut, & Herrmann, 2019). Thus, when we apply TMS protocols such as repetitive transcranial magnetic stimulation

(rTMS) or continuous theta burst stimulation (cTBS) we can interfere with ongoing neural activity in the targeted brain region (Hobot, Klincewicz, Sandberg, & Wierzchoń, 2020; Polania et al., 2018) while participants perform a social decision making task, enabling us to modulate changes in pro-social behaviour (Izuma et al., 2015; Soutschek, Sauter, & Schubert, 2015). Further, when we apply tACS it is possible to entrain brain rythms such as theta oscillations in the targeted brain region to examine the neuronal signature underlying social decision making (Polania et al., 2018). Thus, depending on the stimulation protocol, NIBS provides direct evidence that specific brain regions are causally relevant for socio-cognitive processes underlying the behavioural performance in economic games (Marini et al., 2018).

Although we already have a basic understanding why we as humans show strong fairness preferences, it is still debated which precise neural mechanisms are involved in the trade-off between selfish interests and fairness concerns. Further, it still remains unknown which brain regions are causally involved to implement social decision making depending on the social context. Thus, we can extend our understanding of causally involved brain regions in social decision making by applying neurostimulation methods to identify the neural basis of socio-normative behaviour and it's underlying neuro-psychological mechanisms.

1.5. Aim of the Thesis

Our ability to implement decision making in the social context determines how much we are able to follow social norms while at the same time maintaining our own personal interests. Indeed, it is crucial for extending our social network and build up close relationships which in the long term affect our mental health (Beeney, Hallquist, Clifton, Lazarus, & Pilkonis, 2018). Humans who are suffering from neurological and psychiatric disorders show deviations in social behaviour, which might reflect socio-cognitive deficits underlying decision making in the social context (Padmanabhan, Lynch, Schaer, & Menon, 2017; Schneider et al., 2013).

Thus, to extend our understanding of the precise neuro-cognitive mechanisms in social strategical behaviour can shed new light on our understanding how social decision making might be affected in clinical populations.

The concrete goal of this dissertation is to investigate the brain regions which are causally involved in implementing decision making in the social context, depending on the precise contextual circumstances. Our purpose is to extend our understanding of the crucial role of prefrontal cortex and temporo-parietal cortex to resolve conflicts between our fairness tendencies and selfish interests to promote pro-social behaviour towards others as well as to implement social decision making with or without the threat of social punishment. Additionally, we examine the precise neuro-cognitive mechanisms underlying social decision making to extend our understanding of socio-cognitive processes which might promote fairness preferences or represent beliefs about others when we make decisions in the social context.

Based on inconsistent evidence it is still debated whether the rDLPFC is crucially relevant to either promote norm-guided behavior or implement selfish choices (Baumgartner et al., 2011; Brune et al., 2012; Knoch et al., 2006; Maier et al., 2018; Muller-Leinss, Enzi, Flasbeck, & Brune, 2018; Strang et al., 2015). Brain stimulation studies are often based on relatively small sample sizes which may over- or underestimate the true effect size why a meta-analysis provides a more accurate approach to determine TMS effects on rDLPFC to implement decision making in the social context. Therefore, to determine the causal role of the rDLPFC for social norm enforcement we analysed data from previous TMS studies across social decision making paradigms (dictator game, ultimatum game, trust game, third party punishment game and prisoner's dilemma game) in a meta-analysis and subgroup analysis.

Even though previous findings show that the right temporo-parietal junction (rTPJ) and the right lateral prefrontal cortex (rLPFC) play key roles in pro-social choices (Cutler & Campbell-

Meiklejohn, 2019; Gao et al., 2018; Hutcherson et al., 2015), it remains unknown whether these regions show dissociable roles for advantageous or disadvantageous inequity. Further, it remains unknown whether this is reflected by the same or dissociable brain rythms underlying social decision making in the proactive fairness context. Thus, we applied non-invasive brain stimulation methods (tACS) to examine the causal role of the rLPFC as well as the rTPJ for inequity aversion in the dictator game. Additionally, to further understand the precise neuro-psychological mechanisms which drive advantageous or disadvantageous inequity aversion we analyzed performance in an additional perspective taking task (director task).

While previous findings suggest that the dorsomedial prefrontal cortex (dmPFC) plays a key role in representing the other's mental states and update beliefs about the other's behavioural strategy when our prediction about the other mismatches (Haroush & Williams, 2015; Nicolle et al., 2012), it's precise functional role in social strategical behaviour is still debated. We used electroencephalography (EEG) and transcranial magnetic stimulation (TMS) to examine the role of the medial prefrontal cortex in response to unexpected defection in the prisoner's dilemma game. In the EEG experiment, we tested whether an early medial frontal ERP component, Medial-Frontal Negativity (MFN), reflects negative prediction errors, when the co-player in the prisoner's dilemma game unexpectedly defected. Further, for the same project, by applying cTBS over the dmPFC we examined the crucial role of the dmPFC to implement belief updating and behavioural adaption in response to unexpected defection.

2. Chapter I: Causal role of right dorsolateral prefrontal cortex for norm-guided social decision making: a meta-analysis of TMS studies

This article was published on november, 5th, 2022:

Christian, P., & Soutschek, A. (2022). Causal role of right dorsolateral prefrontal cortex for norm-guided social decision making: A meta-analysis of TMS studies. Neuropsychologia, 176, 108393. https://doi.org/10.1016/j.neuropsychologia.2022.108393

PC and AS designed research; PC performed research; PC analyzed data; PC and AS wrote first draft of manuscript, all authors approved manuscript

Neuropsychologia 176 (2022) 108393

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Causal role of right dorsolateral prefrontal cortex for norm-guided social decision making: A meta-analysis of TMS studies

Patricia Christian [a,b,*], Alexander Soutschek [a,b]

[a] Department of Psychology, Ludwig Maximilians University Munich, Munich, Germany
[b] Graduate School of Systemic Neurosciences, Ludwig Maximilians University Munich, Munich, Germany

ARTICLE INFO

Keywords:
Transcranial magnet stimulation
Social decision making
Social norms
Fairness
Dorsolateral prefrontal cortex
Prosocial giving
Prosocial punishment

ABSTRACT

Theoretical accounts ascribe the right dorsolateral prefrontal cortex (rDLPFC) a crucial role in social decision making, but previous studies assessing the rDLPFC's function with transcranial magnetic stimulation (TMS) provided inconsistent evidence. While some studies suggest that the rDLPFC promotes norm-guided behavior, others report the rDLPFC to implement selfish choices. To decide between these conflicting accounts, we conducted a meta-analysis of studies that investigated the impact of rDLPFC TMS on social decision making. While we observed no significant effect of rDLPFC TMS across all studies, moderator analyses revealed that the rDLPFC's role in social decision making crucially depends on the social context: in particular, we found that rDLPFC promotes norm-guided behavior predominantly when decision makers have to trade-off their interaction partners' intentions and fairness expectations against their selfish interests (reactive fairness). In contrast, there was no evidence that rDLPFC TMS affects prosocial giving (proactive fairness). Our results thus inform theoretical accounts by showing that brain stimulation over rDLPFC does not increase or decrease norm-guided behavior per se; instead, contextual factors determine the role of the rDLPFC in social interactions.

1. Introduction

Social norms are based on widely shared beliefs about what is considered as appropriate behavior in social interactions (Cialdini and Goldstein, 2004; Fehr and Fischbacher, 2004a). Because violations of fairness expectations are often retaliated with social exclusion or punishment (Spitzer et al., 2007), fairness norms strongly influence social interactions and require decision makers to trade-off self-related interests against fairness considerations (Rilling and Sanfey, 2011).

The right dorsolateral prefrontal cortex (rDLPFC) is hypothesized to play a crucial role in implementing fairness-oriented behavior (Buckholtz et al., 2015; Lee and Harris, 2013; Rilling and Sanfey, 2011) and in resolving conflicts between fairness norms and selfish interests (Buckholtz and Marois, 2012). However, the rDLPFC's precise function in social decisions remains a matter of controversial debate. One line of research assumes that rDLPFC promotes norm-guided behavior, which includes both the punishment of unfair others (norm enforcement) and the decision maker's compliance with social norms (norm compliance), e.g. by costly sharing money with others in order to reduce inequity (Buckholtz et al., 2008; Feng et al., 2015; Montague and Lohrenz, 2007; Spitzer et al., 2007). In contrast, other accounts link rDLPFC activation to strengthening selfish interests over fairness norms when facing conflicts between outcome maximization and compliance to social norms (Emonds et al., 2011; Fermin et al., 2016; Sanfey et al., 2003).

Brain stimulation studies testing the causal contribution of rDLPFC in social decision making with transcranial magnetic stimulation (TMS) provided mixed evidence for these conflicting views: While some studies reported disruptive rDLPFC TMS to increase selfish behavior and thus to lower the weight assigned to fairness norms (Baumgartner et al., 2011; Knoch et al., 2006; Knoch et al., 2009; Müller-Leinß et al., 2017; Soutschek et al., 2015; Strang et al., 2015; van't Wout et al., 2005), others suggest that rDLPFC perturbation strengthens norm-guided over selfish choices (Brüne et al., 2012; Christov-Moore et al., 2017; Maier et al., 2018). To resolve this controversy and to clarify the role of the rDLPFC in social decision making, we conducted a meta-analysis on the available evidence from TMS studies. To determine whether the rDLPFC promotes selfish or norm-guided behavior in social interactions, we combined TMS studies that examined the role of rDLPFC for social preferences in paradigms involving economic conflicts between one's own and others' payoff, including the dictator game (DG), the ultimatum game (UG), third-party punishment game (TPPG), trust game (TG), and prisoner's dilemma game (PDG). All of these paradigms involve conflicts between

* Corresponding author. Department of Psychology, Ludwig Maximilians, University Munich, Leopoldstr. 13, 80802, Munich, Germany.
E-mail address: patricia.christian@psy.lmu.de (P. Christian).

https://doi.org/10.1016/j.neuropsychologia.2022.108393
Received 4 February 2022; Received in revised form 4 October 2022; Accepted 4 October 2022
Available online 11 October 2022

compliance to fairness norms and selfish interests (Fehr and Camerer, 2007; Sanfey, 2007). In contrast, we excluded studies using scenario-based moral dilemma paradigms (Buckholtz et al., 2015; Jeurissen et al., 2014) which entailed no conflict between social norms and economic self-interests. While many TMS studies are based on relatively small sample sizes and effect sizes from single studies may over- or underestimate the true effect size (Borenstein et al., 2011), a meta-analysis provides a more precise estimate of TMS effects on normative choices and may allow resolving the controversy on the impact of rDLPFC TMS on social decision making.

One potential reason for the inconsistent findings in the literature is that rDLPFC may not have one unitary role in norm-guided behavior, but that its function depends on the given social context. This notion is supported by findings showing that the rLPFC promotes costly giving in a dictator game if proposers can be punished for unfair offers, whereas rLPFC activation reduces voluntary transfers if no punishment for low transfers is possible (Ruff et al., 2013). This suggests that the rDLPFC's role for social decision making may depend on whether a choice involves proactive or reactive fairness considerations. Proactive fairness considerations influence prosocial giving in situations where the receiver has no opportunity to react to the decision maker's choice (Hallsson et al., 2018). In contrast, in reactive fairness contexts decision makers react to their interaction partners' norm violations and fairness expectations, for example when they can be punished for unfair behavior. Proactive versus reactive fairness considerations are also closely linked to the decision maker's role (i.e., proposer versus responder) in social interactions. In fact, previous studies revealed that lowering the excitability of the rDLPFC reduces responders' acceptance rates of unfair offers in the Ultimatum Game, whereas rDLPFC stimulation had no effect on proposers' offers (Speitel et al., 2019), hinting to a specific role of rDLPFC for responder behavior. This suggests a potential moderating effect of the decision maker's role on the rDLPFC's function for fairness considerations. Note that the variables "fairness type" and "role of the decision maker" are closely linked but nevertheless represent distinct categories which measure for dissociable aspects of social interactions. The role of the decision maker is defined via participants' position in experimental economic games (and thus represents a characteristic of the task paradigm), whereas fairness type represents a theoretical construct indicating whether or not fairness considerations are influenced by the reactions or intentions of others. Testing whether the effects of rDLPFC TMS depend on fairness type (proactive versus reactive) or the decision maker's role (proposer versus receiver) allowed us to determine which of these variables moderate the influence of rDLPFC TMS on social interactions.

2. Methods

We conducted a meta-analysis to determine the impact of rDLPC TMS on norm-guided behavior as well as the contextual factors moderating the effects of rDLPFC TMS. The meta-analysis was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009).

2.1. Literature search

A literature search was carried out using PSYCInfo, PubMed, and Web of Science using the following search terms: ("DLPFC "or "dorsolateral prefrontal cortex" or "prefrontal cortex") AND ("TMS" or "transcranial magnetic stimulaton" or "brain stimulation") AND ("social norms" or "fairness" or "pro-social behavior" or "altruism" or "cooperation" or "punishment" or "norm violation") OR ("social decision making" or "moral decisions" or "Dictator Game" or "Ultimatum Game "or "Third-Party Punishment "or "Trust Game "or "Prisoner's Dilemma") before April 27, 2020. Additional papers were identified by examining the citation indices and reference sections of the articles. If insufficient information about statistical results reported in an article prevented the calculation of effect sizes, we asked the authors for clarification or raw data.

2.2. Study selection criteria

A total of 623 records were identified in the initial search and 13 additional studies were included via reference and citation search (n = 636), which were reduced to 578 studies after the removal of duplicates. Then, we preselected relevant articles by screening the titles and abstracts of all remaining records for social decision making paradigms (n = 578), which resulted in a preselection of n = 68 articles (Moher et al., 2009). In the next step, eligibility assessment of the preselected studies (n = 68) was performed by full text analysis in a standardized manner by two authors to avoid the possibility of rejecting relevant reports and biases in article selection (Liberati et al., 2009). The inter-rater agreement of the selection process was high and disagreements were resolved through further discussion.

We selected studies which (i) included healthy young adults as participants, (ii) applied TMS over right or left DLPFC, (iii) reported data from participants acting as responders or proposers, (iv) reported data from either the dictator game (DG), the ultimatum game (UG), third-party punishment game (TPP), trust game (TG), or prisoner's dilemma game (PDG). We excluded articles from the meta-analysis if they met one of the following exclusion criteria: (i) the article type represented a review, meta-analysis, or commentary, (ii) the study tested clinical populations, (iii) reported neuroimaging results, (iv) applied transcranial electrical stimulation instead of TMS, or (v) applied TMS over brain regions other than DLPFC (Fig. 1). As a result of this selection procedure, 10 studies were included in the meta-analysis (Table 1).

2.3. Statistical analysis

As meta-analyses estimate the magnitude of effects in the population by combining effect sizes from single studies, we first calculated Cohen's d for each study based on the statistical tests reported in the included papers (Lipsey and Wilson, 2001). Based on these individual effect sizes, Hedges'g (pooled estimate of standardized mean difference) was computed as measure of the mean effect size, together with 95% confidence intervals (CIs) and p-values for the random-effects model. Because Strang et al. (2015) reported TMS effects on offers in a dictator game with (reactive fairness) and without (proactive fairness) punishment option, we computed separate effect sizes for proactive and reactive fairness for this study. Moreover, in the studies of Maier et al. (2018) and Müller-Leinß et al. (2017) participants played a dictator game against opponents who had previously shown either fair or unfair behavior in an ultimatum game. For these two studies, we computed an effect size only for the condition involving previously fair others, because only this condition entailed a conflict between (proactive) fairness norms and selfish interests. In contrast, the condition with previously unfair others in these studies included no such conflict, because punishing others for unfair behavior also increased participants' selfish payoff. In our meta-analysis, the direction of the calculated effect size indicates whether TMS changed decisions towards more norm-guided or more selfish behavior. Positive effect sizes indicate that inhibitory rDLPFC TMS enhanced norm-guided behavior, while negative effect sizes indicate that rDLPFC disruption resulted in more selfish behavior. We used the meta and metafor package in R (Schwarzer et al., 2015; Viechtbauer, 2010) to calculate a random-effects meta-analysis. We choose a random-effects model to account for potential heterogenity of the selected studies (Field, 2005). Random-effects models increase the generalizability of results by considering both the within-study and between-study variance (Borenstein et al., 2011; Huizenga et al., 2011; Schmidt et al., 2009). We used a Sidik-Jonkman estimator for the random effects model, which is considered to lead to more precise estimates of error terms compared with other estimators (IntHout et al., 2014), particularly in case of large between-study variability (Veroniki

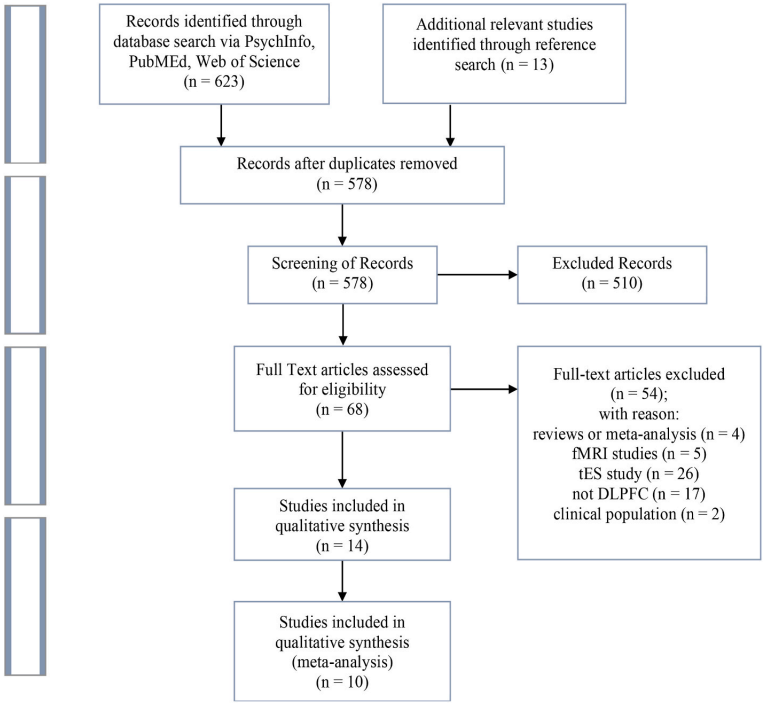

Fig. 1. Flow diagram of each step of the literature search and selection process following the PRISMA guidelines (Moher et al., 2009).

Table 1
Characteristics of included studies (*N = 10*).

Reference	Study Design	N	Stim. Type	Localization strategy	TMS protocol	Active Stim.	Control Stim	Economic Game	Behavioral Outcome (TMS effect)
Strang et al. (2015) (no punishmnent)	within-subject	17	rTMS	neuronavigation	1 Hz, 110% RMT, 900 pulses, 15 min	rDLPFC, lDLPFC	sham	dictator game (without punishment)	less fair offers
Strang et al. (2015) (punishment)	within-subject	17	rTMS	neuronavigation	1 Hz, 110% RMT, 900 pulses, 15 min	rDLPFC, lDLPFC	sham	dictator game (with punishment)	less fair offers
Christov-Moore et al. (2017)	between subject	58	cTBS	neuronavigation	5 Hz, 80% AMT, 600 pulses, 40 s	rDLPFC, DMPFC	control	dictator game (modified)	higher proportion of fair offers to high SES group
Maier et al. (2018)	within-subject	19	cTBS	EEG 10–20 system	5 Hz, 80% AMT, 600 pulses, 40 s	rDLPFC	sham	dictator game (after UG)	higher proportion of fair offers for previously fair players
Müller-Leinß et al. (2017)	between subject	46	rTMS	neuronavigation	1 Hz, 110% RMT, 1200 pulses, 20 min	rDLPFC, lDLPFC	sham	dictator game (after UG)	lower proportion of fair offers for previously fair players
van't Wout et al. (2005)	within-subject	7	rTMS	EEG 10–20 system	1 Hz, 12 min	rDLPFC	sham	ultimatum game	lower acceptance rate of unfair offers
Knoch et al. (2006)	between subject	52	rTMS	neuronavigation	1 Hz, 110% RMT, 900 pulses, 15 min	rDLPFC, lDLPFC	sham	ultimatum game	lower acceptance rate of unfair offers
Baumgartner et al. (2011)	between-subject	32	rTMS	EEG 10–20 system	1 Hz, 110% RMT, 900 pulses, 15 min	rDLPFC	lDLPFC	ultimatum game	lower acceptance rate of unfair offers
Knoch et al. (2009)	between subject	87	rTMS	EEG 10–20 system	1 Hz, 110% RMT, 900 pulses, 15 min	rDLPFC, lDLPFC	sham	trust game (modified)	lower back-transfers in reputation condition
Brüne et al. (2012)	within-subject	20	rTMS	EEG 10–20 system	1 Hz, 110% RMT, 1200 pulses, 20 min	rDLPFC, lDLPFC	sham	third-party punishment	increased costly punishment rate
Soutschek et al. (2015)	between subject	56	rTMS	EEG 10–20 system	1 Hz, 480 pulses, 110 RMT, 8 min	rDLPFC, lDLPFC	sham, control	prisoner's dilemma	decreased cooperation rate

et al., 2016). In addition, we used the modified Knapp-Hartung adjustment (ad hoc correction) which improves the estimation of between-study heterogeneity when relatively few studies are included in the meta-analysis (Knapp and Hartung, 2003; Rover et al., 2015). Because publication bias can distort effect size estimations in meta-analyses, we examined the risk of a potential publication bias with a funnel plot (Sterne et al., 2011) and Egger's regression test (Egger et al., 1997), which quantitatively assesses asymmetry in the data. Finally, heterogeneity was tested using Cochran Q, the I^2 statistics, and τ^2. The Q statistic tests whether heterogeneity between studies is significantly different from zero, whereas τ^2 measures the between-study variance and I^2 indicates the proportion of the variance in effect size estimates that can be explained by study heterogeneity (Borenstein et al., 2011; Higgins and Thompson, 2002).

To test the hypothesis that contextual factors moderate the influence of rDLPFC TMS on social decisions, we calculated subgroup analyses (Borenstein et al., 2011; Borenstein and Higgins, 2013). We defined categorical predictors to test for differences depending on the social context: fairness type (proactive versus reactive) and role of the decision maker (proposer versus responder) (Table 2). Previous literature suggests that also impartial third parties react to observed norm violations by costly punishment of unfair proposers (Fehr and Fischbacher, 2004b; FeldmanHall et al., 2014; Jordan et al., 2016; Krueger and Hoffman, 2016). However, third parties are not directly affected by norm violations (no personal relevance) and different motivations are thought to underlie punishment decisions in third-party versus second-party interactions (Chavez and Bicchieri, 2013; Fehr and Fischbacher, 2004b; Feng et al., 2021; Strobel et al., 2011). We therefore excluded the third-party punishment game (Brüne et al., 2012) from the fairness type subgroup analysis.

3. Results

The meta-analysis included a total of 10 papers (Ultimatum game: $N = 3$; Dictator game: $N = 2$; Dictator game after Ultimatum game: $N = 2$; Third Party Punishment game: $N = 1$; Prisoner's Dilemma game: $N = 1$; Trust game: $N = 1$) with 11 effect sizes from a total of 358 participants. Across all studies, we observed no main effect of rDLPFC stimulation on behavior, $g = -0.37$, 95% CI = [−0.83,0.08], $p = 0.10$, providing no evidence for a causal role of rDLPFC for promoting either selfish or norm-compliant behavior (Fig. 2). A post-hoc power calculation (Jackson and Turner, 2017) revealed that the power of the current meta-analysis was only 30.47%, suggesting that the number of studies in the meta-analysis was not sufficient to reliably detect a main effect of TMS.

Next, we tested for a potential publication bias with a funnel plot (Fig. 3). Visual inspection of the funnel plot indicated no substantial asymmetry. This was further supported by the non-significant result of the Egger's regression test, $\beta 0 = 1.79$, $t = 1.17$, $p = 0.2$. There was thus no evidence for publication bias in the current meta-analysis.

Table 2
Overview over moderator variables ("Fairness type" and "Role of decision maker").

Reference	Fairness Type	Role of Decision Maker
Strang et al. (2015) (no punishment)	proactive	proposer
Strang et al. (2015) (punishment)	reactive	proposer
Christov-Moore et al. (2017)	proactive	proposer
Maier et al. (2018)	proactive	proposer
Müller-Leinß et al. (2017)	proactive	proposer
van't Wout (2005)	reactive	responder
Knoch et al. (2006)	reactive	responder
Baumgartner et al. (2011)	reactive	responder
Knoch et al. (2009)	reactive	responder
Brüne et al. (2012)	NA	NA
Soutschek et al. (2015)	reactive	responder

We observed significant heterogenity among studies, Q = 34.89, $p < 0.01$, $I^2 = 71\%$, $\tau^2 = 0.3210$, which may hint to a potential influence of moderator variables on the rDLPFC's role in social decision making. We therefore calculated subgroup analyses to determine the factors modulating the impact of rDLPFC TMS. We observed a significant moderating effect of fairness type (proactive versus reactive), $k = 10$, $p = 0.007$: While rDLPFC TMS did not alter proactive fairness, $g = 0.12$, 95% CI = [−0.91, 1.15], rDLPFC disruption increased selfishness in studies assessing reactive fairness, $g = -0.86$, 95% CI = [−1.29, −0.43] (Fig. 4). The results of testing for heterogeneity suggest that the proactive fairness subgroup showed significant heterogeneity between studies, $I^2 = 71\%$, $\tau^2 = 0.2984$, $p = 0.02$, whereas in the reactive fairness subgroup heterogeneity measures showed no effect, $I^2 = 0\%$, $\tau^2 = 0.0177$, $p = 0.83$. The results show that rDLPFC studies on reactive fairness consistently show less norm-guided behavior under rDLPFC TMS compared with sham TMS, whereas there were no significant TMS effects on proactive fairness.

We conducted a second subgroup analysis to examine whether the moderating role of fairness type on rDLPFC TMS could alternatively be explained by the role of the decision maker in social bargaining games, given that the variable fairness type is confounded with the decision maker's role (proposer versus receiver). This subgroup analysis revealed a significant moderating effect of the decision maker's role in the game (proposer versus responder), $k = 10$, $p = 0.04$. When participants played in the role of the proposer, rDLPFC TMS showed no significant mean effect, $g = -0.08$, 95% CI = [−0.97, 0.81], whereas rDLPFC TMS increased selfishness when decision makers were in the role of the responder, $g = -0.86$, 95% CI = [−1.38, −0.33] (Fig. 5). Heterogeneity tests revealed significant heterogeneity among effect sizes only in the proposer subgroup, $I^2 = 76\%$, $\tau^2 = 0.3913$, $p < 0.01$, not in the responder subgroup, $I^2 = 0\%$, $\tau^2 = 0.0250$, $p = 0.71$. We note that all of the five studies in which participants played as responder measured reactive fairness, such that the significant TMS effect on responder behavior may not appear surprising given the impact of TMS on reactive fairness. Taken together, these results support the notion rDLPFC TMS promotes selfish behavior in contexts where participants have to respond to unfair others. Finally, we conducted additional subgroup analyses to explore moderating effects of TMS parameters including pattern (1 Hz rTMS versus cTBS), number of pulses, stimulation length, and localization strategy (neuronavigation versus EEG 10–20 system). We found a significant effect of TMS pattern, $p < 0.01$, with 1 Hz rTMS leading to less norm-guided behavior than cTBS. However, the cTBS subgroup included only two studies, both measuring choices in the dictator game, such that this result should be interpreted with caution. No further subgroup analysis showed a significant result, all $p > 0.06$.

4. Discussion

The current meta-analysis suggests a crucial role of rDLPFC for trading-off fairness norms against selfish interests. Interestingly, rather than generally biasing norm-guided or selfish behavior, the moderator analyses show that the social context determines the influence of rDLPFC TMS on social decision making. In reactive fairness contexts, i.e. when decision makers have to consider their interaction partners' norm violations or fairness expectations, rDLPFC enhances punishment of norm violations and sanction-induced norm compliance. In contrast, rDLPFC TMS did not significantly affect trade-offs between self-interests and proactive fairness, i.e. when receivers cannot react to decision makers' norm violations. This suggests that rDLPFC influences social decision making predominantly in reactive, but not proactive, fairness contexts. This interpretation is further supported by the results of the second subgroup analysis suggesting that rDLPFC promotes norm-guided behavior if participants are in the role of the responder and need to react to others' norm violations, but not if they are in the role of the proposer. Because the moderator analyses for both fairness type and decision maker's role yielded significant results, with both variables

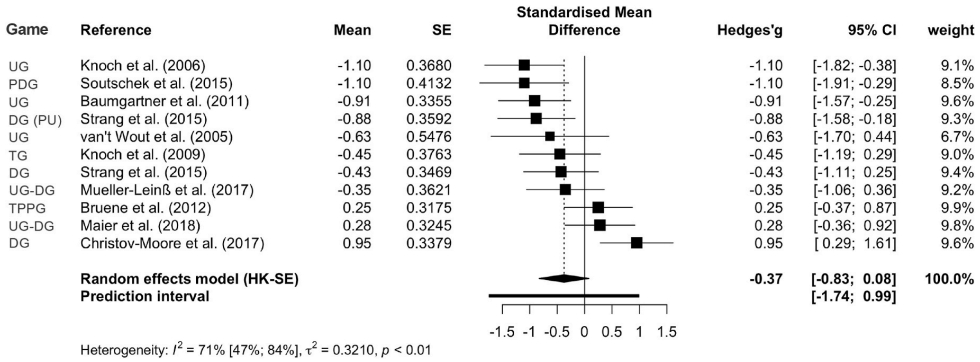

Fig. 2. Effects of TMS on norm-guided choices in social decision making. Forest plot illustrating the results of the meta-analysis including the effect sizes (Mean), standard error of effect sizes (SE) and 95% confidence intervals (CIs) from all studies. The center of the diamond represents the pooled estimate of standardized mean difference (Hedges'g) for TMS effects on social decision making. The following task paradigms were used: dictator game with punishment option (DG (PU)) and without punishment option (DG), dictator game after ultimatum game (UG-DG), ultimatum game (UG), trust game (TG), third party punishment game (TPPG), and prisoner's dilemma game (PDG).

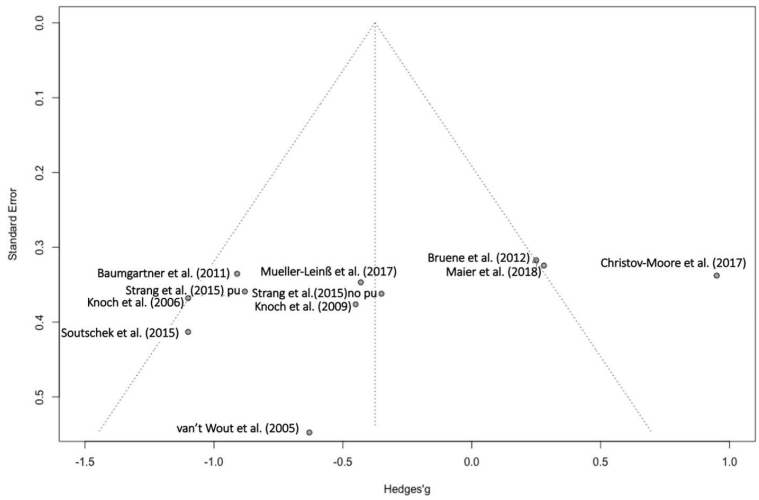

Fig. 3. Funnel plot indicating the variability of effect sizes. Each dot represents an effect size (Hedges'g) as a function of its standard error for each study included in the meta-analysis.

being strongly correlated with each other, the current data do not allow deciding which of these variables is driving the observed effects. In any case, our meta-analysis suggests that the rDLPFC's role in social decision making is highly context-specific, with rDLPFC disruption reducing norm-guided in situations where decision makers have to respond to others' fairness violations in the role of the receiver.

Our results are in line with prevous imaging studies showing that rDLPFC activity correlates with norm violations (Zinchenko and Arsalidou, 2018), lower acceptance rates of unfair offers (Wu et al., 2014), and punishment of norm violators (Buckholtz et al., 2008; Civai et al., 2019; Stallen et al., 2018; Treadway et al., 2014). The finding that rDLPFC promotes norm-guided behavior particularly in reactive fairness contexts when humans have to consider others' fairness intentions could hint to a decisive role of the rDLPFC for integrating beliefs about others' fairness expectations or intentions into the decision process (Güroglu, van den Bos, Rombouts and Crone, 2010; Rilling and Sanfey, 2011). This is supported by previous results showing that rDLPFC is associated with the incorporation of fairness judgments into norm-based decisions (Buckholtz et al., 2015). rDLPFC may thus promote norm-compliant behavior by increasing the weight assigned to fairness norms or potential punishments (Sanfey et al., 2014).

The finding that the rDLPFC affects social decisions when decision makers have to trade-off the goal of maximizing their own selfish payoff against costly punishment of others' norm violations is consistent with the hypothesized role of DLPFC for strategic thinking and higher-order cognition (Friedman and Miyake, 2017; Miller and Cohen, 2001; Smith and Jonides, 1999; Smolker et al., 2015). In contrast, our results do not support the view that rDLPFC biases more "rational", outcome-maximizing decisions by suppressing negative emotional reactions to perceived unfairness (Maier et al., 2018; Sanfey et al., 2003). Rather than top-down suppressing the temptation to act selfishly, rDLPFC might trade-off selfish interests against conflicting fairness norms (Buckholtz, 2015; Buckholtz and Marois, 2012). From this perspective, rDLPFC perturbation impairs the integration of abstract fairness norms into the decision process, thereby increasing the preference for selfish rewards.

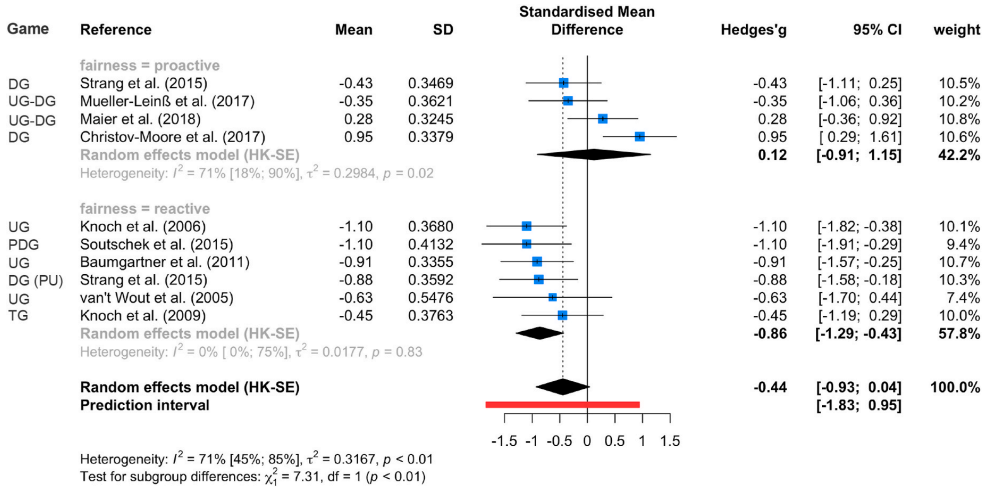

Fig. 4. Effects of TMS on norm-guided choices depending on fairness type (proactive versus reactive). Forest plot illustrating the results of the meta-analysis including effect sizes (Mean), standard error of effect sizes (SE) and 95% confidt ence intervals (CIs) from all studies. The center of the diamond represents the pooled estimate of standardized mean difference (Hedges'g) for TMS effects on social decision making depending on the fairness subgroup. The following task paradigms were used: dictator game with punishment option (DG (PU)) and without punishment option (DG), dictator game after ultimatum game (UG-DG), ultimatum game (UG), trust game (TG) and prisoner's dilemma game (PDG).

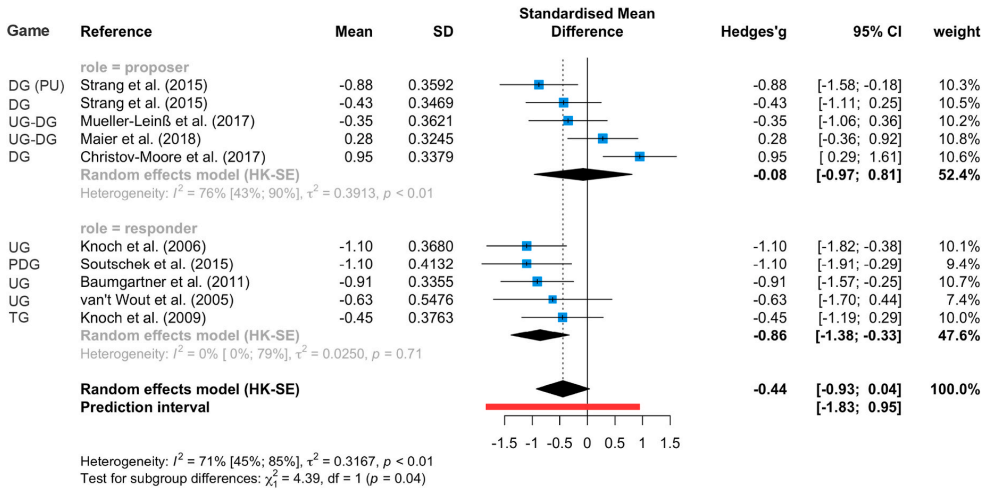

Fig. 5. Effects of TMS on norm-guided choices depending on role of the decision maker (proposer versus responder). Forest plot illustrating the results of the meta-analysis including the effect sizes (Mean), standard error of effect sizes (SE) and 95% confidence intervals (CIs) from all studies. The center of the diamond represents the pooled estimate of standardized mean difference (Hedges'g) for TMS effects on social decision making depending on the decision maker's role. The following task paradigms were used: dictator game with punishment option (DG (PU)) and without punishment option (DG), dictator game after ultimatum game (UG-DG), ultimatum game (UG), trust game (TG) and prisoner's dilemma game (PDG).

Contrary to reactive fairness, conflicts between self-interests and proactive fairness were not significantly affected by rDLPFC TMS. This could be explained by the assumption that the influence of social norms on behavior is weaker when norm violations cannot be punished (Fehr and Fischbacher, 2004a), as evidenced by findings showing that costly sharing in the dictator game (as indicator of proactive fairness) is not affected by the receiver's expectations about how the dictator ought to split the money (Bicchieri and Xiao, 2009). Interestingly, heterogeneity tests indicated that there was significant variability between TMS studies on proactive fairness. It is thus possible that the impact of rDLPFC TMS in proactive fairness contexts depends on further factors that could not be assessed in the current meta-analysis. For example, there is evidence that the rDLPFC's role in prosocial giving might be gender-specific (Rand et al., 2016), with the rDLPFC promoting more selfish behavior in women versus more prosocial behavior in men (Chen et al., 2019). In line with this assumption, evidence suggests that also the role of the dopaminergic reward system in costly giving differs between female and male individuals (Soutschek et al., 2017). Thus, the rDLPFC might influence prosocial giving by inhibiting predominant action impulses encoded by the dopaminergic reward system, consistent with the

rDLPFC's involvement in cognitive control and action inhibition (Friedman and Robbins, 2022; Smith and Jonides, 1999). However, as most studies examined in this meta-analysis did not provide sufficient information to compute separate effect sizes for female and male participants, we could not test this hypothesis in our meta-analysis, such that this assumption remains speculative and will need to be tested by future studies.

A potential limitation of the current study is that our meta-analysis included only a total of 11 effect sizes. Although this is generally considered as sufficient for meta-analyses (Borenstein et al., 2011), the estimated power for the main effect of TMS in the meta-analysis was relatively low. Statistical power was even further reduced for the subgroup analyses where the 11 effect sizes were split up into subgroups, such that only relatively large effects could be detected in the current analyses. It is thus possible that with a higher statistical power we might have observed significant rDLPFC TMS effects on norm enforcement also in the main analysis across all subgroups or in the proactive fairness subgroup. Nevertheless, given the low variability between effect sizes in the reactive fairness and responder subgroups, the results for these subgroups seem likely to be replicated when adding future TMS studies to this meta-analysis.

Another possible limitation hampering the interpretation of the reported results is that the physiological mechanisms underlying TMS are not fully understood. Even though previous literature suggests that cTBS and 1 Hz TMS protocols attenuate cortical excitability of the motor cortex (Fitzgerald et al., 2006; Huang et al., 2005), recent accounts question the assumption of universally inhibitory effects of these TMS protocols (McCalley et al., 2021). Furthermore, there is evidence that cTBS can disrupt or enhance cortical excitability depending on the used stimulation parameter (Gamboa et al., 2010). When interpreting the current results, one should therefore keep in mind that at least some part of the between-study variability might reflect heterogeneous effects of TMS on cortical excitability.

Taken together, our findings inform theoretical models on the rDLPFC's role in social decision making. Existing accounts disagree on whether rDLPFC promotes selfish or prosocial behavior in social interactions, and existing empirical evidence does not clearly favor one alternative over the other. Our meta-analysis provides a solution to this controversy by suggesting that the rDLPFC's role in social decision making is highly context-specific: rDLPFC strengthens norm-guided behavior in reactive fairness contexts where decision makers have to trade-off others fairness expectations aginst their selfish interests. In contrast, we found no conclusive evidence for an impact of rDLPFC TMS on prosocial giving when a decision maker's norm violation could not be sanctionized. The rDLPFC might thus bias either prosocial or selfish behavior depending on the current social context, consistent with the findings of a meta-analysis of neuroimaging studies reporting that both selfish and prosocial decisions correlate with increased DLPFC activation (Cutler and Campbell-Meiklejohn, 2019). We note that in the literature two accounts were proposed to explain the role of the rDLPFC for social decision making. One account posits that rDLPFC implements response inhibition processes that override prepotent selfish interests in order to promote norm-based choices (Knoch et al., 2006; Nash et al., 2013; Steinbeis et al., 2012). Alternatively, the rDLPFC was hypothesized to integrate context-specific information such as blame in order to select a context-appropriate action alternative. Accordingly, rDLPFC disruption might promote selfish behavior by interfering with this integration of social context information (e.g., fairness violations) into the choices process (Buckholtz and Marois, 2012; Buckholtz et al., 2015). While the results of our meta-analysis are in principle consistent with both views, the integration account may appear more plausible given that the response inhibition account provides no reason for why control processes inhibit the temptation to be selfish rather than the emotional reaction to unfairness.

Deficits in social decision making belong to the core symptoms of several psychiatric disorders (Chang et al., 2012; King-Casas and Chiu, 2012), and these deficits are hypothesized to (at least partially) relate to prefrontal dysfunctions (Herpertz et al., 2014; Verdejo-Garcia, Verdejo-Román, Albein-Urios, Martínez-González and Soriano-Mas, 2017). By providing insights into the causal link between rDLPFC activation and social behavior, our findings contribute to improving our understanding of the neural basis of altered social behavior for these psychiatric disorders, which may promote the development of brain-targeting therapeutic interventions.

Credit author statement

Patricia Christian: Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft, Reviewing and Editing. **Alexander Soutschek**: Conceptualization, Methodology, Writing – original draft, Reviewing and Editing, Supervision, Funding acquisition.

Data availability

Data will be made available on request.

Acknowledgements

AS received an Emmy Noether fellowship (SO 1636/2–1) from the German Research Foundation.

References

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., Fehr, E., 2011. Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. Nat. Neurosci. 14 (11), 1468–1474. https://doi.org/10.1038/nn.2933.

Bicchieri, C., Xiao, E., 2009. Do the right thing: but only if others do so. J. Behav. Decis. Making 22 (2), 191–208. https://doi.org/10.1002/bdm.621.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., 2011. Introduction to Meta-Analysis. John Wiley & Sons.

Borenstein, M., Higgins, J.P., 2013. Meta-analysis and subgroups. Prev. Sci. 14 (2), 134–143. https://doi.org/10.1007/s11121-013-0377-7.

Brüne, M., Scheele, D., Heinisch, C., Tas, C., Wischniewski, J., Gunturkun, O., 2012. Empathy moderates the effect of repetitive transcranial magnetic stimulation of the right dorsolateral prefrontal cortex on costly punishment. PLoS One 7 (9), e44747. https://doi.org/10.1371/journal.pone.0044747.

Buckholtz, J.W., 2015. Social norms, self-control, and the value of antisocial behavior. Curr. Opin. Behav. Sci. 3, 122–129. https://doi.org/10.1016/j.cobeha.2015.03.004.

Buckholtz, J.W., Asplund, C.L., Dux, P.E., Zald, D.H., Gore, J.C., Jones, O.D., Marois, R., 2008. The neural correlates of third-party punishment. Neuron 60 (5), 930–940. https://doi.org/10.1016/j.neuron.2008.10.016.

Buckholtz, J.W., Marois, R., 2012. The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. Nat. Neurosci. 15 (5), 655–661. https://doi.org/10.1038/nn.3087.

Buckholtz, J.W., Martin, J.W., Treadway, M.T., Jan, K., Zald, D.H., Jones, O., Marois, R., 2015. From blame to punishment: disrupting prefrontal cortex activity reveals norm enforcement mechanisms. Neuron 87 (6), 1369–1380. https://doi.org/10.1016/j.neuron.2015.08.023.

Chang, S.W., Barack, D.L., Platt, M.L., 2012. Mechanistic classification of neural circuit dysfunctions: insights from neuroeconomics research in animals. Biol. Psychiatr. 72 (2), 101–106. https://doi.org/10.1016/j.biopsych.2012.02.017.

Chavez, A.K., Bicchieri, C., 2013. Third-party sanctioning and compensation behavior: findings from the ultimatum game. J. Econ. Psychol. 39, 268–277. https://doi.org/10.1016/j.joep.2013.09.004.

Chen, W., Zhang, S., Turel, O., Peng, Y., Chen, H., He, Q., 2019. Sex-based differences in right dorsolateral prefrontal cortex roles in fairness norm compliance. Behav. Brain Res. 361, 104–112. https://doi.org/10.1016/j.bbr.2018.12.040.

Christov-Moore, L., Sugiyama, T., Grigaityte, K., Iacoboni, M., 2017. Increasing generosity by disrupting prefrontal cortex. Soc. Neurosci. 12 (2), 174–181. https://doi.org/10.1080/17470919.2016.1154105.

Cialdini, R.B., Goldstein, N.J., 2004. Social influence: compliance and conformity. Annu. Rev. Psychol. 55, 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015.

Civai, C., Huijsmans, I., Sanfey, A.G., 2019. Neurocognitive mechanisms of reactions to second- and third-party justice violations. Sci. Rep. 9 (1), 9271. https://doi.org/10.1038/s41598-019-45725-8.

Cutler, J., Campbell-Meiklejohn, D., 2019. A comparative fMRI meta-analysis of altruistic and strategic decisions to give. Neuroimage 184, 227–241. https://doi.org/10.1016/j.neuroimage.2018.09.009.

Egger, M., Smith, G.D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ 315, 629–634.

Emonds, G., Declerck, C.H., Boone, C., Vandervliet, E.J.M., Parizel, P.M., 2011. Comparing the neural basis of decision making in social dilemmas of people with

different social value orientations, a fMRI study. J. Neurosci. Psychol. Econ. 4 (1), 11–24. https://doi.org/10.1037/a0020151.
Fehr, E., Camerer, C.F., 2007. Social neuroeconomics: the neural circuitry of social preferences. Trends Cognit. Sci. 11 (10), 419–427. https://doi.org/10.1016/j.tics.2007.09.002.
Fehr, E., Fischbacher, U., 2004a. Social norms and human cooperation. Trends Cognit. Sci. 8 (4), 185–190. https://doi.org/10.1016/j.tics.2004.02.007.
Fehr, E., Fischbacher, U., 2004b. Third-party punishment and social norms. Evol. Hum. Behav. 25 (2), 63–87. https://doi.org/10.1016/S1090-5138(04)00005-4.
FeldmanHall, O., Sokol-Hessner, P., Van Bavel, J.J., Phelps, E.A., 2014. Fairness violations elicit greater punishment on behalf of another than for oneself. Nat. Commun. 5, 5306. https://doi.org/10.1038/ncomms6306.
Feng, C., Luo, Y.J., Krueger, F., 2015. Neural signatures of fairness-related normative decision making in the ultimatum game: a coordinate-based meta-analysis. Hum. Brain Mapp. 36 (2), 591–602. https://doi.org/10.1002/hbm.22649.
Feng, C., Yang, Q., Azem, L., Atanasova, K.M., Gu, R., Luo, W., Krueger, F., 2021. An fMRI investigation of the intention-outcome interactions in second- and third-party punishment. Brain Imaging Behaviour 16 (2), 715–727. https://doi.org/10.1007/s11682-021-00555-z.
Fermin, A.S., Sakagami, M., Kiyonari, T., Li, Y., Matsumoto, Y., Yamagishi, T., 2016. Representation of economic preferences in the structure and function of the amygdala and prefrontal cortex. Sci. Rep. 6, 20982 https://doi.org/10.1038/srep20982.
Field, A.P., 2005. Is the meta-analysis of correlation coefficients accurate when population correlations vary? Psychol. Methods 10 (4), 444–467. https://doi.org/10.1037/1082-989X.10.4.444.
Fitzgerald, P.B., Fountain, S., Daskalakis, Z.J., 2006. A comprehensive review of the effects of rTMS on motor cortical excitability and inhibition. Clin. Neurophysiol. 117 (12), 2584–2596. https://doi.org/10.1016/j.clinph.2006.06.712.
Friedman, N.P., Miyake, A., 2017. Unity and diversity of executive functions: individual differences as a window on cognitive structure. Cortex 86, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023.
Friedman, N.P., Robbins, T.W., 2022. The role of prefrontal cortex in cognitive control and executive function. Neuropsychopharmacology 47 (1), 72–89. https://doi.org/10.1038/s41386-021-01132-0.
Gamboa, O.L., Antal, A., Moliadze, V., Paulus, W., 2010. Simply longer is not better: reversal of theta burst after-effect with prolonged stimulation. Exp. Brain Res. 204 (2), 181–187. https://doi.org/10.1007/s00221-010-2293-4.
Güroglu, B., van den Bos, W., Rombouts, S.A., Crone, E.A., 2010. Unfair? It depends: neural correlates of fairness in social context. Soc. Cognit. Affect Neurosci. 5 (4), 414–423. https://doi.org/10.1093/scan/nsq013.
Hallsson, B.G., Siebner, H.R., Hulme, O.J., 2018. Fairness, fast and slow: a review of dual process models of fairness. Neurosci. Biobehav. Rev. 89, 49–60. https://doi.org/10.1016/j.neubiorev.2018.02.016.
Herpertz, S.C., Jeung, H., Mancke, F., Bertsch, K., 2014. Social dysfunctioning and brain in borderline personality disorder. Psychopathology 47 (6), 417–424. https://doi.org/10.1159/000365106.
Higgins, J.P.T., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21, 1539–1558.
Huang, Y.Z., Edwards, M.J., Rounis, E., Bhatia, K.P., Rothwell, J.C., 2005. Theta burst stimulation of the human motor cortex. Neuron 45 (2), 201–206. https://doi.org/10.1016/j.neuron.2004.12.033.
Huizenga, H.M., Visser, I., Dolan, C.V., 2011. Testing overall and moderator effects in random effects meta-regression. Br. J. Math. Stat. Psychol. 64 (Pt 1), 1–19. https://doi.org/10.1348/000711010X522687.
IntHout, J., Ioannidis, J.P., Borm, G.F., 2014. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med. Res. Methodol. 14, 25. https://doi.org/10.1186/1471-2288-14-25.
Jackson, D., Turner, R., 2017. Power analysis for random-effects meta-analysis. Res. Synth. Methods 8 (3), 290–302. https://doi.org/10.1002/jrsm.1240.
Jeurissen, D., Sack, A.T., Roebroeck, A., Russ, B.E., Pascual-Leone, A., 2014. TMS affects moral judgment, showing the role of DLPFC and TPJ in cognitive and emotional processing. Front. Neurosci. 8, 18. https://doi.org/10.3389/fnins.2014.00018.
Jordan, J.J., Hoffman, M., Bloom, P., Rand, D.G., 2016. Third-party punishment as a costly signal of trustworthiness. Nature 530 (7591), 473–476. https://doi.org/10.1038/nature16981.
King-Casas, B., Chiu, P.H., 2012. Understanding interpersonal function in psychiatric illness through multiplayer economic games. Biol. Psychiatr. 72 (2), 119–125. https://doi.org/10.1016/j.biopsych.2012.03.033.
Knapp, G., Hartung, J., 2003. Improved tests for a random effects meta-regression with a single covariate. Stat. Med. 22 (17), 2693–2710. https://doi.org/10.1002/sim.1482.
Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., Fehr, E., 2006. Diminishing reciprocal fairness by disrupting the right prefrontal cortex. Science 314, 829–832. https://doi.org/10.1126/science.1129156.
Knoch, D., Schneider, F., Schunk, D., Hohmann, M., Fehr, E., 2009. Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. Proc. Natl. Acad. Sci. U. S. A 106 (49), 20895–20899 doi:10.1073 pnas.0911619106.
Krueger, F., Hoffman, M., 2016. The emerging neuroscience of third-party punishment. Trends Neurosci. 39 (8), 499–501.
Lee, V.K., Harris, L.T., 2013. How social cognition can inform social decision making. Front. Neurosci. 7, 259. https://doi.org/10.3389/fnins.2013.00259.
Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Ioannidis, J.P., Moher, D., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 6 (7), e1000100 https://doi.org/10.1371/journal.pmed.1000100.
Lipsey, M.W., Wilson, D.B., 2001. Practical Meta-Analysis. Sage Publications, Thousand Oaks, CA.
Maier, M.J., Rosenbaum, D., Haeussinger, F.B., Brüne, M., Enzi, B., Plewnia, C., Ehlis, A. C., 2018. Forgiveness and cognitive control - provoking revenge via theta-burst-stimulation of the DLPFC. Neuroimage 183, 769–775. https://doi.org/10.1016/j.neuroimage.2018.08.065.
McCalley, D.M., Lench, D.H., Doolittle, J.D., Imperatore, J.P., Hoffman, M., Hanlon, C.A., 2021. Determining the optimal pulse number for theta burst induced change in cortical excitability. Sci. Rep. 11 (1), 8726. https://doi.org/10.1038/s41598-021-87916-2.
Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Group, P., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339, b2535. https://doi.org/10.1136/bmj.b2535.
Montague, P.R., Lohrenz, T., 2007. To detect and correct: norm violations and their enforcement. Neuron 56 (1), 14–18. https://doi.org/10.1016/j.neuron.2007.09.020.
Müller-Leinß, J.M., Enzi, B., Flasbeck, V., Brüne, M., 2017. Retaliation or selfishness? An rTMS investigation of the role of the dorsolateral prefrontal cortex in prosocial motives. Soc. Neurosci. 13 (6), 701–709. https://doi.org/10.1080/17470919.2017.1411828.
Nash, K., Schiller, B., Gianotti, L.R., Baumgartner, T., Knoch, D., 2013. Electrophysiological indices of response inhibition in a Go/NoGo task predict self-control in a social context. PLoS One 8 (11), e79462. https://doi.org/10.1371/journal.pone.0079462.
Rand, D.G., Brescoll, V.L., Everett, J.A., Capraro, V., Barcelo, H., 2016. Social heuristics and social roles: intuition favors altruism for women but not for men. J. Exp. Psychol. Gen. 145 (4), 389–396. https://doi.org/10.1037/xge0000154.
Rilling, J.K., Sanfey, A.G., 2011. The neuroscience of social decision-making. Annu. Rev. Psychol. 62, 23–48. https://doi.org/10.1146/annurev.psych.121208.131647.
Rover, C., Knapp, G., Friede, T., 2015. Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. BMC Med. Res. Methodol. 15, 99. https://doi.org/10.1186/s12874-015-0091-1.
Ruff, C.C., Ugazio, G., Fehr, E., 2013. Changing social norm compliance with noninvasive brain stimulation. Science 342 (6157), 482–484. https://doi.org/10.1126/science.1241399.
Sanfey, A.G., 2007. Social decision-making: insights from game theory and neuroscience. Science 318 (5850), 598–602. https://doi.org/10.1126/science.1142996.
Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2003. The neural basis of economic decision-making in the Ultimatum Game. Science 300 (5626), 1755–1758. https://doi.org/10.1126/science.1082976.
Sanfey, A.G., Stallen, M., Chang, L.J., 2014. Norms and expectations in social decision-making. Trends Cognit. Sci. 18 (4), 172–174. https://doi.org/10.1016/j.tics.2014.01.011.
Schmidt, F.L., Oh, I.S., Hayes, T.L., 2009. Fixed- versus random-effects models in meta-analysis: model properties and an empirical comparison of differences in results. Br. J. Math. Stat. Psychol. 62 (Pt 1), 97–128. https://doi.org/10.1348/000711007X255327.
Schwarzer, G., Carpenter, J.R., Rücker, G., 2015. Meta-Analysis with R. Springer International Publishing, Cham.
Smith, E.E., Jonides, J., 1999. Storage and executive processes in the frontal lobes. Science 283 (5408), 1657–1661.
Smolker, H.R., Depue, B.E., Reineberg, A.E., Orr, J.M., Banich, M.T., 2015. Individual differences in regional prefrontal gray matter morphometry and fractional anisotropy are associated with different constructs of executive function. Brain Struct. Funct. 220 (3), 1291–1306. https://doi.org/10.1007/s00429-014-0723-y.
Soutschek, A., Burke, C.J., Raja Beharelle, A., Schreiber, R., Weber, S.C., Karipidis II, Tobler, P.N., 2017. The dopaminergic reward system underpins gender differences in social preferences. Nat. Human Behav. 1 (11), 819–827. https://doi.org/10.1038/s41562-017-0226-y.
Soutschek, A., Sauter, M., Schubert, T., 2015. The importance of the lateral prefrontal cortex for strategic decision making in the prisoner's dilemma. Cognit. Affect Behav. Neurosci. 15 (4), 854–860. https://doi.org/10.3758/s13415-015-0372-5.
Speitel, C., Traut-Mattausch, E., Jonas, E., 2019. Functions of the right DLPFC and right TPJ in proposers and responders in the ultimatum game. Soc. Cognit. Affect Neurosci. 14 (3), 263–270. https://doi.org/10.1093/scan/nsz005.
Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., Fehr, E., 2007. The neural signature of social norm compliance. Neuron 56 (1), 185–196. https://doi.org/10.1016/j.neuron.2007.09.011.
Stallen, M., Rossi, F., Heijne, A., Smidts, A., De Dreu, C.K.W., Sanfey, A.G., 2018. Neurobiological mechanisms of responding to injustice. J. Neurosci. 38 (12), 2944–2954. https://doi.org/10.1523/JNEUROSCI.1242-17.2018.
Steinbeis, N., Bernhardt, B.C., Singer, T., 2012. Impulse control and underlying functions of the left DLPFC mediate age-related and age-independent individual differences in strategic social behavior. Neuron 73 (5), 1040–1051. https://doi.org/10.1016/j.neuron.2011.12.027.
Sterne, J.A., Sutton, A.J., Ioannidis, J.P., Terrin, N., Jones, D.R., Lau, J., Higgins, J.P., 2011. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 343, d4002. https://doi.org/10.1136/bmj.d4002.
Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B., Sack, A.T., 2015. Be nice if you have to - the neurobiological roots of strategic fairness. Soc. Cognit. Affect Neurosci. 10 (6), 790–796. https://doi.org/10.1093/scan/nsu114.

Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S., Kirsch, P., 2011. Beyond revenge: neural and genetic bases of altruistic punishment. Neuroimage 54 (1), 671–680. https://doi.org/10.1016/j.neuroimage.2010.07.051.

Treadway, M.T., Buckholtz, J.W., Martin, J.W., Jan, K., Asplund, C.L., Ginther, M.R., Marois, R., 2014. Corticolimbic gating of emotion-driven punishment. Nat. Neurosci. 17 (9), 1270–1275. https://doi.org/10.1038/nn.3781.

van't Wout, M., Kahn, R.S., Sanfey, A.G., Aleman, A., 2005. Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex a¡ects strategic decision-making. Neuroreport 16 (16), 1849–1852. https://doi.org/10.1097/01.wnr.0000183907.08149.14.

Verdejo-Garcia, A., Verdejo-Román, J., Albein-Urios, N., Martínez-González, J.M., Soriano-Mas, C., 2017. Brain substrates of social decision-making in dual diagnosis: cocaine dependence and personality disorders. Addiction Biol. 22 (2), 457–467.

Veroniki, A.A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Salanti, G., 2016. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res. Synth. Methods 7 (1), 55–79. https://doi.org/10.1002/jrsm.1164.

Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Software 36 (3), 1–48.

Wu, Y., Yu, H., Shen, B., Yu, R., Zhou, Z., Zhang, G., Zhou, X., 2014. Neural basis of increased costly norm enforcement under adversity. Soc. Cognit. Affect Neurosci. 9 (12), 1862–1871. https://doi.org/10.1093/scan/nst187.

Zinchenko, O., Arsalidou, M., 2018. Brain responses to social norms: meta-analyses of fMRI studies. Hum. Brain Mapp. 39 (2), 955–970. https://doi.org/10.1002/hbm.23895.

3. Chapter II: Causal roles of prefrontal and temporo-parietal theta oscillations for inequity aversion

Manuscript submitted to *Social Cognitive and Affective Neuroscience*, Christian, P., Kapetaniou, G. E. & Soutschek, A.

PC and AS designed research; PC performed research; PC, AS and GEK analyzed data; PC and AS wrote first draft of manuscript.

Causal roles of prefrontal and temporo-parietal theta oscillations for inequity aversion

Patricia Christian[1,2], Georgia E. Kapetaniou[1,2], and Alexander Soutschek[1,2]

[1] Department of Psychology, Ludwig Maximilians University Munich, Munich, Germany

[2] Graduate School of Systemic Neurosciences, Department of Biology, Ludwig Maximilians University Munich, Munich, Germany

Corresponding author:

Patricia Christian

Department of Psychology

Ludwig Maximilians University Munich

Leopoldstr. 13

80802 Munich, Germany

Email: patricia.christian@psy.lmu.de

Abstract

The right temporo-parietal junction (rTPJ) and the right lateral prefrontal cortex (rLPFC) are known to play prominent roles in human social behavior. However, it remains unknown which brain rhythms in these regions contribute to trading-off fairness norms against selfish interests as well as whether the influence of these oscillations depends on whether fairness violations are advantageous or disadvantageous for a decision maker. To answer these questions, we used noninvasive transcranial alternating current stimulation (tACS) to determine which brain rhythms in rTPJ and rLPFC are causally involved in moderating aversion to advantageous and disadvantageous inequity. Our results show that theta oscillations in rTPJ strengthen the aversion to unequal splits, which is statistically mediated by the rTPJ's role for perspective taking. Entrainment of theta oscillations in rLPFC, in contrast, showed dissociable effects depending on the type of inequity aversion: theta oscillations enhanced the preference for outcome-maximising choices more strongly when outcome distributions were disadvantageous compared to advantageous for the decision maker. Taken together, we provide evidence that neural oscillations in rTPJ and rLPFC have distinct causal roles in implementing inequity aversion, which can be explained by their involvement in distinct psychological processes.

Key words: transcranial alternating current stimulation (tACS), social decision making, perspective-taking, temporo-parietal junction, lateral prefrontal cortex.

Introduction

Fairness motives play an important role in guiding human social behavior by determining which payoff allocations are considered as desirable. Previous findings suggest that humans are averse to inequity both when they receive lower (disadvantageous inequity) and higher payoffs than others (advantageous inequity) (Fehr & Schmidt, 1999). Aversion to advantageous and disadvantageous inequity are hypothesized to relate to distinct psychological processes: While advantageous inequity aversion is discussed to rely on mentalizing processes enabling humans to take the perspective of others (Imuta, Henry, Slaughter, Selcuk, & Ruffman, 2016; Underwood & Moore, 1982), overcoming aversion to disadvantageous inequity may require downregulating the negative emotional reactions to unfair allocations (McAuliffe, Blake, Steinbeis, & Warneken, 2017). Despite the evidence that distinct psychological motives underlie prosocial behavior in the domains of advantageous and disadvantageous inequity, less is known about whether social decision making in these domains is implemented by dissociable brain mechanisms. While previous research ascribes the right temporo-parietal junction (rTPJ) and the right lateral prefrontal cortex (rLPFC) central roles in trading-off selfish interests against fairness norms (Cutler & Campbell-Meiklejohn, 2019; Maier et al., 2018; Speer & Boksem, 2019; Strang et al., 2015; Strombach et al., 2015; Will, Crone, & Guroglu, 2015; Yamagishi et al., 2016), the precise roles of these regions for advantageous or disadvantageous inequity aversion are poorly understood.

The rTPJ is thought to promote prosociality towards others, but there is disagreement on whether the rTPJ generally encodes reward values for others (Hare, Camerer, Knoepfle, O'Doherty, & Rangel, 2010; Hutcherson, Bushong, & Rangel, 2015; Park et al., 2017) or whether the rTPJ is more specifically involved in resolving conflicts between self- and other-regarding motives under advantageous inequity (Morishima, Schunk, Bruhin, Ruff, & Fehr, 2012; Obeso, Moisa, Ruff, & Dreher, 2018; Soutschek, Ruff, Strombach, Kalenscher, & Tobler, 2016). The rTPJ's function for fairness-guided behavior is often explained by its more general

role for perspective taking (Morishima et al., 2012; Soutschek et al., 2016; Strombach et al., 2015). Previous electrophysiological findings on brain rythms underlying higher-level cognitive functioning suggest that perspective taking is associated with theta oscillations in the rTPJ (Gooding-Williams, Wang, & Kessler, 2017; Seymour, Wang, Rippon, & Kessler, 2018; Wang, Callaghan, Gooding-Williams, McAllister, & Kessler, 2016). Consistent with this, previous research linked prosocial choices to temporo-parietal theta oscillations (Billeke et al., 2014), though other studies reported correlations between prosociality and beta, rather than theta, oscillations in the rTPJ (Gianotti, Dahinden, Baumgartner, & Knoch, 2019). These inconsistent findings raise the question as to whether perspective taking and social decision making are implemented by the same or dissociable brain rhythms within the rTPJ.

Likewise, also the role of the rLPFC for prosocial choices is controversially debated. The rLPFC has been hypothesized to play a key role in resolving conflicts between selfish interests and fairness considerations when being confronted with unfairness (Buckholtz et al., 2008; Buckholtz et al., 2015). Previous finding suggest that rLPFC promotes the rejection of unfair offers even if this reduces one's own payoff (Baumgartner, Knoch, Hotz, Eisenegger, & Fehr, 2011; Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006), which is consistent with a more general role of the rLPFC for goal-directed actions and cognitive control (Mansouri, Tanaka, & Buckley, 2009; Miller & Cohen, 2001; Muhle-Karbe, Jiang, & Egner, 2018; Yamagata, Nakayama, Tanji, & Hoshi, 2012). However, imaging studies directly comparing rLPFC activation between advantageous and disadvantageous inequity inconsistently reported stronger rLPFC activation either during advantageous inequity (Gao et al., 2018) or, in contrast, during disadvantageous inequity (Fliessbach et al., 2012). Thus, the rLPFC's role for moderating aversion to disadvantageous or advantageous inequity is far from understood. Further, even though previous research suggests that control processes in the rLPFC during negative feedback and conflict processing are associated with theta oscillations (Oehrn et al., 2014; van de Vijver, Ridderinkhof, & Cohen, 2011), the specific brain rythms underlying the

rLPFC's role for conflicts between fairness-guided behavior and selfish interests remain unkown.

To resolve the controversy between conflicting accounts on rTPJ and rLPFC functioning in social decision making, we conducted two experiments assessing the causal roles of rTPJ and rLPFC oscillations for social decision making with transcranial alternating current stimulation (tACS). In particular, we tested the following hypotheses: First, given the crucial role of theta oscillations for perspective taking in the rTPJ (Gooding-Williams et al., 2017; Seymour et al., 2018; Wang et al., 2016), we hypothesized that theta tACS over rTPJ strengthens inequity aversion by increasing the sensitivity to conflicts between selfish and other-regarding interests. Second, we expected that entrainment of theta oscillations in rLPFC reduces aversion to disadvantageous rather than advantageous outcomes due to the involvement of prefrontal theta in cognitive control (Oehrn et al., 2014; van de Vijver et al., 2011).

Materials and Methods

Participants

We tested 64 volunteers who were recruited at the Ludwig Maximilians University. We excluded data from three participants due to technical issues with tACS and from one participant due to electrode movement during the experiment, leaving 30 participants for the rTPJ experiment (17 female, M_{age} = 25.2 years, SD_{age} = 3.8 years) and 30 participants for the rLPFC experiment (14 female, M_{age} = 23.4 years, SD_{age} = 4.2 years). According to a power analysis assuming the effect size of Cohen's d = 0.54 observed in a previous tACS study on decision making (Soutschek, Moisa, Ruff, & Tobler, 2021), 29 participants are sufficient to detect significant effects (p = 0.05, two-tailed) with a power of 80%. All participants were healthy volunteers, without any known psychiatric or neurological disorders or contra-indications for tACS. The study was approved by the local ethics committee and conducted following the principles of the Declaration of Helsinki (2013) as well as the safety guidelines

for tACS (Bikson et al., 2016). All participants gave informed written consent prior to participation and received a payment of 10 Euro/hour as well as additional earnings from the social decision task (dictator game).

Dictator game

Participants played an adapted version of the dictator game (Hutcherson et al., 2015; Kapetaniou et al., 2021) implemented in Matlab 2019a (Mathworks, Inc., Sherborn, MA). In this task participants played in the role of the proposer ("dictator") and decided how to split a sum of coins between themselves (M_{self}) and another anonymous player (M_{other}). The monetary split could either be advantageous (proposer obtains higher payoff than receiver) or disadvantageous (proposer obtains less than receiver) for the participant. The participants were instructed to decide whether to accept or reject this proposed payoff; rejecting the offer resulted in an equal split for both players (Figure 1A). For the unequal choice option, the amounts for M_{self} and M_{other} varied from 1 to 31 coins (see Supplementary material), allowing us to disentangle efficiency concerns (combined payoff for both participants: $M_{self} + M_{other}$) and absolute inequity ($|M_{self} - M_{other}|$) (Gao et al., 2018; Kapetaniou et al., 2021). The dictator game included a total of 96 trials with equal numbers of advantageous and disadvantageous choice options. We also included catch trials where the unequal option was replaced by another equal option involving either higher (e.g. "18 coins for you, 18 for other") or lower stakes than the standard equal option (10 coins for both) to test participants' task understanding. Choice options were presented in random order to avoid repetition effects. Participants were informed that their choices had real consequences: One choice was randomly selected and the participant (M_{self}) and as well as the next participant coming to the lab (M_{other}) received a monetary bonus based on the participant's decision (with an exchange rate of 5 coins = 1 euro).

Director task

We used an adapted version of the director task to measure perspective taking (Dumontheil, Apperly, & Blakemore, 2010; Santiesteban, Banissy, Catmur, & Bird, 2012, 2015; Symeonidou, Dumontheil, Chow, & Breheny, 2016; Tamnes et al., 2018). In this task, a 4×4 set of shelves containing eight different objects and a human agent ("director") standing behind the shelves were displayed on the screen. All objects were visible fom the participant's perspective, whereas some objects were occluded from the director's perspective. Some of the presented objects belonged to the same category (e.g., balls), but differed in position (upper vs. lower shelves) or size (large vs. small). Participants had to follow the auditory instructions of the director (e.g., "Where is the small ball?"; instructions were given in German) and decide which target object was visible from the director's perspective. Following auditory instructions participants had to indicate within 3.6 seconds whether the target stimulus (e.g. small ball) was positioned on the right or the left side (Figure 1B). In control trials, the target object was visible from both the participant's and the director's perspective (congruent perspectives), such that participants could stick with their own perspective to select the target object. In contrast, in experimental trials the object which fitted to the auditory instructions from the participant's perspective (e.g., the smallest ball in the shelf) was occluded from the director's view (incongruent perspectives). Thus, to resolve the conflict between the incongruent perspectives, participants had to switch to the directors perspective to select the target object visible from the director's position (Dumontheil et al., 2010). Stimuli were presented in counterbalanced order across participants. To avoid repetition effects, no stimulus was presented more than once. The director task included a total of 96 trials (48 experimental and 48 control trials in random order).

tACS protocol

We used a 16-channel tDCS stimulator (neuroConn, Ilmenau, Germany) to apply tACS with sham, theta (6 Hz), or beta (20 Hz) stimulation frequency. For rTPJ stimulation, a smaller (5 × 5 cm) saline-soaked sponge electrode was placed vertically over electrode position CP6

and a larger (10 × 10 cm) electrode was placed horizontally over electrode position FP1 (Santiesteban et al., 2012) according to the international EEG 10/20 system. For rLPFC stimulation, the smaller electrode was placed horizontally over electrode position F4 and the larger electrode was placed horizontally over the occipital lobe (Frings, Brinkmann, Friehs, & van Lipzig, 2018). We applied online tACS during task performance with a current strength of 1 mA (peak-to-peak). Following previous procedures (Moisa, Polania, Grueschow, & Ruff, 2016; Soutschek et al., 2021; Soutschek, Nadporozhskaia, & Christian, 2022), tACS was administered during task performance in miniblocks lasting less than 3 min in order to minimize the risk of tACS-induced aftereffects. Each stimulation block started with a ramp-up phase for the tACS current for 15 seconds, followed by a buffer interval of 15 seconds before the start of the task to allow stimulation effects on brain activity to build up before task performance (Nitsche & Paulus, 2001; Vosskuhl, Huster, & Herrmann, 2016). In the sham condition, the current was ramped down directly after the ramp-up phase. During task performance, participants received online stimulation either for 152 seconds (dictator game) or for 122 seconds (director task). After each miniblock, participants had to indicate the perceived aversiveness of the stimulation on a rating scale from 0 (not aversive at all) to 20 (very aversive) within 5 seconds as measure of tACS-induced discomfort. The following block started after a stimulation-free interval of 35 seconds (including the ramp down phase of 5 s) to minimize carry-over effects between tACS conditions. In both experiments, the order of stimulation conditions was counterbalanced using latin square methods.

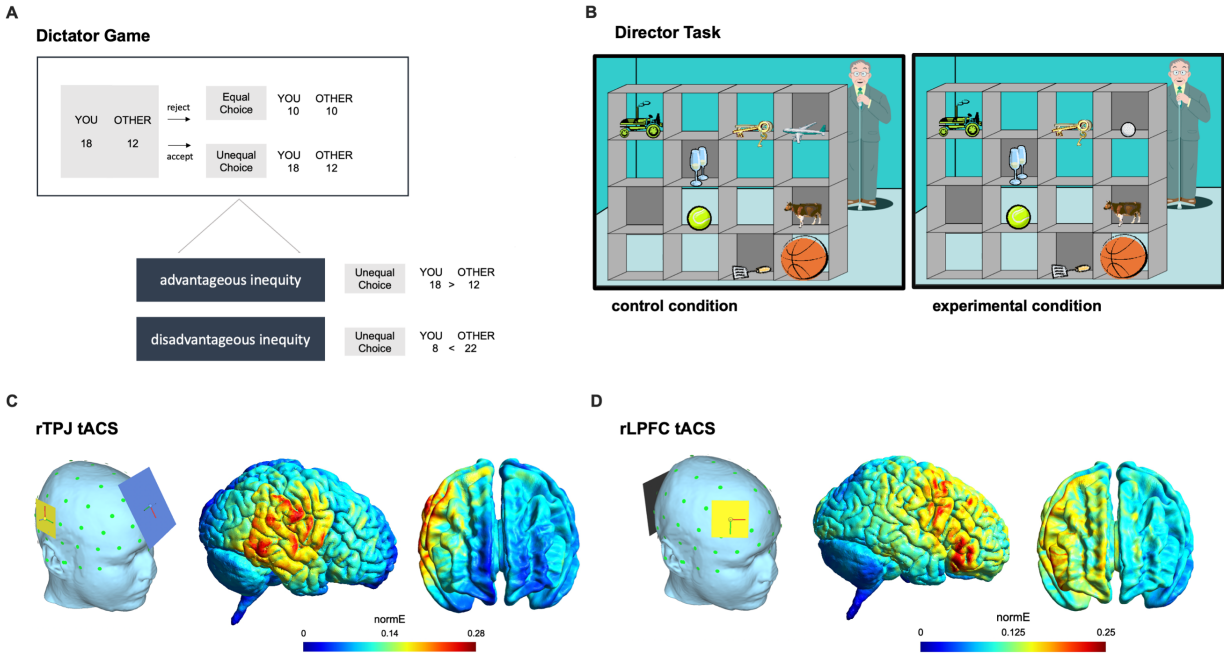

Figure 1. (A) Example trial of the dictator game. In this task participants were instructed to decide how to split an amount of money between themselves (M_{self}) and another co-player (M_{other}). In every trial participants were confronted with an unequal proposed monetary payoff which could either be advantageous ($M_{self} > M_{other}$) or disadvantageous ($M_{other} > M_{self}$) for themselves. Participants had to decide within 4 seconds whether to accept or reject the unequal split. If participants accepted the proposed payoff, the participant and the designated co-player received this monetary payoff at the end of the experiment. If participants rejected the unequal split, both the participant and the other player received a fixed amount of 10 coins (equal choice option). If participants failed to respond within 4 seconds, both players gained 0 coins. (B) Example trial of the director task: participants had to follow auditory instructions of the director ("Where is the small ball?") and select the designated target object visible from the director's view. In control trials, two objects belonging to the same category were presented (e.g., two balls), but only one of the objects matched the exact instruction of the director and was visible from both perspectives (in this example, the small yellow ball). In experimental trials, the director's view was incongruent with the participant's one (here, the smallest, white ball is occluded from director's view). To identify the target object (in the example, the yellow ball), participants had to inhibit their own perspective and take the director's perspective instead. (C/D) Simulations of electric current flow with the SimNIBS toolbox (Saturnino et al., 2019) for the (C) rTPJ and (D) rLPFC electrode placement.

Experimental Design

Experimental procedures were identical for the rTPJ and rLPFC experiments (apart from the electrode placement). Both experiments followed a within-subject design in which participants performed two experimental tasks (dictator game and director task) while undergoing sham, theta, or beta tACS. During tACS, participants performed 6 miniblocks of the dictator game (22 trials each) and 6 miniblocks (16 trials each) of the director task. The task order was counterbalanced across participants. In total, one session lasted approximately one hour and 30 minutes.

Statistical Analysis

We analysed data of the dictator game with Bayesian generalized linear mixed models (GLMMs) as implemented in the brms package in R (Bürkner, 2017). We analysed choice data with model-free GLMMs rather than with model-based parameter estimates for the Fehr-Schmidt model because the Fehr-Schmidt model was shown to provide a poor fit of dictator game data (Engelmann & Strobel, 2004) and trial numbers in the current experiment might not be sufficient for a reliable estimate of individual model parameters. Following the procedures we had established in our previous study (Kapetaniou et al., 2021), we therefore assessed the impact of rTPJ and rLPFC tACS on social decision making as a function of the degree of inequity between the participant's and the receiver's payoff ($Inequity_{absolute} = |M_{self} - M_{other}|$) as well as the efficiency of an offer, i.e. the overall payoff for both participants ($Efficiency = M_{self} + M_{other}$). In more detail, for both the rTPJ and the rLPFC experiment, we performed Bayesian GLMMs regressing binary choices (0 = equal option, 1 = unequal option) on fixed-effect predictors for $tACS_{theta\text{-}sham}$, $tACS_{beta\text{-}sham}$, $Inequity_{type}$ (0 = advantageous inequity, 1 = disadvantageous inequity), $Inequity_{absolute}$, Efficiency, and the interaction terms. We also included discomfort ratings after each tACS block as predictors of no interest to control for potential confounding effects of tACS-induced discomfort on choices. All fixed-effect

predictors were also modelled as random slopes in addition to participant-specific random intercepts. Continuous predictors were z-transformed. We assessed the statistical significance of model parameters with the 95% highest density interval (HDI) of the posterior distributions (Ahn, Haines, & Zhang, 2017; Kruschke, 2010). Parameter values falling within the 95% HDI are considered as more credible than parameter values outside of the HDI (Kruschke, 2013, 2018). If the 95% HDI does not overlap with zero, parameter erstimates are considered as statistically significant (Kruschke, 2013), in analogy to frequentist statistics. To minimize the impact of priors on the parameter estimates, we used weakly informative flat uniform distributions as priors as implemented in the brms package (Bürkner, 2017). The model was fitted with 2 Markov chain Monte Carlo (MCMC) chains with 3000 iterations, including 1000 warm-up iterations. We used $\hat{R}$ as measure of model convergence: $\hat{R}$ was below 1.01 for all parameter estimates, suggesting model convergence.

In the director task, we used Bayesian GLMMs to analyse performance accuracy. Accurate responses in experimental trials reflect the participant's ability to take the perspective of the director in case of conflict between one's own and the director's incongruent perspectives, whereas in control trials the participants' and the director's perspectives were congruent. We regressed binary responses (1 = correct, 0 = incorrect response) on fixed-effect predictors for $tACS_{theta\text{-}sham}$, $tACS_{beta\text{-}sham}$, Condition (1 = experimental, 0 = control), and the interaction terms. Again, we entered discomfort ratings as covariate of no interest. All fixed effects were modelled also as random slopes in addition to participant-specific intercepts. For the analysis of the director task, we used the same model fitting procedures as for the dictator game. $\hat{R}$ values were below 1.01 for all parameter estimates, suggesting that all models converged.

Results

Theta tACS over rTPJ and rLPFC affect dissociable aspects of inequity aversion

First, we tested the causal involvement of the rTPJ and rLPFC in advantageous and disadvantageous inequity aversion. For both the rTPJ and the rLPFC experiment, we regressed binary choices on predictors for $tACS_{theta\text{-}sham}$, $tACS_{beta\text{-}sham}$, $Inequity_{type}$ (advantageous inequity = 0, disadvantageous inequity = 1), $Inequity_{absolute}$, Efficiency, and the interaction terms, controlling for tACS-induced discomfort.

In the rTPJ experiment, sanity checks revealed that participants accepted unequal splits less often if inequity was disadvantageous compared to advantageous for them, $Inequity_{type}$: HDI_{mean} = -2.65, $HDI_{95\%}$ = [-4.49, -0.87], indicating a stronger aversion against disadvantageous than advantageous inequity (Fehr & Schmidt, 1999). They also preferred more efficient choices, Efficiency: HDI_{mean} = 4.14, $HDI_{95\%}$ = [2.71, 5.77], as well as options with smaller absolute differences between M_{self} and M_{other}, $Inequity_{absolute}$: HDI_{mean} = -1.32, $HDI_{95\%}$ = [-2.08, -0.58], the latter suggesting that participants were inequity averse. Furthermore, participants were more averse to increasing disadvantageous compared to advantageous inequity, $Inequity_{absolute} \times Inequity_{type}$: HDI_{mean} = -2.15, $HDI_{95\%}$ = [-3.34, -1.05], and showed a stronger preference for efficient (i.e., payoff-maximising) outcomes under disadvantageous in contrast to advantageous inequity, Efficiency × $Inequity_{type}$: HDI_{mean} = 4.13, $HDI_{95\%}$ = [2.25, 6.42]. Taken together, participants' preferences for unequal splits strongly depended on whether inequity was advantageous or disadvantageous for them.

Next, we assessed how rTPJ tACS affected choices in the dictator game. We observed that $tACS_{theta\text{-}sham}$ significantly increased inequity aversion, $tACS_{theta\text{-}sham} \times Inequity_{absolute}$: HDI_{mean} = -0.73, $HDI_{95\%}$ = [-1.48, -0.04] (Table 1), irrespective of whether inequity was advantageous or disadvantageous, $tACS_{theta\text{-}sham} \times Inequity_{type} \times Inequity_{absolute}$: HDI_{mean} = 0.58, $HDI_{95\%}$ = [-0.47, 1.65]. This suggests that theta tACS promotes the rejection of unequal splits independently of whether the participant or the other were worse off (Figure 2). We observed no significant effects of $tACS_{beta\text{-}sham}$ on $Inequity_{absolute}$, Efficiency, or $Inequity_{type}$ (Table 1), and also a further GLMM directly comparing theta and beta tACS revealed no significant

stimulation effects. Thus, entrainment of theta oscillations in rTPJ increases aversion to unequal splits independently of whether these payoff splits are advantageous or disadvantageous for the participant.

A different pattern of stimulation effects emerged in the rLPFC experiment: In the sham condition, we again observed significant main effects of $Inequity_{type}$: HDI_{mean} = -4.56, $HDI_{95\%}$ = [-7.36, -1.99], $Inequity_{absolute}$: HDI_{mean} = -1.57, $HDI_{95\%}$ = [-2.46, -0.64], and Efficiency: HDI_{mean} = 4.06, $HDI_{95\%}$ = [2.53, 5.97]. As in the rTPJ experiment, participants also showed stronger preferences for more efficient, Efficiency × $Inequity_{type}$: HDI_{mean} = 4.59, $HDI_{95\%}$ = [2.34, 7.35], and less unequal options, $Inequity_{absolute}$ × $Inequity_{type}$: HDI_{mean} = -2.21, $HDI_{95\%}$ = [-3.74, -0.92], for disadvantageous relative to advantageous unequal splits. When we assessed the effects of rLPFC tACS on choice behaviour, we observed that $tACS_{theta\text{-}sham}$ significantly increased the impact of efficiency on choices depending on inequity type, $tACS_{theta\text{-}sham}$ × Efficiency × $Inequity_{type}$: HDI_{mean} = 1.67, $HDI_{95\%}$ = [0.22; 3.13] (Figure 2, Table 2). This suggests that theta tACS over rLPFC strengthens the preference for efficient choices more strongly under disadvantageous than under advantageous inequity: When participant receive less payoff than the other, they show a stronger preference for the payoff-maximizing unequal option under rDLPFC theta tACS compared with sham. We observed no significant effects of $tACS_{beta\text{-}sham}$ on $Inequity_{absolute}$, Efficiency, or $Inequity_{type}$ (Table 2), and also a further GLMM comparing theta versus beta tACS yielded no significant stimulation effects.

Taken together, our results provide evidence that theta oscillations in rLPFC are causally involved in increasing the preference for options that maximize the overall welfare particularly when the decision maker is worse, rather than better, off than the other receiver. In contrast, theta tACS over the rTPJ increased inequity aversion independently of whether inequity is advantageous or disadvantageous for the decision maker.

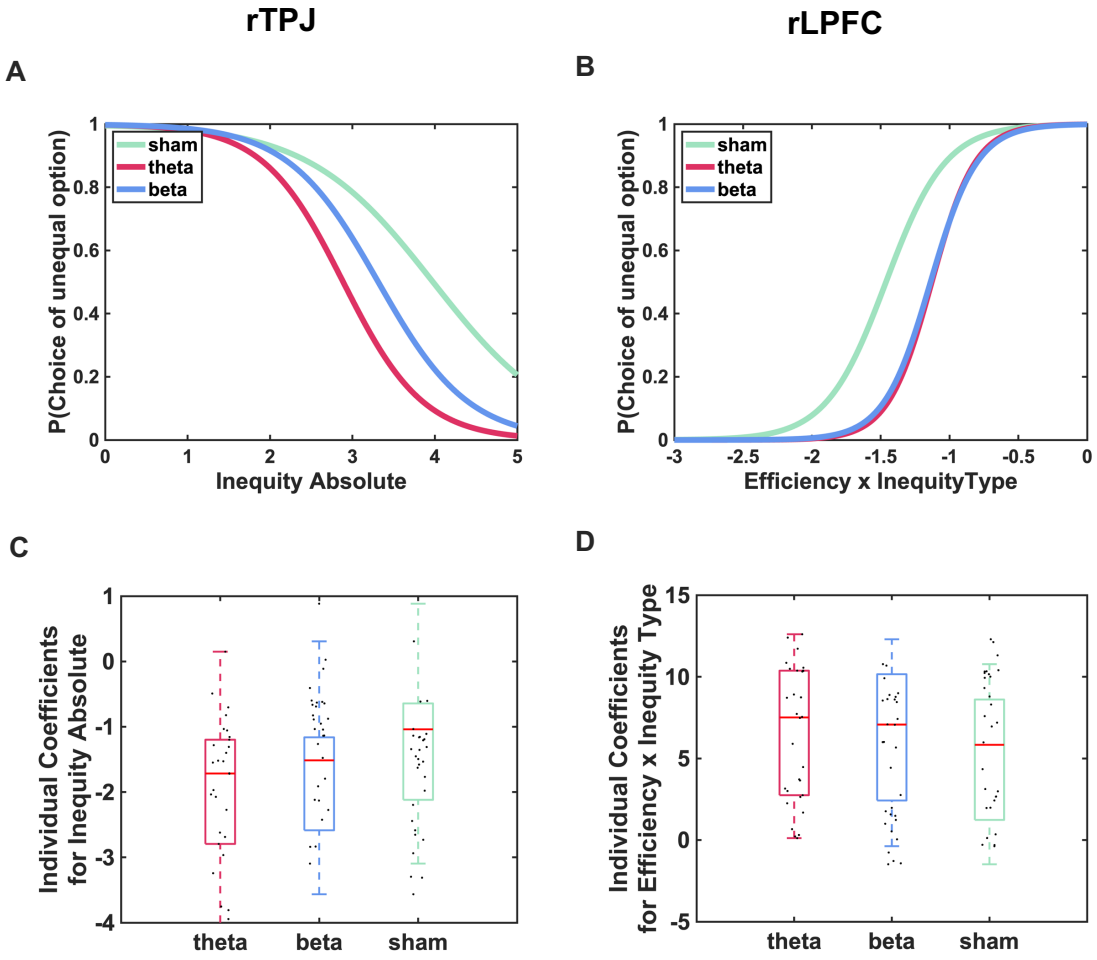

Figure 2. Stimulation effects on inequity aversion based on the results of bayesian generalized linear mixed models in the dictator game. (A) theta tACS over rTPJ lowers acceptance rates of unequal splits with increasing inequity between the players (which indicates increased inequity aversion), irrespective of whether inequity is advantageous or disadvantageous for the participant. (B) In contrast, theta tACS over rLPFC increases the acceptance of efficient options (i.e., unequal choice option that maximize the overall payoff for both players) more strongly under disadvantageous compared with advantageous inequity (C) Individual regression coefficients for the impact of $Inequity_{absolute}$ in the rTPJ experiment, separately for each tACS condition. More negative values indicate a stronger aversion against unequal choice options. Colored boxes indicate median and interquartile range, black dots indicate individual data points (N = 30 participants). (D) Individual regression coefficients for the impact of efficiency as function of inequity type in the rLPFC experiment, separately for each tACS condition. Higher positive values indicate a stronger preference for more efficient outcomes (i.e. maximizing the overall welfare) when the participant is worse compared to better off than the other. Colored boxes indicate median and interquartile range, black dots indicate individual data points (N = 30 participants).

Theta oscillations in rTPJ causally implement perspective tasking

The observed effects of rTPJ stimulation on inequity aversion raise the question as to whether these findings can be explained by the established role of the rTPJ for mentalizing and perspective taking (Frith & Frith, 2006; Saxe & Kanwisher, 2003; Schurz, Aichhorn, Martin, & Perner, 2013; Van Overwalle, 2009). Despite the evidence that the rTPJ is causally relevant for perspective taking (Santiesteban et al., 2012, 2015), the underlying brain oscillations causally implementing the ability to differentiate between one's own and others' perspectives remain unknown. Therefore, in line with recent findings (Gooding-Williams et al., 2017), we tested whether rTPJ theta tACS promotes perspective taking and, if so, whether the link between rTPJ theta oscillations and inequity aversion can be explained by their role for perspective taking. We tested the causal involvement of tACS over rTPJ and rLPFC on performance in the director task. We regressed performance in the director task (correct versus incorrect responses) on predictors for $tACS_{theta\text{-}sham}$, $tACS_{beta\text{-}sham}$, Condition (control = 0, experimental = 1), and the interaction terms.

In line with previous studies (Santiesteban et al., 2012), we found a significant main effect of Condition on accuracy in the rTPJ experiment, HDI_{mean} = -2.21, $HDI_{95\%}$ = [-3.37; -0.94], suggesting that participants committed more errors when their perspective was incongruent, compared to congruent, with the director's perspective. The significant $tACS_{theta\text{-}sham}$ × Condition interaction, HDI_{mean} = 1.56, $HDI_{95\%}$ = [0.55; 2.54], suggested that (as hypothesized) rTPJ theta tACS effects on accuracy depended on whether perspectives were congruent or incongruent, whereas we observed no significant $tACS_{beta\text{-}sham}$ × Condition interaction, HDI_{mean} = 0.53, $HDI_{95\%}$ = [-0.50; 1.58] (Table 3). Post-hoc GLMMs showed that in the experimental condition theta tACS increased accuracy in contrast to sham, HDI_{mean} = 1.57, $HDI_{95\%}$ = [1.07; 2.14] (Figure 3), whereas we could not find significant effects of theta tACS in the control condition: $tACS_{theta\text{-}sham}$, HDI_{mean} = 0.92, $HDI_{95\%}$ = [-0.65; 3.11]. $tACS_{beta\text{-}sham}$ affected performance neither in experimental, HDI_{mean} = 0.25, $HDI_{95\%}$ = [-0.47; 1.10], nor

in control trials, HDI_{mean} = -0.33, $HDI_{95\%}$ = [-1.36; 0.80]. Aditionally, GLMMs comparing the effects of theta versus beta tACS revealed that the influence of theta tACS on performance in experimental versus control trials was significantly stronger than the influence of beta tACS, $tACS_{theta\text{-}beta}$ × Condition, HDI_{mean} = -1.28, $HDI_{95\%}$ = [-2.28; -0.23] (Figure 3 and Table 4). Thus, rTPJ theta tACS, relative to both sham tACS and beta tACS, improved participants' ability to inhibit their own perspective in order to resolve conflicts between their own and the director's perspective.

When conducting the same GLMMs for the rLPFC experiment, neither theta nor beta tACS significantly affected performance in the director task in contrast to sham tACS, $tACS_{theta\text{-}sham}$ × Condition: HDI_{mean} = -0.03, $HDI_{95\%}$ = [-1.06; 1.00], $tACS_{beta\text{-}sham}$ × Condition: HDI_{mean} = 0.46, $HDI_{95\%}$ = [-0.60; 1.52] (Table 5), and we also observed no significant differences between theta and beta tACS, $tACS_{theta\text{-}beta}$ × Condition: HDI_{mean} = 0.41, $HDI_{95\%}$ = [-0.54; 1.42] (Table 6). Consequently, contrary to the rTPJ, there was no evidence for rLPFC involvement in perspective taking.

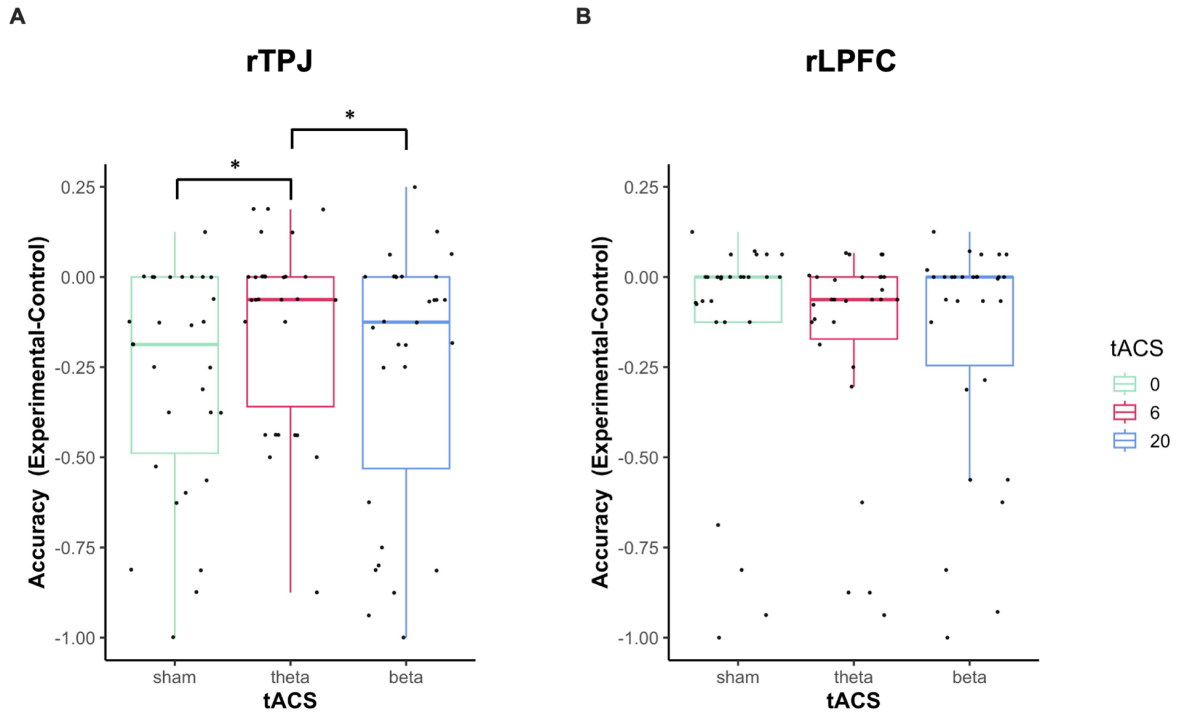

Figure 3. Results of tACS effects on accuracy (perspective taking) in the director task. (A) rTPJ theta tACS, compared with sham tACS and beta tACS, increased accuracy in the experimental condition relative to the control condition. (B) rLPFC tACS showed no significant effects on accuracy in the director task. Colored boxes indicate median and interquartile range, black dots indicate individual data points (N = 30 participants). Values are calculated based on the difference between each participant's accuracy rates in the experimental and the control condition: In control trials, participants could stick with their own perspective to identify the object designated by the director, whereas in experimental trials participants needed to switch from their own to the director's perspective. More negative values indicate worse performance in the experimental relative to the control condition, reflecting the need to resolve conflicts between one's own and the director's perspective.

Impact of rTPJ tACS on perspective taking mediates stimulation effects on inequity aversion

Given that theta tACS over the rTPJ enhanced both perspective taking and inequity aversion, we conducted a mediation analysis to test whether the rTPJ's involvement in inequity aversion can statistically be explained by its more general role for perspective taking. For this purpose, we entered mean individual accuracy differences between experimental and control

trials under theta versus sham tACS from the director task as additional predictors to the GLMM we had used to analyse rTPJ tACS effects on choices in the dictator game (Baron & Kenny, 1986; Preacher & Hayes, 2008). Re-computing this GLMM revealed that, contrary to the original GLMM results, the effect of $tACS_{theta\text{-}sham}$ on inequity aversion no longer passed the statistical threshold, $tACS_{theta\text{-}sham} \times Inequity_{absolute}$: HDI_{mean} = -0,89, $HDI_{95\%}$ = [-1.95, 0.08], which suggests that the effect of $tACS_{theta\text{-}sham}$ on inequity aversion is reduced when controlling for stimulation effects on perspective taking. Crucially, the marginally significant result of the Sobel test (Sobel, 1982, 2008), $z = 1.92$, $p = 0.05$, suggests that the impact of rTPJ theta tACS on inequity aversion can statistically be explained by tACS-induced changes in perspective taking. In contrast, we found no significant mediation effects in the rLPFC experiment, Sobel test: $z = -0.06$, $p = 0.95$. Thus, the role of the rTPJ, though not of the rLPFC, in inequity aversion can be explained by its more general contribution to perspective taking.

Discussion

Both rTPJ and rLPFC are thought to play central roles in social decision making, but their causal contributions to moderating aversion to advantageous versus disadvantageous inequity as well as the brain rhythms underlying these functions remained unknown so far. Here, we advance the field by determining the specific roles of neural oscillations in rTPJ and rLPFC for advantageous and disadvantageous inequity aversion. While entrainment of theta oscillations in rTPJ increased inequity aversion irrespective of whether the unequal splits were advantageous or disadvantageous for an individual, theta tACS over rLPFC showed dissociable effects depending on the type of inequity involved: theta tACS increased the preference for welfare-maximizing efficient choices more strongly for disadvantageous than for advantageous unequal splits. Moreover, our data suggest that rTPJ and rLPFC affect social decisions via dissociable cognitive mechanisms, as only theta stimulation of rTPJ, but not rLPFC, changed perspective taking processes, which statistically explained the rTPJ tACS effects on inequity

aversion. Taken together, our study provides evidence for dissociable neuro-cognitive roles of theta oscillations in rTPJ and rLPFC for weighing inequity concerns against selfish interests, improving our understanding of the functions of these brain mechanisms in social decision making.

Although previous evidence suggested a causal role of rTPJ for prosocial giving (Obeso et al., 2018; Soutschek et al., 2016), these brain stimulation studies did not differentiate between different types of inequity. A neuroimaging study dissociating between advantageous and disadvantageous inequity reported that grey matter volume in the rTPJ predicted individual differences in advantageous, but not disadvantageous, inequity aversion (Morishima et al., 2012). While our findings suggest that rTPJ stimulation indeed enhances inequity aversion, there was no evidence for dissociables effects on advantageous and disadvantageous inequity aversion, though we note that the lack of a significant difference must not be interpreted as evidence against inequity-specific contributions of the rTPJ to decision making (Obeso et al., 2018; Soutschek et al., 2016). Nevertheless, our findings point to a general function of the rTPJ for integrating own and others' needs into the choice process, which increases the preference for equal splits in order to reduce the conflict between selfish and other-regarding interests. From a psychological perspective, this function may rely on the ability to distinguish between own and others' mental states, which on the neural level is implemented by the rTPJ (Martin, Huang, Hunold, & Meinzer, 2019; Martin, Kessler, Cooke, Huang, & Meinzer, 2020; Santiesteban et al., 2012; Zhang, Chen, Hu, & Mai, 2019). While a link between the rTPJ's roles for perspective taking and social decision making has often been discussed in the literature (Baumgartner, Schiller, Rieskamp, Gianotti, & Knoch, 2014; Soutschek et al., 2016; Strombach et al., 2015), our mediation analysis provides conclusive evidence that the rTPJ's causal involvement in perspective taking indeed statistically explains its contribution to social decision making. Thus, rTPJ theta oscillations enable us to put ourselves into the shoes of others, which

increases the sensitivity for conflicts between selfish and other-regarding interests, thereby motivating choices that reduce inequity between the players' payoffs.

A different result pattern emerged in the rLPFC tACS experiment, where theta entrainment in rLPFC increased the preference for efficient (i.e., welfare-maximizing) choice options more strongly for disadvantageous than for advantageous inequity. As under disadvantageous inequity the more efficient option is likely to include a lower payoff for the decision maker than the less efficient option, this finding suggests that rLPFC theta oscillations lower aversion to disadvantageous inequity if the unequal option maximizes the overall welfare. While this appears to speak in favor of the hypothesis that rLPFC promotes outcome-maximizing choices by downregulating negative emotional responses to unfairness (Maier et al., 2018; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003), this view appears inconsistent with findings according to which the rLPFC activation is associated with rejection of unfair offers (which reduces the overall welfare) (Baumgartner et al., 2011; Knoch et al., 2006) or with the punishment of norm violations (Buckholtz et al., 2008; Buckholtz et al., 2015). To reconcile these findings, we posit that the rLPFC generally strengthens norm-guided behavior, which can lead either to the punishment of others' norm violations or to the reduction of envy if others are better off than oneself without being responsible for the unequal outcomes.

In addition to determining the roles of rTPJ and rLPFC in inequity aversion, our findings also provide insights into the brain oscillations implementing these functions. The causal involvement of theta oscillations in both rTPJ and rLPFC in social decision making is consistent with previous electrophysiological findings linking theta oscillations in these regions to perspective taking (Gooding-Williams et al., 2017; Rodrigues, Ulrich, & Hewig, 2015; Seymour et al., 2018; Wang et al., 2016) and cognitive control processes (Oehrn et al., 2014; van de Vijver et al., 2011), respectively. Note that only in the director task we found that rTPJ theta tACS affected behavior relative to both sham and beta tACS, whereas in the dictator game we observed no significant differences between theta and beta tACS. One possible reason for

this is that also beta oscillations might play a role in social decision making (though these effects did not pass the statistical threshold in the current study). Previous findings linked rTPJ beta oscillations to individual differences in prosociality (Gianotti et al., 2019), which might hint to dissociable roles of theta and beta oscillations in the rTPJ for social decisions. Thus, we are cautious with any conclusions regarding the frequency-specific of our stimulation effects on inequity aversion. As further limitation, it is worth keeping the relatively low spatial specificity of tDCS in mind, allowing no inferences regarding which precise subregions in the prefrontal and the parietal cortices are responsible for the observed effects. Nevertheless, our findings provide first evidence for a causal contribution of prefrontal and parietal theta oscillations to social decision making.

Deficits in social decision making belong to the core symptoms of several psychiatric disorders (Chang, Barack, & Platt, 2012; King-Casas & Chiu, 2012) and previous findings suggest that these deficits in social interactions are linked with dysfunctions in the pefrontal cortex and parietal regions (Bitsch, Berger, Nagels, Falkenberg, & Straube, 2019; Horat et al., 2018; Hu et al., 2021; Schneider et al., 2013). Studying cortical oscillatory dynamics can lead to a better understanding of the neuronal mechanisms underlying deficits in cognitive and social impairments in psychiatric disorders (Kirihara, Rissling, Swerdlow, Braff, & Light, 2012). By providing insights into how brain rythms in prefrontal and parietal brain regions implement social decision making, our findings contribute to improving our understanding of the neural basis of altered social behavior in psychiatric disorders.

To sum up, our findings demonstrate that theta oscillations in rTPJ and rLPFC causally moderate aversion to unequal outcomes in social interactions. The dissociable effects of rTPJ and rLPFC stimulation on inequity aversion are consistent with the idea that (at least partially) different neuro-cognitive mechanisms underlie aversion to disadvantageous and advantageous inequity. These insights into the brain rhythms causally implementing perspective taking and

inequity aversion extend our understanding of the neuronal signature of the processes underlying social behavior.

Acknowledgements

We kindly thank the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA) for support with recruitment and data collection.

Funding:

This work was supported by the German Research Foundation: AS received the Emmy Noether fellowship with the Grant reference number: SO 1636/2-1.

Ethics

The study protocol was approved by the Ethics Committee of the Department of Psychology, LMU Munich. For this study we collected data from humans subjects, which gave written informed consent before their participation in the study.

Declaration of competing interest

The authors declared that there were no conflicts of interest in relation to the subject of this study.

Data availability statement

The data that support the findings of this study will be available on Open Science Framework (https://osf.io/8ybp7/?view_only=11e1775f38224deaa2311aa9f93d1f80).

Supplementary materials

Supplementary Table S1/S2: Overview of unequal choice options for advantageous and disadvantageous inequity with values for $M_{Self,}$ M_{Other}, $Inequity_{absolute}$, Efficiency and $Inequity_{type}$.

References

Ahn, W. Y., Haines, N., & Zhang, L. (2017). Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. *Computational Psychiatry, 1*, 24-57.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173 - 1182.

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience, 14*(11), 1468-1474.

Baumgartner, T., Schiller, B., Rieskamp, J., Gianotti, L. R., & Knoch, D. (2014). Diminishing parochialism in intergroup conflict by disrupting the right temporo-parietal junction. *Social Cognitive and Affective Neuroscience, 9*(5), 653-660.

Bikson, M., Grossman, P., Thomas, C., Zannou, A. L., Jiang, J., Adnan, T., . . . Woods, A. J. (2016). Safety of Transcranial Direct Current Stimulation: Evidence Based Update 2016. *Brain Stimulation, 9*(5), 641-661.

Billeke, P., Zamorano, F., Lopez, T., Rodriguez, C., Cosmelli, D., & Aboitiz, F. (2014). Someone has to give in: theta oscillations correlate with adaptive behavior in social bargaining. *Social Cognitive and Affective Neuroscience, 9*(12), 2041-2048.

Bitsch, F., Berger, P., Nagels, A., Falkenberg, I., & Straube, B. (2019). Impaired Right Temporoparietal Junction-Hippocampus Connectivity in Schizophrenia and Its Relevance for Generating Representations of Other Minds. *Schizophrenia Bulletin, 45*(4), 934-945.

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron, 60*(5), 930-940.

Buckholtz, J. W., Martin, J. W., Treadway, M. T., Jan, K., Zald, D. H., Jones, O., & Marois, R. (2015). From Blame to Punishment: Disrupting Prefrontal Cortex Activity Reveals Norm Enforcement Mechanisms. *Neuron, 87*(6), 1369-1380.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1-28.

Chang, S. W., Barack, D. L., & Platt, M. L. (2012). Mechanistic classification of neural circuit dysfunctions: insights from neuroeconomics research in animals. *Biological Psychiatry, 72*(2), 101-106.

Cutler, J., & Campbell-Meiklejohn, D. (2019). A comparative fMRI meta-analysis of altruistic and strategic decisions to give. *Neuroimage, 184*, 227-241.

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science, 13(2)*, 331-338.

Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review, 94*, 857 - 869.

Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics, 114(*(3), 817-868.

Fliessbach, K., Phillipps, C., Trautner, P., Schnabel, M., Elger, C., Falk, A., & Weber, B. (2012). Neural responses to advantageous and disadvantageous inequity. *Frontiers in Human Neuroscience, 6*.

Frings, C., Brinkmann, T., Friehs, M. A., & van Lipzig, T. (2018). Single session tDCS over the left DLPFC disrupts interference processing. *Brain and Cognition, 120*, 1-7.

Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron, 50*(4), 531-534.

Gao, X., Yu, H., Saez, I., Blue, P. R., Zhu, L., Hsu, M., & Zhou, X. (2018). Distinguishing neural correlates of context-dependent advantageous- and disadvantageous-inequity

aversion. *Proceedings of the National Academy of Sciences of the United States of America, 115*(33), 7680-7689.

Gianotti, L. R. R., Dahinden, F. M., Baumgartner, T., & Knoch, D. (2019). Understanding Individual Differences in Domain-General Prosociality: A Resting EEG Study. *Brain Topography, 32*(1), 118-126.

Gooding-Williams, G., Wang, H., & Kessler, K. (2017). THETA-Rhythm Makes the World Go Round: Dissociative Effects of TMS Theta Versus Alpha Entrainment of Right pTPJ on Embodied Perspective Transformations. *Brain Topography, 30*(5), 561-564.

Hare, T. A., Camerer, C. F., Knoepfle, D. T., O'Doherty, J. P., & Rangel, A. (2010). Value Computations in Ventral Medial Prefrontal Cortex during Charitable Decision Making Incorporate Input from Regions Involved in Social Cognition. *Journal of Neuroscience, 30*(2), 583-590.

Horat, S. K., Favre, G., Prevot, 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.

Hu, Y., Pereira, A. M., Gao, X., Campos, B. M., Derrington, E., Corgnet, B., . . . Dreher, J. C. (2021). Right Temporoparietal Junction Underlies Avoidance of Moral Transgression in Autism Spectrum Disorder. *Journal of Neuroscience, 41*(8), 1699-1715.

Hutcherson, C. A., Bushong, B., & Rangel, A. (2015). A Neurocomputational Model of Altruistic Choice and Its Implications. *Neuron, 87*(2), 451-462.

Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., & Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: A meta-analytic review. *Developmental Psychology, 52*(8), 1192-1205.

Kapetaniou, G. E., Reinhard, M. A., Christian, P., Jobst, A., Tobler, P. N., Padberg, F., & Soutschek, A. (2021). The role of oxytocin in delay of gratification and flexibility in non-social decision making. *eLife, 10*.

King-Casas, B., & Chiu, P. H. (2012). Understanding interpersonal function in psychiatric illness through multiplayer economic games. *Biological Psychiatry, 72*(2), 119-125.

Kirihara, K., Rissling, A. J., Swerdlow, N. R., Braff, D. L., & Light, G. A. (2012). Hierarchical organization of gamma and theta oscillatory dynamics in schizophrenia. *Biological Psychiatry, 71*(10), 873-880.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex. *Science, 314*, 829-832.

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences, 14*(7), 293-300.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of experimental psychology. General, 142*(2), 573-603.

Kruschke, J. K. (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation. *Advances in Methods and Practices in Psychological Science, 1*(2), 270-280.

Maier, M. J., Rosenbaum, D., Haeussinger, F. B., Brune, M., Enzi, B., Plewnia, C., . . . Ehlis, A. C. (2018). Forgiveness and cognitive control - Provoking revenge via theta-burst-stimulation of the DLPFC. *Neuroimage, 183*, 769-775.

Mansouri, F. A., Tanaka, K., & Buckley, M. J. (2009). Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. *Nature Review Neuroscience, 10*(2), 141-152.

Martin, A. K., Huang, J., Hunold, A., & Meinzer, M. (2019). Dissociable Roles Within the Social Brain for Self-Other Processing: A HD-tDCS Study. *Cerebral Cortex, 29*(8), 3642-3654.

Martin, A. K., Kessler, K., Cooke, S., Huang, J., & Meinzer, M. (2020). The Right Temporoparietal Junction Is Causally Associated with Embodied Perspective-taking. *Journal of Neuroscience, 40*(15), 3089-3095.

McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017). The developmental foundations of human fairness. Nature Human Behaviour, 1(2), 0042. doi:10.1038/s41562-016-0042. *Nature Human Behaviour, 1*, 0042.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review Neuroscience, 24*, 167–202.

Moisa, M., Polania, R., Grueschow, M., & Ruff, C. C. (2016). Brain Network Mechanisms Underlying Motor Enhancement by Transcranial Entrainment of Gamma Oscillations. *Journal of Neuroscience, 36*(47), 12053-12065.

Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., & Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron, 75*(1), 73-79.

Muhle-Karbe, P. S., Jiang, J., & Egner, T. (2018). Causal Evidence for Learning-Dependent Frontal Lobe Contributions to Cognitive Control. *Journal of Neuroscience, 38*(4), 962-973.

Nitsche, M. A., & Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology, 57*(10), 1899-1901.

Obeso, I., Moisa, M., Ruff, C. C., & Dreher, J.-C. (2018). A causal role for right temporo-parietal junction in signaling moral conflict. *eLife, 7*, e40671.

Oehrn, C. R., Hanslmayr, S., Fell, J., Deuker, L., Kremers, N. A., Do Lam, A. T., . . . Axmacher, N. (2014). Neural communication patterns underlying conflict detection, resolution, and adaptation. *Journal of Neuroscience, 34*(31), 10438-10452.

Park, S. Q., Kahnt, T., Dogan, A., Strang, S., Fehr, E., & Tobler, P. N. (2017). A neural link between generosity and happiness. *Nature Communications, 8*(1), 15964.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavioural Research Methods, 40*(3), 879-891.

Rodrigues, J., Ulrich, N., & Hewig, J. (2015). A neural signature of fairness in altruism: a game of theta? *Social Neuroscience, 10*(2), 192-205.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science, 300*(5626), 1755-1758.

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012). Enhancing social ability by stimulating right temporoparietal junction. *Current Biology, 22*(23), 2274-2277.

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2015). Functional lateralization of temporoparietal junction - imitation inhibition, visual perspective-taking and theory of mind. *European Journal of Neuroscience, 42*(8), 2527-2533.

Saturnino, G. B., Puonti, O., Nielsen, J. D., Antonenko, D., Madsen, K. H., & Thielscher, A. (2019). SimNIBS 2.1: a comprehensive pipeline for individualized electric field modelling for transcranial brain stimulation. In S. Makarov, M. Horner, & G. Noetscher (Eds.), *Brain and human body modeling* (pp. 3-25).

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking peopleThe role of the temporo-parietal junction in "theory of mind". *Neuroimage, 19*(4), 1835-1842.

Schneider, K., Pauly, K. D., Gossen, A., Mevissen, L., Michel, T. M., Gur, R. C., . . . Habel, U. (2013). Neural correlates of moral reasoning in autism spectrum disorder. *Social Cognitive and Affective Neuroscience, 8*(6), 702-710.

Schurz, M., Aichhorn, M., Martin, A., & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Frontiers in Human Neuroscience, 7*, 712.

Seymour, R. A., Wang, H., Rippon, G., & Kessler, K. (2018). Oscillatory networks of high-level mental alignment: A perspective-taking MEG study. *Neuroimage, 177*, 98-107.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13*, 290–312.

Sobel, M. E. (2008). Identification of Causal Parameters in Randomized Studies With Mediating Variables. *Journal of Educational and Behavioral Statistics, 33*(2), 230-251.

Soutschek, A., Moisa, M., Ruff, C. C., & Tobler, P. N. (2021). Frontopolar theta oscillations link metacognition with prospective decision making. *Nature Communications, 12*(1), 3943.

Soutschek, A., Nadporozhskaia, L., & Christian, P. (2022). Brain stimulation over dorsomedial prefrontal cortex modulates effort-based decision making. *Cognitive, Affective, & Behavioral Neuroscience, 22*(6), 1264-1274.

Soutschek, A., Ruff, C. C., Strombach, T., Kalenscher, T., & Tobler, P. N. (2016). Brain stimulation reveals crucial role of overcoming self-centeredness in self-control. *Science Advances, 2(10)*, e1600992.

Speer, S. P. H., & Boksem, M. A. S. (2019). Decoding fairness motivations from multivariate brain activity patterns. *Social Cognitive and Affective Neuroscience, 14*(11), 1197-1207.

Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B., & Sack, A. T. (2015). Be nice if you have to--the neurobiological roots of strategic fairness. *Social Cognitive and Affective Neuroscience, 10*(6), 790-796.

Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, II, Tobler, P. N., & Kalenscher, T. (2015). Social discounting involves modulation of neural value signals by temporoparietal junction. *Proceedings of the National Academy of Sciences of the United States of America, 112*(5), 1619-1624.

Symeonidou, I., Dumontheil, I., Chow, W. Y., & Breheny, R. (2016). Development of online use of theory of mind during adolescence: An eye-tracking study. *Journal of Experimental Child Psychology, 149*, 81-97.

Tamnes, C. K., Overbye, K., Ferschmann, L., Fjell, A. M., Walhovd, K. B., Blakemore, S. J., & Dumontheil, I. (2018). Social perspective taking is associated with self-reported prosocial behavior and regional cortical thickness across adolescence. *Developmental Psychology, 54*(9), 1745-1757.

Underwood, B., & Moore, B. (1982). Perspective-taking and altruism. *Psychological Bulletin, 91*(1), 143–173.

van de Vijver, I., Ridderinkhof, K. R., & Cohen, M. X. (2011). Frontal Oscillatory Dynamics Predict Feedback Learning and Action Adjustment. *Journal of Cognitive Neuroscience, 23*(12), 4106-4121.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping, 30*(3), 829-858.

Vosskuhl, J., Huster, R. J., & Herrmann, C. S. (2016). BOLD signal effects of transcranial alternating current stimulation (tACS) in the alpha range: A concurrent tACS-fMRI study. *Neuroimage, 140*, 118-125.

Wang, H., Callaghan, E., Gooding-Williams, G., McAllister, C., & Kessler, K. (2016). Rhythm makes the world go round: An MEG-TMS study on the role of right TPJ theta oscillations in embodied perspective taking. *Cortex, 75*, 68-81.

Will, G. J., Crone, E. A., & Guroglu, B. (2015). Acting on social exclusion: neural correlates of punishment and forgiveness of excluders. *Social Cognitive and Affective Neuroscience, 10*(2), 209-218.

Yamagata, T., Nakayama, Y., Tanji, J., & Hoshi, E. (2012). Distinct information representation and processing for goal-directed behavior in the dorsolateral and ventrolateral prefrontal cortex and the dorsal premotor cortex. *Journal of Neuroscience, 32*(37), 12934-12949.

Yamagishi, T., Takagishi, H., Fermin Ade, S., Kanai, R., Li, Y., & Matsumoto, Y. (2016). Cortical thickness of the dorsolateral prefrontal cortex predicts strategic choices in

economic games. *Proceedings of the National Academy of Sciences of the United States of America, 113*(20), 5582-5587.

Zhang, Y., Chen, S., Hu, X., & Mai, X. (2019). Increasing the Difference in Decision Making for Oneself and for Others by Stimulating the Right Temporoparietal Junction. *Frontiers in Psychology, 10*, 185.

Table 1. Results of Bayesian GLMM for the dictator game in the rTPJ experiment. We report the upper and lower borders of the 95% HDI of the posterior distributions. Standard errors of the mean are in brackets.

Predictor	Estimate (SE)	2.5%	97.5%
Intercept	5.25 (0.95)	3.53	7.23
$tACS_{theta\text{-}sham}$	0.67 (0.43)	-0.13	1.57
$tACS_{beta\text{-}sham}$	0.78 (0.44)	-0.07	1.65
$Inequity_{absolute}$	-1.32 (0.38)	-2.08	-0.58
Efficiency	4.14 (0.78)	2.71	5.77
$Inequity_{type}$	-2.65 (0.89)	-4.49	-0.87
discomfort	-0.22 (0.21)	-0.66	0.18
$Inequity_{absolute} \times Inequity_{type}$	-2.15 (0.58)	-3.34	-1.05
$Inequity_{absolute} \times Efficiency$	-0.31 (0.37)	-1.07	0.37
$Efficiency \times Inequity_{type}$	4.13 (1.09)	2.25	6.42
$tACS_{theta\text{-}sham} \times Inequity_{type}$	-0.45 (0.54)	-1.55	0.58
$tACS_{beta\text{-}sham} \times Inequity_{type}$	0.01 (0.56)	-1.12	1.09
$tACS_{theta\text{-}sham} \times Inequity_{absolute}$	-0.73 (0.38)	-1.48	-0.04
$tACS_{beta\text{-}sham} \times Inequity_{absolute}$	-0.50 (0.35)	-1.18	0.16
$tACS_{theta\text{-}sham} \times Efficiency$	0.70 (0.51)	-0.27	1.77
$tACS_{beta\text{-}sham} \times Efficiency$	0.73 (0.51)	-0.25	1.71
$tACS_{theta\text{-}sham} \times Inequity_{absolute} \times Inequity_{type}$	0.58 (0.54)	-0.47	1.65
$tACS_{beta\text{-}sham} \times Inequity_{absolute} \times Inequity_{type}$	0.43 (0.54)	-0.61	1.52
$tACS_{theta\text{-}sham} \times Efficiency \times Inequity_{type}$	0.02 (0.76)	-1.47	1.47
$tACS_{beta\text{-}sham} \times Efficiency \times Inequity_{type}$	0.46 (0.80)	-1.11	2.02
$tACS_{theta\text{-}sham} \times Inequity_{absolute} \times Efficiency$	-0.27 (0.51)	-1.28	0.71
$tACS_{beta\text{-}sham} \times Inequity_{absolute} \times Efficiency$	-0.62 (0.53)	-1.69	0.40
$tACS_{theta\text{-}sham} \times Inequity_{absolute} \times Efficiency \times Inequity_{type}$	-0.32 (0.75)	-1.77	1.16
$tACS_{beta\text{-}sham} \times Inequity_{absolute} \times Efficiency \times Inequity_{type}$	0.35 (0.80)	-1.30	1.89

Table 2. Results of Bayesian GLMM for the dictator game in the rLPFC experiment. We report the upper and lower borders of the 95% HDI of the posterior distributions. Standard errors of the mean are in brackets.

Predictor	Estimate (SE)	2.5%	97.5%
Intercept	6.71 (1.25)	4.52	9.40
$tACS_{theta\text{-}sham}$	0.36 (0.41)	-0.40	1.18
$tACS_{beta\text{-}sham}$	-0.03 (0.42)	-0.86	0.80
$Inequity_{absolute}$	-1.57 (0.46)	-2.46	-0.64
Efficiency	4.06 (0.88)	2.53	5.97
$Inequity_{type}$	-4.56 (1.37)	-7.36	-1.99
discomfort	-0.29 (0.33)	-0.92	0.37
$Inequity_{absolute} \times Inequity_{type}$	-2.21 (0.72)	-3.74	-0.92
$Inequity_{absolute} \times Efficiency$	-0.55 (0.52)	-1.59	0.47
$Efficiency \times Inequity_{type}$	4.59 (1.30)	2.34	7.35
$tACS_{theta\text{-}sham} \times Inequity_{type}$	-0.14 (0.51)	-1.14	0.86
$tACS_{beta\text{-}sham} \times Inequity_{type}$	-0.39 (0.49)	-1.35	0.61
$tACS_{theta\text{-}sham} \times Inequity_{absolute}$	0.38 (0.41)	-0.40	1.23
$tACS_{beta\text{-}sham} \times Inequity_{absolute}$	0.23 (0.38)	-0.49	0.99
$tACS_{theta\text{-}sham} \times Efficiency$	-0.68 (0.51)	-1.70	0.30
$tACS_{beta\text{-}sham} \times Efficiency$	-0.57 (0.49)	-1.56	0.35
$tACS_{theta\text{-}sham} \times Inequity_{absolute} \times Inequity_{type}$	-0.66 (0.59)	-1.82	0.48
$tACS_{beta\text{-}sham} \times Inequity_{absolute} \times Inequity_{type}$	-0.17 (0.54)	-1.23	0.86
$tACS_{theta\text{-}sham} \times Efficiency \times Inequity_{type}$	1.67 (0.73)	0.22	3.13
$tACS_{beta\text{-}sham} \times Efficiency \times Inequity_{type}$	1.28 (0.75)	-0.15	2.90
$tACS_{theta\text{-}sham} \times Inequity_{absolute} \times Efficiency$	-0.42 (0.62)	-1.61	0.81
$tACS_{beta\text{-}sham} \times Inequity_{absolute} \times Efficiency$	0.45 (0.56)	-0.61	1.58
$tACS_{theta\text{-}sham} \times Inequity_{absolute} \times Efficiency \times Inequity_{type}$	1.49 (0.80)	-0.08	3.09
$tACS_{beta\text{-}sham} \times Inequity_{absolute} \times Efficiency \times Inequity_{type}$	0.03 (0.73)	-1.36	1.48

Table 3. Results of Bayesian GLMM for the director task in the rTPJ experiment. We report the upper and lower borders of the 95% HDI of the posterior distributions. Standard errors of the mean are in brackets.

Predictor	Estimate (SE)	2.5%	97.5%
Intercept	3.62 (0.31)	3.04	4.31
$tACS_{theta\text{-}sham}$	-0.02 (0.44)	-0.85	0.90
$tACS_{beta\text{-}sham}$	-0.33 (0.43)	-1.15	0.59
Condition	-2.21 (0.63)	-3.37	-0.94
discomfort	-0.10 (0.20)	-0.49	0.31
$tACS_{theta\text{-}sham} \times$ Condition	1.56 (0.51)	0.55	2.54
$tACS_{beta\text{-}sham} \times$ Condition	0.53 (0.54)	-0.50	1.58

Table 4. Results of Bayesian GLMM for the director task in the rTJP experiment comparing theta with beta tACS. We report the upper and lower borders of the 95% HDI of the posterior distributions. Standard errors of the mean are in brackets.

Predictor	Estimate (SE)	2.5%	97.5%
Intercept	3.52 (0.34)	2.93	4.25
$tACS_{theta\text{-}beta}$	-0.26 (0.40)	-1.02	0.55
Condition	-0.32 (0.70)	-1.69	1.11
discomfort	0.05 (0.22)	-0.36	0.50
$tACS_{theta\text{-}beta} \times$ Condition	-1.28 (0.51)	-2.28	-0.23

Table 5. Results of Bayesian GLMM for the director task in the rLPFC experiment. We report the upper and lower borders of the 95% HDI of the posterior distributions. Standard errors of the mean are in brackets.

Predictor	Estimate (SE)	2.5%	97.5%
Intercept	3.56 (0.35)	2.93	4.29
$tACS_{theta\text{-}sham}$	-0.15 (0.41)	-0.95	0.64
$tACS_{beta\text{-}sham}$	-0.22 (0.40)	-0.99	0.54
Condition	-1.31 (0.73)	-2.71	0.11
discomfort	-0.20 (0.20)	-0.61	0.18
$tACS_{theta\text{-}sham} \times$ Condition	-0.03 (0.51)	-1.06	1.00
$tACS_{beta\text{-}sham} \times$ Condition	0.46 (0.55)	-0.60	1.52

Table 6. Results of Bayesian GLMM for the director task in the rLPFC experiment comparing theta with beta tACS. We report the upper and lower borders of the 95% HDI of the posterior distributions. Standard errors of the mean are in brackets.

Predictor	Estimate (SE)	2.5%	97.5%
Intercept	3.31 (0.35)	2.68	4.05
$tACS_{theta\text{-}beta}$	0.05 (0.43)	-0.75	0.98
Condition	-1.38 (0.66)	-2.64	-0.03
discomfort	-0.23 (0.31)	-0.91	0.34
$tACS_{theta\text{-}beta} \times$ Condition	0.41 (0.49)	-0.54	1.42

4. Chapter III: The causal role of medial prefrontal cortex for updating of mental model representations in social interactions

Manuscript in preparation, Christian, P., Kaiser, J., Taylor, P., George, M., Schütz-Bosbach, S. & Soutschek, A.

Author contributions: PC, PT, SSB, and AS designed research; PC and MG conducted research; PC, JK, and MG analysed data; PC and AS drafted manuscript.

The causal role of medial prefrontal cortex for updating of mental model representations in social interactions

Patricia Christian[1,2], Jakob Kaiser[1], Paul Taylor[1,2] , Michelle George[1], Simone Schütz-Bosbach[1,2], and Alexander Soutschek[1,2]

[1] Department of Psychology, Ludwig Maximilians University Munich, Munich, Germany

[2] Graduate School of Systemic Neurosciences, Ludwig Maximilians University Munich, Munich, Germany

Author contributions: PC, PT, SSB, and AS designed research; PC and MG conducted research; PC, JK, and MG analysed data; PC and AS drafted manuscript; all authors contributed to final manuscript version

Correspondence address:

Patricia Christian
Department of Psychology
Ludwig Maximilians University Munich
Leopoldstr. 13
80802 Munich, Germany
Email: patricia.christian@psy.lmu.de

Abstract

In competitive interactions, humans have to flexibly update their beliefs about another person's intentions in order to adjust their own choice strategy, such as when believing that the other may exploit their cooperativeness. Here we investigate both the neural dynamics and the causal neural substrate of belief updating processes. We used an adapted prisoner's dilemma task in which participants explicitly predicted the co-player's actions, which allowed us to quantify the prediction error between expected and actual behaviour. First, in a EEG experiment we found a stronger medial frontal negativity (MFN) for negative than positive prediction errors, suggesting that this medial-frontal ERP component may encode unexpected defection of the co-player. The MFN also predicted subsequent belief updating after negative prediction errors. In a second experiment we used transcranial magnetic stimulation (TMS) to investigate whether the dorsomedial prefrontal cortex (dmPFC) causally implements belief updating after unexpected outcomes. Our results show that dmPFC TMS impaired belief updating and strategic behavioural adjustments after negative prediction errors. Taken together, our findings reveal the time-course of the use of prediction errors in social decisions, and suggest that the dmPFC plays a crucial role in updating mental representations of others' intentions.

Key words: Electroencephalography (EEG), Medial-Frontal Negativity (MFN), transcranial magnetic stimulation (TMS), dorsomedial prefrontal cortex (dmPFC), Prisoner's Dilemma Game, Belief Updating, Prediction Errors

Significance statement

For successful social interactions, humans must be able to reliably predict their interaction partners' actions. Previous research has linked this capacity mainly to the temporo-parietal junction. Here, we show that the dorsomedial prefrontal cortex also plays a causal role for belief updating in social interactions: Perturbing the dorsomedial prefrontal cortex with brain stimulation impaired the ability to modify expectations about the interaction partner's next actions based on past experiences and to adjust one's choice behaviour in accordance with these updated expectations. Our findings highlight the role of belief updating for strategic social interactions, and identify the dorsomedial prefrontal cortex and its underlying neural dynamics as neural substrate of the ability to successfully learn others' strategies.

Introduction

In social interactions humans need to flexibly adapt their behavioural strategy to their partner's intentions. For example, we may not help a colleague at work if we believe that the colleague will not return this favour when we request support. Mental representations of others' intentions allow humans to compare the expected behaviour of the other with what actually arises, and to adjust their expectations based on the results of this comparison (Stallen & Sanfey, 2013). Even though our decisions are guided by our predictions about others' behaviour, it is still debated which precise neural mechanisms are involved in adjusting beliefs about others' intentions.

The dorsomedial prefrontal cortex (dmPFC) and temporo-parietal junction (TPJ) have been related to mentalizing processes that allow inferring others' intentions (Burnett & Blakemore, 2009; Gallagher & Frith, 2003). These regions were linked to predictions of others' behaviour (Frith & Frith, 2010; Hampton, Bossaerts, & O'Doherty, 2008; Rilling & Sanfey, 2011) but also to the strength of prediction errors, that is the mismatch between others' predicted and observed behaviour (Behrens, Hunt, Woolrich, & Rushworth, 2008; Hampton et al., 2008). In particular, disrupting TPJ activation impaired the updating of mental models in dmPFC (Hill et al., 2017), suggesting that dmPFC may represent and adjust beliefs about others' intentions based on prediction error signals provided by TPJ. Unlike with the TPJ, however, the dmPFC's causal role in mediating the influence of unexpected social outcomes on strategy adjustments remains unknown.

Preliminary evidence for such a role of dmPFC is provided by different lines of correlational research: First, dmPFC is sensitive to predictions errors about others' actions in social interactions (Dungan, Stepanovic, & Young, 2016), particularly when the other free-rides (i.e., unilaterally defects) (Bitsch, Berger, Nagels, Falkenberg, & Straube, 2018; Hertz et al., 2017). Such prediction errors might be reflected by an early event-related potential, the medial frontal negativity (MFN) (Billeke, Zamorano, Cosmelli, & Aboitiz, 2013; Martin, Potts,

Burton, & Montague, 2009), which encodes a prediction error signal in a similar way to dmPFC (Martin et al., 2009). Despite the evidence for a role of the MFN in representing reward prediction errors in the non-social domain (Holroyd & Coles, 2002), in social interactions so far the MFN has been linked to negative social events like unfair outcomes per se, rather than specifically expectation violations (Billeke et al., 2013; Boksem & De Cremer, 2010; Fernandes et al., 2019; Van der Veen & Sahibdin, 2011). Because it has never directly been tested whether the MFN encodes social prediction errors, we analysed the MFN to reveal the time course of the detection of unexpected negative events in social interactions.

A further knowledge gap is whether the detection of prediction errors by dmPFC leads to an updating of mental models of others' intentions, which in turn may affect an agent's own decision strategy. Again, there is correlative evidence that the dmPFC is involved in representing others' mental states (Hill et al., 2017; Nicolle et al., 2012; Zhu, Mathewson, & Hsu, 2012), which forms the basis for predictions about others' behaviour (Kang, Lee, Sul, & Kim, 2013). In particular, enhanced dmPFC activity was linked to the updating of mental model representations (Haroush & Williams, 2015; Nicolle et al., 2012) as well as the adaption of behavioural strategies according to the updated model (Suzuki et al., 2012). However, the available correlative evidence does not allow concluding that dmPFC is indeed causally relevant for the updating of mental model representations and the adjustment of choice strategies following unexpected outcomes.

To address these issues, we conducted two independent studies with EEG and transcranial magnetic stimulation (TMS) in which participants played an adapted version of the prisoner's dilemma game. In the EEG experiment, we tested whether the MFN reflects unexpected defective behaviour and predicts subsequent belief updating. Based on this, we tested in the second experiment whether dmPFC perturbation with TMS impairs belief updating and behavioural adjustments after unexpected defection.

Materials and Methods

Participants

For the EEG study, we tested 35 healthy volunteers (M_{age} = 23,5 years, SD_{age} = 2,75 years, range 18-35 years). We excluded data from three participants who did not believe the cover story that the co-player was a human (see below). The EEG data of two further participants were lost due to technical issues, leaving 30 participants for the statistical analyses (12 male,18 female). For the TMS study, we tested 21 new volunteers (M_{age} = 21,5 years, SD_{age} = 3,25 years, range 18-35 years); we excluded data from one participant due to lack of task understanding, lowering the sample to 20 participants (8 male, 12 female). An apriori power analysis based on the effects size of Cohen's d = 0.86 reported in a meta-analysis on the impact of TMS on social interactions suggests that 17 participants are sufficient to detect significant effects (alpha = 5%) with a power of 90%. (Christian & Soutschek, 2022). Participants were recruited at the Ludwig Maximilian University (LMU) Munich. We included only participants without any known psychiatric or neurological disorders, and participants in the TMS study were moreover screened for counterindications to TMS. They were all naive with respect to the aims of the study. Ethical approval was granted by the ethics committee of the psychology department at the LMU Munich. All participants gave written informed consent prior to participation in the study. Participants received a show-up fee of 10 euro/hour as well as additional earnings depending on the outcome in the prisoner's dilemma game (see below).

Task design

Participants played a version of the prisoner's dilemma game (PDG) which required the participant to repeatedly choose whether to cooperate or defect with the same anonymous interaction partner (Axelrod & Hamilton, 1981; Rilling et al., 2002). So that we could control strategy, the co-player was in fact a computer: so that participants engaged fully, they were told it was a human (see below). During the experiment, the participant and the co-player (computer

algorithm) made their decisions simultaneously, such that outcome-maximising decision making required reliable predictions about the other's next actions. We measured participants' predictions about the co-players' choices: Participants had to indicate their belief whether the other would cooperate or defect on a continuous rating scale from 0 to 20, within a 4 second time-window prior to each choice. After that, participants had to decide within 4 seconds whether to cooperate or defect with the other player by using the left and right arrow keys (for the options presented on the left and right screen side, respectively) on a standard keyboard. At the end of each trial, participants received feedback for 1 second about both their own payoff and that of the co-player, from which participants could then infer what the co-player had chosen (Figure 1A). Based on the payoff matrix (Kapetaniou, Deroy, & Soutschek, 2023), both players gained 4 virtual coins each for mutual cooperation and 2 coins each for mutual defection; free-riders (i.e., players who unilaterally defected) gained 7 coins while the exploited player gained 1 coin (Figure 1B). Comparing the predicted with the actual choice of the co-player allowed us to quantify the degree to which participants either overestimated (negative prediction error) or underestimated their co-player's cooperativeness (positive prediction error). The co-player's choices were determined by a computer algorithm that varied between a tit-for-tat, cooperative, and defective strategy. This procedure was crucial for our study goals as it induced a sufficient number of positive and negative prediction errors that required participants to adjust their own behaviour to the co-player's strategy changes. On tit-for-tat trials (which represented 60% of all trials), the algorithm adopted the participants' choice to cooperate or defect in the previous trial N-1 with a mean probability of 70%. In cooperative trials (20% of all trials), the algorithm cooperated with a mean probability of 70% irrespective of the participant's choice, whereas in defective trials (20% of all trials) the computer chose defection with a mean probability of 70%. The algorithm switched between these strategies after 10-15 trials to make the co-player behaviour less predictable without making it appear random.

Experimental Procedure

Participants took part in one experimental session for the EEG study, whereas the TMS study involved two testing sessions where TMS was administered (in counterbalanced order) either over dmPFC or the vertex as control site (within-subject crossover design). As a cover story, participants were told that they would play the prisoner's dilemma game against an anonymous other person sitting in a separate experimental cabin. To increase the credibility of the cover story, the experimenter moved between the rooms of the participant and the co-player. In the TMS study, participants were told that they would play with different co-players in the two testing sessions to minimize learning effects. The participants could not see, but hear the co-player in the room next door, which was in fact a confederate of the experimenter (Soutschek, Weinreich, & Schubert, 2018). The experimental task lasted for approximately 20 minutes and included a total of 100 trials of the prisoner's dilemma task. The order of the blocks (where the algorithm used either a tit-for-tat or a more cooperative or defective strategy) was pseudo-randomized within the experiment and counterbalanced between sessions. At the end of the experiment we assessed the credibility of the cover story by asking participants whether they believed they had played against a human.

EEG protocol

Continuous EEG data were recorded from 65 active electrodes (actiCAP system; Brain Products, Munich, Germany) and one additional ground electrode, in accordance with the international 10-20 system. For data acquisition all electrodes were referenced to FCz. EEG data were recorded with a Brain Products QuickAmp amplifier, employing a 500 Hz sampling rate. The impedance of all electrodes was kept below 25 KΩ to ensure good signal to noise ratio.

TMS protocol

TMS was administered using a 75 mm outer diameter figure-8 coil (MCF-B65) with MagPro X100 biphasic stimulator (MagVenture, Alpharetta, GA, USA). The experiment included two TMS conditions: dmPFC versus vertex TMS (Figure 1C). The dmPFC site (MNI coordinates: X = -9; Y = 41, Z = 40) was determined based on a previous imaging study on strategic social interactions (Hill et al., 2017). We defined the dmPFC stimulation coordinates for each participant based on their individual structural T1 images and warped the target coordinates into the space of the individual T1 scan using inverse normalization with trilinear interpolation as implemented in SPM12 (Wellcome Trust Centre for Neuroimaging). The vertex was defined as the point at the midline over the central sulcus based on each participant's T1 scan (Soutschek, Moisa, Ruff, & Tobler, 2020; Soutschek & Tobler, 2020). For both stimulation targets, the coil was held tangentially to the skull, parallel to the midline with the handle pointing backwards. We used neuronavigation software (Brainsight, Rogue Research, Montreal, Canada) to determine and monitor coil placement. The individual motor threshold for each participant was obtained by administering single-pulse TMS over the motor cortex (coil was placed over M1 with the handle pointing backwards and perpendicular to the precentral gyrus). The resting motor threshold was defined as the lowest stimulus intensity that induced contractions of the index finger in at least five of ten pulses while the subject rested their hands. For determination of the active motor threshold the same procedure was applied while participants were instructed to exert constant pressure between the index finger and the thumb with 20% of their maximum strength (Groppa et al., 2012; Rossini et al., 2015). We applied a continuous theta-burst protocol with 80% of the active motor threshold for 40 seconds with continuous trains of 600 pulses in bursts of three pulses at 50 Hz, repeated at intervals of 5 Hz (200 ms), which is thought to disrupt cortical excitability at the stimulation site for at least 30 min (Huang, Edwards, Rounis, Bhatia, & Rothwell, 2005). Directly after the stimulation we asked participants to indicate how aversive they experienced the stimulation (based on a 7-point

Likert scale) to control for individual differences in perceived aversiveness of dmPFC versus vertex TMS.

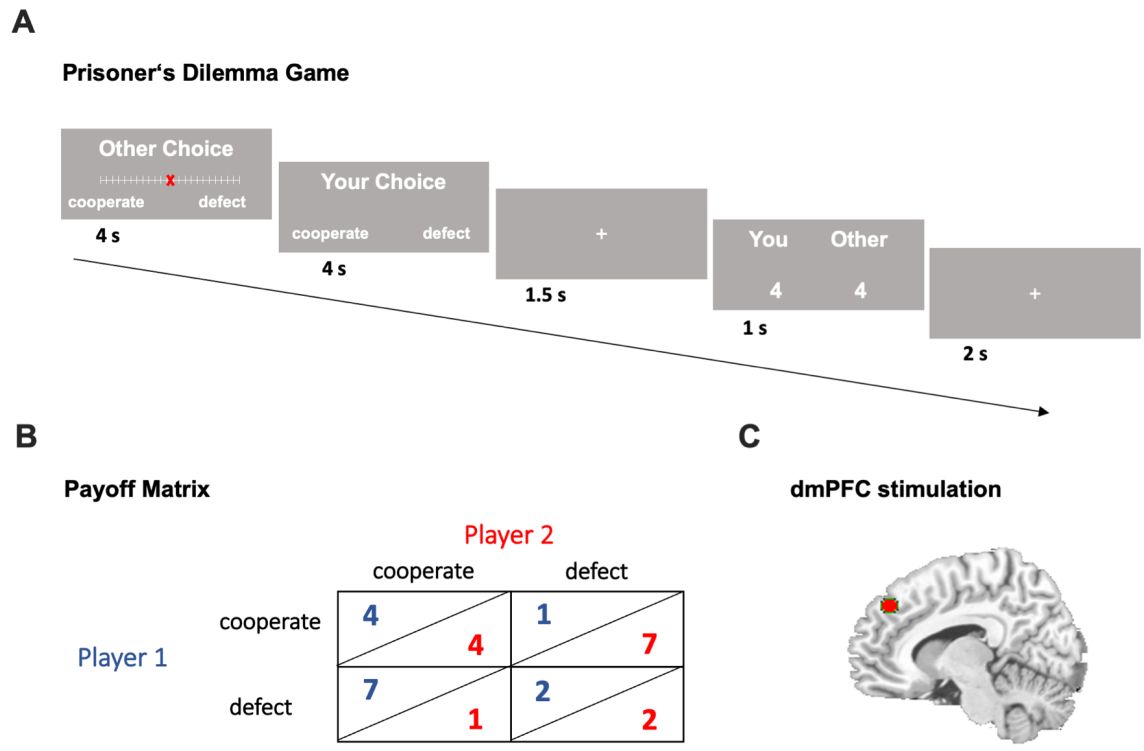

Figure 1. (A) Example trial of the prisoner's dilemma game in the EEG and TMS study: At the beginning of each trial, participants were asked to predict whether the other player would cooperate or defect. Next, participants decided to cooperate or defect. At the outcome stage participants were informed about the payoff for themselves and the co-player, which allowed participants to infer whether their prediction was correct or incorrect (B) Payoff Matrix: Players obtained 4 coins in case of mutual cooperation and 2 coins for mutual defection, whereas unilateral cooperation and defection yielded 1 anf 7 coins, respectively. (C) Illustration of dmPFC TMS site modeled with MRIcon based on the MNI coordinates for dmPFC cTBS stimulation in an example participant.

Statistical analysis

Behavioural analysis. In both the EEG and TMS experiments, we computed linear mixed models (LMMs) using the lme4 package in the R (version 3.6.3.) statistical software environment (Bates, Mächler, Bolker, & Walker, 2015). As dependent variable, we analysed

trial-by-trial changes in continuous predictions ($prediction_{trial\ N}$ minus $prediction_{trial\ N-1}$): Positive values indicated that participants considered it as more likely that the other would cooperate on the current compared with the previous trial, whereas negative values reflected an increased subjective likelihood that the other would defect compared with the previous trial. In the EEG study, we regressed continuous prediction changes on fixed-effect predictors for the predicted choice of the co-player (Predicted choice; 0 = cooperate, 1 = defect), the absolute (unsigned) prediction error ($PE_{absolute}$), the sign (direction) of the prediction error (PE_{sign}; -1 = negative, 1 = positive), and all interaction terms. In the TMS study, we added fixed-effect predictors for TMS (0 = vertex, 1 = dmPFC) to the model. All fixed effects were also modelled as random slopes in addition to participant-specific random intercepts. The absolute prediction error ($PE_{absolute}$) was defined as the absolute difference between the other's choice and participants' prediction ratings on the previous trial (N-1) as a measure of the magnitude of the prediction error, whereas the sign of the prediction error (PE_{sign}) indicated its direction (with positive and negative prediction errors reflecting unexpected defection or unexpected cooperation, respectively). The variable Predicted choice indicated the predicted behaviour of the co-player on the previous trial. We also included discomfort ratings as a predictor of no interest to control for potential confounding effects of TMS-induced discomfort.

Furthermore, we analyzed effects on choice behaviour with generalized linear mixed models (GLMMs) where binary choices (cooperate = 0, defect = 1) were regressed on fixed-effect predictors for TMS, $PE_{absolute}$, PE_{sign}, Prediction change, and all interaction terms. We included the variable Prediction change in the model to test whether trial-by-trial changes in predictions moderated TMS effects on choices. Again, all predictors were also modelled as random slopes in addition to participant-specific intercepts.

EEG analysis.

Preprocessing of EEG data was performed using the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). EEG data were filtered using a 50 Hz low-pass filter and 0.5 Hz high-pass filter within the recommended guidelines to avoid distortion of the time course of MFN waveform (Luck & Gaspelin, 2017). Data were re-referenced offline to the average of both mastoids. Noisy electrodes were removed and reintegrated with spherical spline interpolations applying the Fieldtrip function ft_channelrepair (Oostenveld et al., 2011). We segmented the data in epochs ranging from -1500 ms to +2000 ms relative to the feedback window and baseline- corrected with the baseline interval defined from −200 ms to 0 ms. We performed independent component analysis (ICA) for artefact removal of eye blinks, horizontal eye movements, muscle movements and high skin potentials (Delorme, Sejnowski, & Makeig, 2007; Mennes, Wouters, Vanrumste, Lagae, & Stiers, 2010). Finally, we excluded trials with missing data due to technical issues during EEG recording.

To examine whether the MFN component reflects prediction errors, we recorded the MFN component at the medial prefrontal electrodes sites in the feedback window between 200 – 400 ms in line with previous findings (Billeke et al., 2013; Boksem & De Cremer, 2010; Campanha, Minati, Fregni, & Boggio, 2011; Wang et al., 2022). The electrode positions were predefined (FCz, FC1, FC2, Fz, F1, F2) based on previous studies on the MFN (Campanha et al., 2011; Wang et al., 2022; Wu, Hu, van Dijk, Leliveld, & Zhou, 2012) to reduce bias towards statistical significance (Luck & Gaspelin, 2017). To test our hypothesis whether negative prediction errors are associated with more negative MFN amplitudes than positive prediction errors, we calculated the difference between the average feedback-locked ERPs on trials for negative and positive prediction errors with permutation tests based on cluster statistics, as implemented in the Fieldtrip function ft_freqstatistics (Maris & Oostenveld, 2007). Based on a Monte Carlo randomization procedure, the p-values for each cluster were estimated to compute the significance probability. Average feedback-locked ERP data for each participant were randomly shuffled between the conditions for 2000 iterations. The cluster candidates with the

highest summed values were compared against the permutation distribution for each of these permutations. Differences between negative and positive prediction errors were considered as significant if the p-value calculated for the largest cluster-level statistic was smaller than the critical alpha-level (0.05). For each significant cluster, we report the cluster weight, p-value, and its start and end time.

Results

Mediofrontal negativity signals unexpected defection (EEG study)

As a sanity check, we first assessed whether participants updated their predictions of the co-player's actions and adjusted their own choice strategy following prediction errors. As expected, participants indeed more strongly expected the co-player to cooperate or defect after positive (unexpected cooperation) or negative (unexpected defection) prediction errors, respectively, PE_{sign}: $\beta = 0.06$, $z = 4.26$, $p < 0.001$. Furthermore, participants updated their beliefs more strongly after larger prediction errors, $PE_{absolute}$: $\beta = 0.20$, $z = 2.62$, $p < 0.01$, with larger positive and negative prediction errors being associated with stronger updating of the expectation that the co-player would cooperate or defect, respectively, $PE_{sign} \times PE_{absolute}$: $\beta = 0.23$, $z = 13.41$, $p < 0.001$ (Figure 2A, Table 1). These changes in predictions were also associated with adjustments of choice behavior: the results of the GLMM on choices revealed that participants chose to defect more often after stronger negative prediction errors, while participants were more likely to cooperate following larger positive prediction errors, $PE_{sign} \times PE_{absolute}$: $\beta = 1.41$, $z = 6.30$, $p < 0.01$. Moreover, choice behavior was predicted by changes in prediction, Prediction change: $\beta = -1.76$, $z = -6.36$, $p < 0.001$. Specifically, participants chose to defect more often the stronger they updated their prediction that the co-player would defect after negative prediction errors, while participants cooperated more often the more they updated their predictions towards cooperation after positive prediction errors, $PE_{sign} \times$ Prediction change: $\beta = -0.24$, $z = -2.80$, $p < 0.01$ (Figure 2B, Table 2). Thus, participants both updated

their expectations about the co-player's actions and adjusted their behavioural strategy in response to mismatches between predicted and actual choices of the co-player.

To test whether negative prediction errors are linked with more pronounced negative MFN amplitudes than positive prediction errors, we calculated permutation tests based on cluster statistics. This analysis revealed stronger higher negative amplitudes for negative relative to positive prediction errors at medial prefrontal electrodes between 240 and 340 ms (t_mass = -76.36, $p = 0.003$; Figure 2C). To exclude the alternative explanation that the effects could be driven by negative social outcomes per se (Billeke et al., 2013; Boksem & De Cremer, 2010; Campanha et al., 2011; Polezzi, Sartori, Rumiati, Vidotto, & Daum, 2010; Wang et al., 2022) rather then prediction errors about the other's decisions, we calculated the difference between the average feedback-locked ERPs for CD (co-player unilaterally defected) trials and CC (mutual cooperation) trials. The results of the cluster-based permutation analysis showed no significant clusters at medial prefrontal electrodes between 200 and 400 ms for CD in contrast to CC trials. Thus, our findings suggest that the MFN component encodes negative prediction errors (i.e., unexpected defection) rather than negative outcomes per se.

Based on the MFN's role in encoding prediction errors, we next asked whether larger MFN amplitudes for negative compared with positive prediction errors are linked with belief updating on the behavioural level. For this purpose we calculated a Spearman rank correlation between individual MFN amplitudes and trial-by-trial changes in predictions. We extracted the individual coefficients for the intercept from the linear mixed model on prediction changes and calculated the relative distribution of the MFN ($MFN_{rel} = MFN_{negative} - MFN_{positive} / ((MFN_{negative} + MFN_{positive}) / 2)$) to quantify MFN amplitudes for negative relative to positive prediction errors. The MFN components were extracted based on the results of the cluster-based permutation analysis of average feedback-locked ERPs. More negative MFN amplitudes for negative in contrast to positive prediction errors were correlated with stronger prediction changes to defection (negative values for Prediction change), Spearman's $r_s = 0.32$, $p = 0.049$

(Figure 2D). To sum up, our findings suggest that this ERP component recorded at electrodes over dmPFC is linked with mismatched predictions about others' actions in cooperative-competitive contexts, with the MFN being more sensitive to unexpected defection than unexpected cooperation. The MFN then reflects when people have been treated badly, but only if this was unexpected – and furthermore predicts whether participants proceed on the basis of this negative experience to update their expectations about the other's cooperativeness.

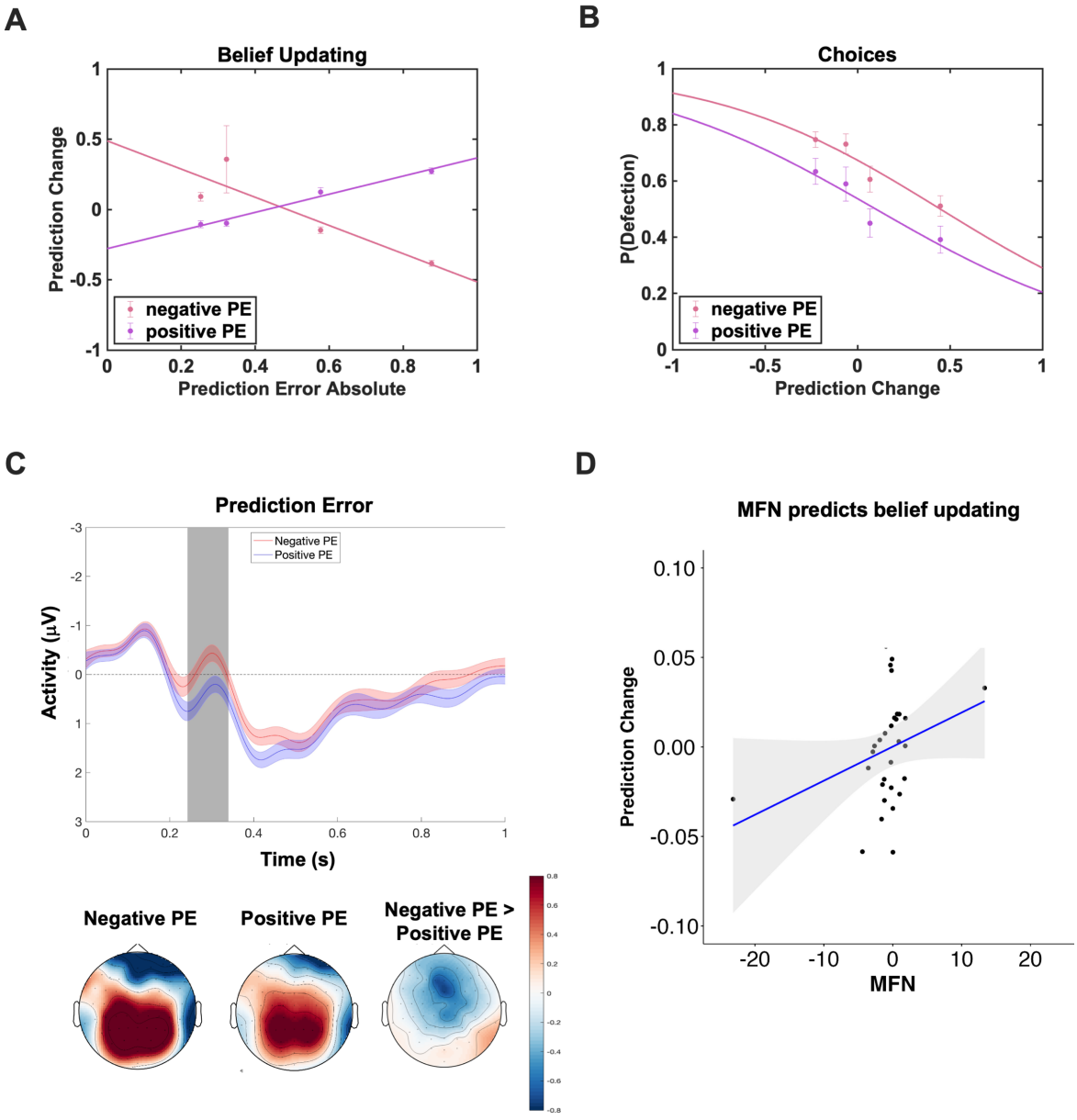

Figure 2. Behavioural and ERP Results of the EEG Experiment. (A) Higher negative values indicate a prediction change to defection (increased expectation that the co-player will defect) while more positive values indicate a stronger prediction change to cooperation: Participants updated their beliefs that the co-player would defect more strongly following negative prediction errors (unexpected defection) than following positive prediction errors (unexpected cooperation). (B) After negative prediction errors participants showed a stronger tendency to defect in contrast to positive prediction errors, driven by prediction changes that the co-player is more likely to defect. (C) Results of cluster-based permutation analysis for event-related potentials show that the medial frontal negativity (MFN) is more pronounced for negative in contrast to positive prediction errors between 240 and 340 ms. Topographic maps displaying the contrast between negative and positive prediction errors at medio-prefrontal electrode sites. (D) Correlation between MFN and prediction change: Larger negative MFN values were linked with stronger updating of expectations that the co-player will defect.

dmPFC TMS impairs belief updating after negative prediction errors (TMS study)

While our EEG findings suggest that activity recorded from medial prefrontal electrodes is sensitive to prediction errors and is associated with belief updating, the correlative nature of these findings leave open whether medial prefrontal activity causally contributes to belief updating and adjustments of one's choice behaviour in response to negative prediction errors. We therefore conducted a second experiment assessing whether dmPFC disruption with TMS interferes with these processes. If the dmPFC is causally involved in moderating the influence of unexpected defection on strategic social decisions, we expected TMS over dmPFC to reduce the impact of negative prediction errors on belief updating, impairing the ability to adjust one's own choice behaviour. To test this hypothesis, we regressed trial-by-trial changes in predictions on fixed-effect predictors for TMS, $PE_{absolute}$, PE_{sign}, Predicted choice, and all interaction terms, controlling for TMS-induced discomfort. The significant $PE_{absolute} \times PE_{sign}$ interaction, $\beta = 0.20$, $z = 6.46$, $p < 0.001$, suggested that participants updated the predicted likelihood that the co-player would cooperate or defect depending on whether the other co-player unexpectedly cooperated or defected, respectively, replicating our behavioural findings in the EEG study. This updating of beliefs after prediction errors was significantly reduced after dmPFC

compared with vertex TMS: TMS × $PE_{absolute}$ × PE_{sign} , β = -0.09, z = -2.19, p = 0.04. Moreover, the significant four-way TMS × $PE_{absolute}$ × PE_{sign} × Predicted choice interaction, β = 0.18, z = 4.01, p < 0.001, suggested the impact of dmPFC TMS on belief updating after prediction errors depended on whether participants had predicted the co-player to cooperate or defect (Table 3). To distinguish between the TMS effects on positive versus negative prediction errors, we calculated separate post-hoc analyses for negative and positive prediction errors. We split these models depending on participants' predictions of the other's behaviour (Predicted choice) to differentiate between expected or unexpected cooperation as well as expected or unexpected defection. The results show that dmPFC TMS relative to vertex TMS reduced the extent to which participants would normally defect more after unexpected defection, TMS × $PE_{absolute}$: β = 0.06, z = 2.57, p = 0.02 (Figure 3A, Table 3). In contrast, there were no significant stimulation effects following negative prediction errors when the participant had expected the co-player to defect, TMS × $PE_{absolute}$: β = -0.02, z = -1.41, p = 0.18. The post-hoc tests for positive prediction errors revealed that with increasing positive prediction errors (unexpected cooperation) participants more strongly updated the expectation that the co-player would cooperate under dmPFC in contrast to vertex TMS, TMS × $PE_{absolute}$: β = 0.08, z = 2.62, p = 0.02 (Figure 3B, Table 3), whereas there were no significant effects after expected cooperation, TMS × $PE_{absolute}$: β = -0.04, z = -1.55, p = 0.14. Thus, dmPFC TMS strengthened the updating of the belief that the other would cooperate after unexpected cooperation, whereas after unexpected defection participants were less likely to update their beliefs towards defection under dmPFC TMS. In both situations participants therefore became more likely to predict that their co-player would cooperate after dmPFC TMS. Together, this provides causal evidence for the hypothesized role of dmPFC for belief updating.

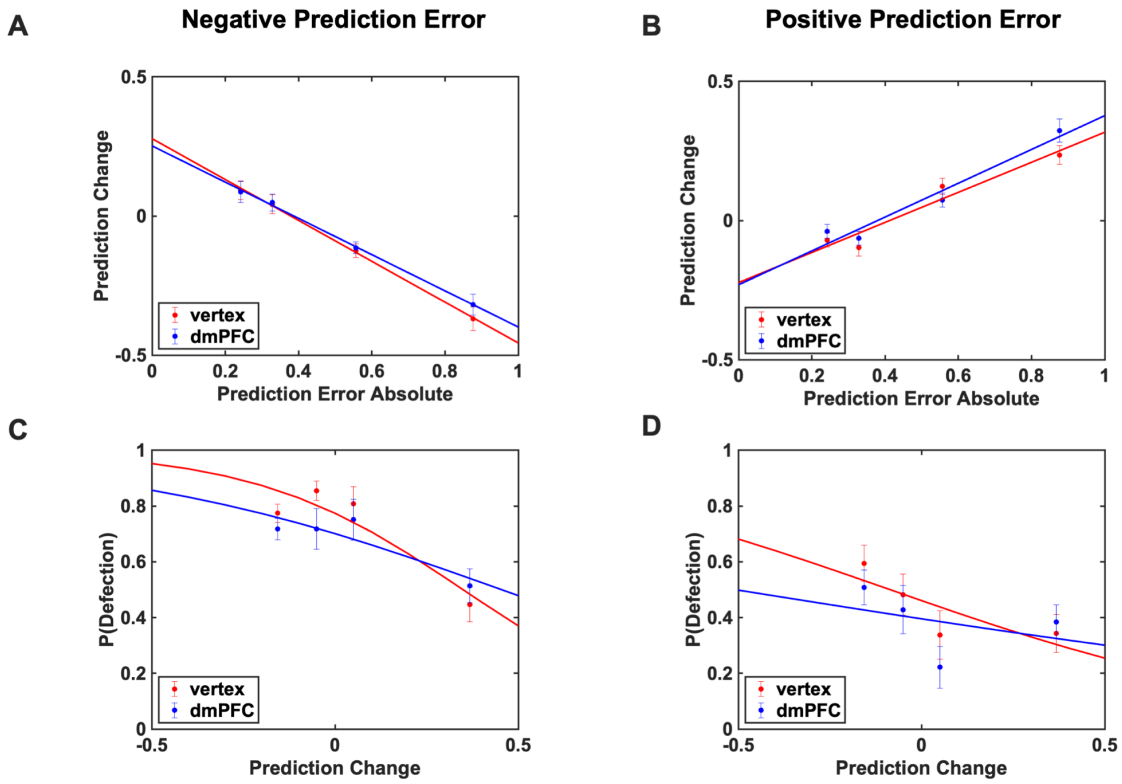

Figure 3. Results of TMS effects on belief updating (A,B) and choices (C,D) in the prisoner's dilemma game. (A) Higher negative values indicate a prediction change to defection while more positive values indicate a stronger prediction change to cooperation. dmPFC TMS reduced prediction changes to defection after negative prediction errors compared to vertex TMS: participants were less likely to change their prediction to defection when the co-player unexpectedly defected. (B) Under dmPFC TMS compared with vertex TMS, participants more strongly predicted the co-player to cooperate after positive prediction errors, i.e. when the co-player unexpectedly cooperated. (C, D) dmPFC TMS in comparison to vertex TMS more strongly decreased defection rates when participants updated the expectation that the co-player would defect after (C) negative prediction errors compared with (D) positive prediction errors.

Impact of dmPFC TMS on strategic decision making is moderated by stimulation effects on belief updating

Based on the dmPFC's causal role for belief updating after prediction errors, we asked whether TMS-induced impairments in belief updating also affect the ability to flexibly adjust one's choice behaviour to the co-player's actions. We expected that dmPFC TMS reduces

behavioural adjustments after unexpected defection, particularly when participants failed to update their expectation that the co-player would defect. To test this hypothesis, we regressed binary choices (cooperate = 0, defect = 1) on fixed-effect predictors for TMS, $PE_{absolute}$, PE_{sign}, prediction change, and all interaction terms. Participants were more likely to defect when changing their predictions towards the expectation that the other will defect, Prediction change: $\beta = -3.22$, $z = -4.83$, $p < 0.001$, as well as after negative compared with positive prediction errors, PE_{sign}: $\beta = -0.36$, $z = -2.25$, $p = 0.03$, with the latter effect being more pronounced the more strongly participants updated their beliefs after prediction errors, PE_{sign} × Prediction change: $\beta = -0.47$, $z = -2.08$, $p = 0.04$. Importantly, this interaction effect was significantly reduced under dmPFC compared with vertex TMS, TMS × PE_{sign} × Prediction change, $\beta = 0.75$, $z = 2.21$, $p = 0.03$, suggesting that dmPFC disruption reduced the preference for defection after negative prediction errors as a function of how strongly participants updated their beliefs (Figure 3C/D, Table 4). That is, under sham TMS people are more likely to defect if they predict the co-player to defect after unexpected defection, and dmPFC TMS weakened this relationship. In addition, participants cooperated more often after larger absolute prediction errors, $PE_{absolute}$: $\beta = -0.47$, $z = -2.08$, $p = 0.04$, and this effect was enhanced under dmPFC TMS in contrast to vertex TMS, TMS × $PE_{absolute}$: $\beta = -0.58$, $z = -2.05$, $p = 0.04$, such that after larger prediction errors (independently of the direction of the prediction error) participants cooperated more often under dmPFC TMS in contrast to vertex TMS. Taken together, our results suggest that dmPFC TMS impaired not only belief updating but also adjustments of choice behaviour in response to prediction errors. This provides causal evidence for the dmPFC's hypothesized role in moderating the influence of mental model updatings on choice behaviour in social interactions.

Discussion

The dmPFC has been ascribed a central role for decision making in social interactions, but its precise functional role remained unclear. Here, we provide converging EEG and TMS evidence for dmPFC involvement in updating mental models about others' intentions and actions based upon these updated representations. We show that the MFN ERP component is a marker signaling unexpected defection and predicts whether participants will update their expectations about the other's intentions. Our findings show that dmPFC TMS made participants both less likely to defect and to predict that their co-player would defect, which suggests that the dmPFC is indeed causally involved in updating mental model representations and adjusting choices in response to negative prediction errors.

We found that the MFN, a medial prefrontal ERP component, reflects negative prediction errors (and more so than positive prediction errors). While in the literature on social interactions the MFN has been linked to negative events like unfairness or free-riding behaviour (Boksem & De Cremer, 2010; Miraghaie et al., 2022; Wang et al., 2022), we show that the MFN does not encode negative events per se but rather only if they are worse than expected. Even though the MFN has been interpreted as a neuronal marker of violated social expectancies (Billeke et al., 2013; Chen, Zhao, & Lai, 2019; Hu & Mai, 2021; Wang et al., 2022), to the best of our knowledge it has never directly been assessed whether mismatches between expectations about the others' social strategy and the actual outcome drive the MFN component in social interactions: previous studies did not explicitly measure participants' predictions about the co-player's actions. Our interpretation is in line with previous findings from non-social reward prediction errors where the MFN was reported to be associated with unexpected outcomes (Martin & Potts, 2011; Soder & Potts, 2018). Thus, our results demonstrate that the MFN is driven by negative prediction errors rather than negative outcomes in social interactions. Our findings extend our understanding of the chronometry and neural dynamics of social decision making by revealing an early neuronal correlate of social negative prediction errors which predicts subsequent belief updating.

While we must be cautious with drawing inferences from the MFN findings to the dmPFC's role in social interactions (as another region than dmPFC might in theory be the source of the MFN signal), our TMS findings provide direct evidence that the dmPFC plays a causal role for strategic social decision making. Our findings suggest that dmPFC causally contributes to integrating and updating beliefs about others' intentions, enabling decision makers to flexibly adjust their behaviour after unexpected actions of their interaction partners. This extends previous correlational findings which showed that dmPFC activity is enhanced during social interactions requiring representations of others' mental states (Andrews-Hanna, Reidler, Sepulcre, Poulin, & Buckner, 2010; Behrens, Hunt, & Rushworth, 2009; Li, Mai, & Liu, 2014; Sul & Kim, 2021). Our results suggest that dmPFC does not only encode such mental models of others' intentions but contributes to updating the the mental model if the predictions from the model mismatch the actual behaviour of the interaction partner. More specifically, after negative prediction errors (unexpected defection), decision makers less strongly updated their expectation that the co-player would defect after dmPFC TMS compared with vertex TMS. Interestingly, after positive prediction errors participants more strongly expected the co-player to cooperate following dmPFC perturbation. Thus, while dmPFC TMS reduced the impact of unexpected defection on belief updating, it increased (rather than impaired) belief updating after positive prediction errors. This pattern is consistent with the MFN's increased sensitivity for unexpected negative compared with positive outcomes and suggests that interfering with dmPFC activity might amplify the positive evaluation of both unexpected defection and cooperation. If positive prediction errors are represented by reduced dmPFC activity, then experimentally lowering dmPFC activity with TMS might strengthen neural representations of positive prediction errors. In any case, we provide conclusive evidence that the dmPFC is causally relevant for belief updating in response to prediction errors in social interactions.

Crucially, dmPFC TMS disturbed not only the ability to update beliefs about the co-player's intentions but also the ability to act upon these mental model representations. It has been proposed that the dmPFC represents others' mental states during strategic social interactions (Andrews-Hanna et al., 2010; Behrens et al., 2009; Li et al., 2014; Sul & Kim, 2021). Our findings provide direct evidence that the dmPFC plays a causal role in promoting flexible behavioural adjustment based on updated mental models of others' intentions in response to unexpected events. The influence of dmPFC TMS on strategic decision making was moderated by the degree of belief updating after prediction errors, linking our stimulation effects on belief updating and on decision making. The dmPFC may then not only update mental models about others' intentions but may forward these representations to other brain regions involved in strategic social decisions, such as ventromedial prefrontal cortex (Hill et al., 2017) or dorsolateral prefrontal cortex (Soutschek, Sauter, & Schubert, 2015), enabling these regions to compute the optimal choice based on these mental representations.

The dmPFC is likely to implement belief updating processes not in isolation but in interaction with other parts of the mentalizing network, including the TPJ. A large body of evidence links the TPJ to simulate the mental states of others (Saxe & Wexler, 2005; Schuwerk, Grosso, & Taylor, 2021; Van Overwalle & Baetens, 2009), but the TPJ was also found to show enhanced activation for mismatches between others' predicted and observed actions (Carter, Bowling, Reeck, & Huettel, 2012; Park, Fareri, Delgado, & Young, 2021). These prediction error signals in the TPJ might be forwarded to the dmPFC, as suggested by a study showing that TPJ inhibition reduces its connectivity with dmPFC during social interactions (Hill et al., 2017). The dmPFC in turn might use these prediction error signals to update the mental model of others' intentions.

To sum up, our findings ascribe the dmPFC a causal role for updating beliefs about others' intentions after unexpected defection and for adjusting choice behaviour according to

these updated beliefs. This clarifies the dmPFC's function in strategic social interactions, assigning it a central role for flexible behavioural adjustments when interacting with others.

Data availability statement

The data that support the findings of this study will be available on Open Science Framework (https://osf.io/pjen2/).

Funding:

AS received an Emmy Noether fellowship (SO 1636/2-1) from the German Research Foundation, a research grant from the Hetzler foundation, and an Exploration grant from the Boehringer Ingelheim Foundation. PT received a grant (TA 857/3-2) from the German Research Foundation.

Ethics

The study protocol was approved by the Ethics Committee of the Department of Psychology at the LMU Munich. All volunteers gave written informed consent before participating in the study.

Declaration of competing interest

The authors declared that there were no conflicts of interest in relation to the subject of this study.

References

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron, 65*(4), 550-562. doi:10.1016/j.neuron.2010.02.005

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*, 1390 - 1396. doi:10.1126/science.74663

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. *Journal of Statistical Software, 67*(1). doi:10.18637/jss.v067.i01

Behrens, T. E., Hunt, L. T., & Rushworth, M. F. (2009). The Computation of Social Behavior. *Science, 324*(5931), 1160-1164. doi:10.1126/science.116969

Behrens, T. E., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. (2008). Associative learning of social value. *Nature, 456*(7219), 245-249. doi:10.1038/nature07538

Billeke, P., Zamorano, F., Cosmelli, D., & Aboitiz, F. (2013). Oscillatory brain activity correlates with risk perception and predicts social decisions. *Cerebral Cortex, 23*(12), 2872-2883. doi:10.1093/cercor/bhs269

Bitsch, F., Berger, P., Nagels, A., Falkenberg, I., & Straube, B. (2018). The role of the right temporo-parietal junction in social decision-making. *Human Brain Mapping, 39*(7), 3072-3085. doi:10.1002/hbm.24061

Boksem, M. A., & De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Social Neuroscience, 5*(1), 118-128. doi:10.1080/17470910903202666

Burnett, S., & Blakemore, S. J. (2009). Functional connectivity during a social emotion task in adolescents and in adults. *European Journal of Neuroscience, 29*(6), 1294-1301. doi:10.1111/j.1460-9568.2009.06674.x

Campanha, C., Minati, L., Fregni, F., & Boggio, P. S. (2011). Responding to unfair offers made by a friend: neuroelectrical activity changes in the anterior medial prefrontal cortex. *Journal of Neuroscience, 31*(43), 15569-15574. doi:10.1523/JNEUROSCI.1253-11.2011

Carter, R. M., Bowling, D. L., Reeck, C., & Huettel, S. A. (2012). A distinct role of the temporal-parietal junction in predicting socially guided decisions. *Science, 337*(6090), 109-111. doi:10.1126/science.1219681

Chen, M., Zhao, Z., & Lai, H. (2019). The time course of neural responses to social versus non-social unfairness in the ultimatum game. *Social Neuroscience, 14*(4), 409-419. doi:10.1080/17470919.2018.1486736

Christian, P., & Soutschek, A. (2022). Causal role of right dorsolateral prefrontal cortex for norm-guided social decision making: A meta-analysis of TMS studies. *Neuropsychologia, 176*, 108393. doi:10.1016/j.neuropsychologia.2022.108393

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage, 34*(4), 1443-1449. doi:10.1016/j.neuroimage.2006.11.004

Dungan, J. A., Stepanovic, M., & Young, L. (2016). Theory of mind for processing unexpected events across contexts. *Social Cognitive and Affective Neuroscience, 11*(8), 1183-1192. doi:10.1093/scan/nsw032

Fernandes, C., Goncalves, A. R., Pasion, R., Ferreira-Santos, F., Barbosa, F., Martins, I. P., & Marques-Teixeira, J. (2019). Age-related changes in social decision-making: An electrophysiological analysis of unfairness evaluation in the Ultimatum Game. *Neuroscience Letters, 692*, 122-126. doi:10.1016/j.neulet.2018.10.061

Frith, U., & Frith, C. D. (2010). The social brain: allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1537), 165-176. doi:10.1098/rstb.2009.0160

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences, 7*(2), 77-83. doi:10.1016/s1364-6613(02)00025-6

Groppa, S., Oliviero, A., Eisen, A., Quartarone, A., Cohen, L. G., Mall, V., . . . Siebner, H. R. (2012). A practical guide to diagnostic transcranial magnetic stimulation: report of an IFCN committee. *Clinical Neurophysiology, 123*(5), 858-882. doi:10.1016/j.clinph.2012.01.010

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences of the United States of America, 105*(18), 6741-6746. doi:10.1073/pnas.0711099105

Haroush, K., & Williams, Z. M. (2015). Neuronal prediction of opponent's behavior during cooperative social interchange in primates. *Cell, 160*(6), 1233-1245. doi:10.1016/j.cell.2015.01.045

Hertz, U., Palminteri, S., Brunetti, S., Olesen, C., Frith, C. D., & Bahrami, B. (2017). Neural computations underpinning the strategic management of influence in advice giving. *Nature Communication, 8*(1), 2191. doi:10.1038/s41467-017-02314-5

Hill, C. A., Suzuki, S., Polania, R., Moisa, M., O'Doherty, J. P., & Ruff, C. C. (2017). A causal account of the brain network computations underlying strategic social behavior. *Nature Neuroscience, 20*(8), 1142-1149. doi:10.1038/nn.4602

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*(4), 679-709. doi:10.1037/0033-295X.109.4.679

Hu, X., & Mai, X. (2021). Social value orientation modulates fairness processing during social decision-making: evidence from behavior and brain potentials. *Social Cognitive and Affective Neuroscience, 16*(7), 670-682. doi:10.1093/scan/nsab032

Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron, 45*(2), 201-206. doi:10.1016/j.neuron.2004.12.033

Kang, P., Lee, J., Sul, S., & Kim, H. (2013). Dorsomedial prefrontal cortex activity predicts the accuracy in estimating others' preferences. *Frontiers in Human Neuroscience, 7*, 686. doi:10.3389/fnhum.2013.00686

Kapetaniou, G. E., Deroy, O., & Soutschek, A. (2023). Social metacognition drives willingness to commit. *Journal of Experimental Psychology: General*. doi:10.1037/xge0001419

Li, W., Mai, X., & Liu, C. (2014). The default mode network and social understanding of others: what do brain connectivity studies tell us. *Frontiers in Human Neuroscience, 8*, 74. doi:10.3389/fnhum.2014.00074

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*(1), 146-157. doi:10.1111/psyp.12639

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177-190. doi:10.1016/j.jneumeth.2007.03.024

Martin, L. E., & Potts, G. F. (2011). Medial frontal event-related potentials and reward prediction: do responses matter? *Brain and Cognition, 77*(1), 128-134. doi:10.1016/j.bandc.2011.04.001

Martin, L. E., Potts, G. F., Burton, P. C., & Montague, P. R. (2009). Electrophysiological and hemodynamic responses to reward prediction violation. *Neuroreport, 20*(13), 1140-1143. doi:10.1097/WNR.0b013e32832f0dca

Mennes, M., Wouters, H., Vanrumste, B., Lagae, L., & Stiers, P. (2010). Validation of ICA as a tool to remove eye movement artifacts from EEG/ERP. *Psychophysiology*. doi:10.1111/j.1469-8986.2010.01015.x

Miraghaie, A. M., Pouretemad, H., Villa, A. E. P., Mazaheri, M. A., Khosrowabadi, R., & Lintas, A. (2022). Electrophysiological Markers of Fairness and Selfishness Revealed by a Combination of Dictator and Ultimatum Games. *Frontiers in Systems Neuroscience, 16*, 765720. doi:10.3389/fnsys.2022.765720

Nicolle, A., Klein-Flugge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron, 75*(6), 1114-1121. doi:10.1016/j.neuron.2012.07.023

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computer Intelligence and Neuroscience, 2011*, 156869. doi:10.1155/2011/156869

Park, B., Fareri, D., Delgado, M., & Young, L. (2021). The role of right temporoparietal junction in processing social prediction error across relationship contexts. *Social Cognitive and Affective Neuroscience, 16*(8), 772-781. doi:10.1093/scan/nsaa072

Polezzi, D., Sartori, G., Rumiati, R., Vidotto, G., & Daum, I. (2010). Brain correlates of risky decision-making. *Neuroimage, 49*(2), 1886-1894. doi:10.1016/j.neuroimage.2009.08.068

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron, 35*(18), 395–405.

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology, 62*, 23-48. doi:10.1146/annurev.psych.121208.131647

Rossini, P. M., Burke, D., Chen, R., Cohen, L. G., Daskalakis, Z., Di Iorio, R., . . . Ziemann, U. (2015). Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: Basic principles and procedures for routine clinical and research application. An updated report from an I.F.C.N. Committee. *Clinical Neurophysiology, 126*(6), 1071-1107. doi:10.1016/j.clinph.2015.02.001

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia, 43*(10), 1391-1399. doi:10.1016/j.neuropsychologia.2005.02.013

Schuwerk, T., Grosso, S. S., & Taylor, P. C. J. (2021). The influence of TMS of the rTPJ on attentional control and mentalizing. *Neuropsychologia, 162*, 108054. doi:10.1016/j.neuropsychologia.2021.108054

Soder, H. E., & Potts, G. F. (2018). Medial frontal cortex response to unexpected motivationally salient outcomes. *International Journal of Psychophysiology, 132*(Pt B), 268-276. doi:10.1016/j.ijpsycho.2017.11.003

Soutschek, A., Moisa, M., Ruff, C. C., & Tobler, P. N. (2020). The right temporoparietal junction enables delay of gratification by allowing decision makers to focus on future events. *PLOS Biology, 18*(8), e3000800. doi:10.1371/journal.pbio.3000800

Soutschek, A., Sauter, M., & Schubert, T. (2015). The Importance of the Lateral Prefrontal Cortex for Strategic Decision Making in the Prisoner's Dilemma. *Cognitive Affective & Behavioural Neuroscience, 15*(4), 854-860. doi:10.3758/s13415-015-0372-5

Soutschek, A., & Tobler, P. N. (2020). Causal role of lateral prefrontal cortex in mental effort and fatigue. *Human Brain Mapping, 41*(16), 4630-4640. doi:10.1002/hbm.25146

Soutschek, A., Weinreich, A., & Schubert, T. (2018). Facial electromyography reveals dissociable affective responses in social and non-social cooperation. *Motivation and Emotion, 42*(1), 118-125. doi:10.1007/s11031-017-9662-2

Stallen, M., & Sanfey, A. G. (2013). The cooperative brain. *Neuroscientist, 19*(3), 292-303. doi:10.1177/1073858412469728

Sul, S., & Kim, M. J. (2021). Human dorsomedial prefrontal cortex delineates the self and other against the tendency to form interdependent social representations. *Neuron, 109*(14), 2209-2211. doi:10.1016/j.neuron.2021.06.029

Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., Haruno, M., . . . Nakahara, H. (2012). Learning to simulate others' decisions. *Neuron, 74*(6), 1125-1137. doi:10.1016/j.neuron.2012.04.030

Van der Veen, F. M., & Sahibdin, P. P. (2011). Dissociation between medial frontal negativity and cardiac responses in the ultimatum game: Effects of offer size and fairness. *Cognitive Affective & Behavioural Neuroscience, 11*(4), 516-525. doi:10.3758/s13415-011-0050-1

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage, 48*(3), 564-584. doi:10.1016/j.neuroimage.2009.06.009

Wang, A., Zhu, L., Lyu, D., Cai, D., Ma, Q., & Jin, J. (2022). You are excusable! Neural correlates of economic neediness on empathic concern and fairness perception. *Cognitive Affective & Behavioural Neuroscience, 22*(1), 99-111. doi:10.3758/s13415-021-00934-5

Wu, Y., Hu, J., van Dijk, E., Leliveld, M. C., & Zhou, X. (2012). Brain activity in fairness consideration during asset distribution: does the initial ownership play a role? *PLoS One, 7*(6), e39627. doi:10.1371/journal.pone.0039627

Zhu, L., Mathewson, K. E., & Hsu, M. (2012). Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proceedings of the National Academy of Sciences of the United States of America, 109*(5), 1419-1424. doi:10.1073/pnas.1116783109

Tables

Table 1. EEG Experiment: Results of computed linear mixed models (LMMs) for belief updating in the Prisoner's Dilemma Game. Standard errors of the mean are in brackets.

	Beta (SEM)	t	df	p
Intercept	-0.00 (0.02)	-0.10	11	0.92
$PE_{absolute}$	-0.03 (0.02)	-2.03	27	0.05
PE_{sign}	0.06 (0.01)	4.26	23	<0.001
Predicted choice	-0.00 (0.02)	-0.16	9	0.87
$PE_{absolute} \times PE_{sign}$	0.23 (0.02)	13.41	21	<0.001
$PE_{absolute} \times$ Predicted choice	0.03(0.02)	1.30	33	0.20
$PE_{sign} \times$ Predicted choice	-0.00 (0.02)	-0.39	61	0.70
$PE_{absolute} \times PE_{sign} \times$ Predicted choice	-0.04 (0.02)	-1.83	26	0.08

Table 2. EEG Experiment: Results of computed generalized linear mixed models (GLMMs) for choice behaviour in the Prisoner's Dilemma Game. Standard errors of the mean are in brackets.

	Estimate (SEM)	z	p
Intercept	0.60 (0.17)	3.58	<0.001
$PE_{absolute}$	0.32 (0.08)	4.10	<0.001
PE_{sign}	-0.02 (0.13)	-0.16	0.87
Prediction change	-1.77 (0.28)	-6.36	<0.001
$PE_{absolute} \times PE_{sign}$	1.41 (0.02)	0.22	<0.001
$PE_{absolute} \times$ Prediction change	-0.05 (0.06)	-0.89	0.37
$PE_{sign} \times$ Prediction change	-0.24 (0.09)	-2.80	0.01
$PE_{absolute} \times PE_{sign} \times$ Prediction change	-0.07 (0.07)	-1.04	0.30

Table 3. TMS Experiment: Results of computed linear mixed models (LMMs) for belief updating in the Prisoner's Dilemma Game. Standard errors of the mean are in brackets.

	Beta (SEM)	t	df	p
Intercept	-0.03 (0.04)	-0.81	6	0.45
$PE_{absolute}$	-0.05 (0.02)	-2.76	54	0.01
PE_{sign}	0.04 (0.02)	1.86	10	0.09
TMS	-0.03 (0.02)	-0.86	7	0.42
Predicted choice	0.01 (0.04)	0.36	7	0.73
Discomfort	-0.03 (0.02)	-1.91	3	0.15
$PE_{absolute} \times PE_{sign}$	0.20 (0.03)	6.46	14	<0.001
$PE_{absolute} \times$ TMS	0.01 (0.03)	0.23	25	0.82
$PE_{signed} \times$ TMS	-0.01 (0.02)	-0.38	34	0.70
$PE_{absolute} \times$ Predicted choice	0.04 (0.03)	1.15	19	0.26
$PE_{signed} \times$ Predicted choice	-0.00 (0.03)	-0.18	11	0.86
TMS × Predicted choice	0.00 (0.05)	0.00	11	0.99
$PE_{absolute} \times PE_{sign} \times$ TMS	-0.09 (0.04)	-2.19	15	0.04
$PE_{absolute} \times PE_{sign} \times$ Predicted choice	-0.03 (0.03)	-0.83	12	0.43
$PE_{absolute} \times$ TMS × Predicted choice	0.04 (0.05)	0.83	16	0.42
$PE_{sign} \times$ TMS × Predicted choice	-0.01 (0.03)	-0.25	20	0.81
$PE_{absolute} \times PE_{sign} \times$ TMS × Predicted choice	0.18 (0.04)	4.01	16	<0.001

Table 4. TMS Experiment: Results of computed generalized linear mixed models (GLMMs) for choice behaviour in the Prisoner's Dilemma Game. Standard errors of the mean are in brackets.

	Estimate (SEM)	z	p
Intercept	0.53 (0.22)	2.36	0.02
$PE_{absolute}$	0.32 (0.20)	1.61	0.11
PE_{sign}	-0.36 (0.16)	-2.25	0.02
TMS	-0.27 (0.20)	-1.40	0.16
Prediction change	-3.22 (0.67)	-4.83	<0.001
discomfort	0.08 (0.16)	0.49	0.62
$PE_{absolute} \times PE_{sign}$	2.46 (0.61)	4.02	<0.001
$PE_{absolute} \times$ TMS	-0.58 (0.28)	-2.05	0.04
$PE_{sign} \times$ TMS	0.01 (0.16)	0.06	0.95
$PE_{absolute} \times$ Prediction change	0.11 (0.18)	0.62	0.53
$PE_{sign} \times$ Prediction change	-0.47 (0.23)	-2.08	0.04
TMS × Prediction change	0.26 (0.36)	0.74	0.46
$PE_{absolute} \times PE_{sign} \times$ TMS	0.19 (0.29)	0.65	0.52
$PE_{absolute} \times PE_{sign} \times$ Prediction change	-0.05 (0.18)	-0.25	0.81
$PE_{absolute} \times$ TMS × Prediction change	-0.18 (0.22)	-0.85	0.39
$PE_{sign} \times$ TMS × Prediction change	0.74 (0.34)	2.21	0.03
$PE_{absolute} \times PE_{sign} \times$ TMS × Prediction change	0.44 (0.28)	1.58	0.11

5. General Discussion

The goal of this dissertation was to investigate the causal roles of the prefrontal cortex (mPFC, rLPFC) and temporo-parietal cortex (rTPJ) in social decision making. Hereby, we examined the precise neuro-computational roles of these regions for implementing decision making depending on the contextual factors which drive our tendency to promote fairness behaviour towards others. We investigated the neuro-cognitive and neurobiological mechanisms underlying social decision making by applying brain stimulation and electrophysiological recordings. Based on the summary of the main findings from each study I will review the most important insights and conclusion. Further, I will discuss the main theoretical and clinical implications across all research projects and consider the limitations of neuroscientific methods.

5.1. Main findings

5.1.1. Causal role of right dorsolateral prefrontal cortex for norm-guided social decision making: a meta-analysis of TMS studies

In this first research project we analysed data from rDLPFC TMS studies across economic games to determine the causal role of the rDLPFC for norm enforcement in a healthy population. We calculated a meta-analysis and subgroup analysis for TMS studies over the rDLPFC studying the behavioural effects for social conflicts between one's own selfish interests and fairness concerns in the ultimatum game, dictator game, trust game, third party punishment and prisoner's dilemma game. Although the causal role of the rDLPFC for implementing decision making during social conflicts has been established, it's precise function is still debated, whether the rDLPFC either enhances norm enforcement or increases the preference to rational, selfish choice options (Baumgartner et al., 2011; Christov-Moore et al., 2017; Knoch et al., 2006; Maier et al., 2018). By applying this meta-analytic approach we were

able to provide an accurate estimate of stimulation effects on social decision making to investigate the crucial role of the rDLPFC across social conflicts in economic games. The distinction between the subgroups allowed us to disentangle the brain stimulation effects depending on the social contextual factors. The subgroup fairness type was defined as either reactive fairness (included social contexts in which threat of punishment was expected as well as norm violators could be punished) or proactive fairness (included economic games which assess pro-social, altruistic preferences; without punishment options). The subgroup role of the responder was defined by the participants role in the economic game as either playing in the role of the proposer or responder.

Our results based on the meta-analysis across all types of economic games showed no significant effects on norm guided choice behaviour while the subgroup analysis revealed that rDLPFC implements fairness-oriented behavior depending on the social context. Importantly, our findings suggest that rDLPFC has no unitary role when being confronted with social conflicts but implements norm enforcement depending on the type of fairness behaviour. More specifically, replicating previous findings we found that rDLPFC enhances norm-guided behavior in reactive fairness contexts, i.e. when deviations from social norms can be sanctionized and participants usually act in the role of the responder (Buckholtz et al., 2008; Cheng et al., 2022; Wu et al., 2014). Interestingly, we found no stimulation effects on proactive fairness suggesting that the rDLPFC has a crucial role for implementing decision making when the threat of social norm violations is present rather then promoting norm compliant behaviour.

Our results provide new insights into the contextual factors that drive the effects of the rDLPFC on normative choice behaviour in social decision making. In particular, our results demonstrate that the rDLPFC implements norm-guided behaviour when undergoing the threat of social punishment while the rDLPFC is not causally involved for prosocial giving. Future research could extend these findings by examining the precise neuro-psychological

mechanisms in the rDLPFC which promote the implementation of normative choices when threat of punishment is present.

5.1.2. Causal roles of prefrontal and temporo-parietal theta oscillations for inequity aversion

In this second research project we investigated whether the rTPJ and rLPFC are involved in aversion to advantageous or disadvantageous inequity in a healthy population. We used non-invasive transcranial alternating current stimulation (tACS) with theta (6 Hz), beta (20 Hz) and sham condition while participants performed the dictator game to examine proactive fairness preferences and the director task to assess accuracy performance in perspective taking. While previous findings show that the rTPJ and rLPFC play key roles in the implementation of pro-social choice behaviour in economic games (Gao et al., 2018; Hutcherson et al., 2015; Morishima et al., 2012; Speer & Boksem, 2019), it still remains unclear whether these brain regions crucially implement inequity aversion per se or modulate the conflict between our own selfish interests and our fairness concerns when oneself or the others are worse off. By using tACS stimulation with either theta, beta or sham stimulation over the rTPJ and rLPFC allowed us to disentangle the causal contribution of these brain regions for proactive fairness preferences (advantageous and disadvantageous inequity) and to investigate the underlying brain rythms reflecting inequity aversion. Additionally, we assessed whether the rTPJ's role for perspective taking promotes inequity aversion reflected by the same or dissociable brain rythms.

Our results demonstrate that the rTPJ and the rLPFC play key roles in implementing inequity aversion to promote pro-social choices, in line with previous findings (Gao et al., 2018; Hutcherson et al., 2015). In contrast to previous accounts which claim that the rTPJ advances advantageous inequity when we are better off than others (Morishima et al., 2012), our results show that the rTPJ implement inequity aversion per se. Further, our result also supports the notion that the rLPFC is functionally relevant for increasing aversion to disadvantageous rather than advantageous inequity which replicates previous findings (Fliessbach et al., 2012). It has

been proposed that the rLPFC has a key role to implement decision making in response to unfairness in social interactions (Baumgartner et al., 2011; Buckholtz et al., 2015). Thus, it is possible that the rLPFC promotes the aversion to disadvantageous inequity based on emotion- and conflict-related processes in response to perceived unfairness. Importantly, our results demonstrate that theta oscillations underlie the aversion to increasing unequal splits in the rTPJ and enhance the preference for welfare maximizing choice outcomes for disadvantageous inequity in the rLPFC. Hereby, our study extend previous findings by demonstrating that prefrontal and parietal theta oscillations underlie inequity aversion.

Further, we replicate previous findings showing that the rTPJ has a a causal role for perspective taking and self-other distinction (Santiesteban, Banissy, Catmur, & Bird, 2012). Altough it has been hypothesized that the rTPJ's role to differentiate between our own and the others perspective promotes pro-social choices (Soutschek et al., 2016; Strombach et al., 2015), to the best of our knowledge our study is the first to show that inequity aversion in the rTPJ's can be explained by it's more general role for perspective taking. Our findings are in line with previous evidence that advantageous and disadvantageous inequity emerge at a developmental stage linked with the maturation of higher-order social cognition such as mentalizing (Ulber, Hamann, & Tomasello, 2017). Thus, our results provide first direct evidence of the precise neuro-psychological mechanisms underlying inequity aversion in the rTPJ.

Here, we demonstrate causal evidence for dissociable roles of theta oscillations in rTPJ and rLPFC for resolving conflicts between selfish and other-regarding motives. Thus, we provide insights how neural oscillations in different brain regions moderate prosocial behavior by implementing dissociable psychological mechanisms. Although our results show that inequity aversion can be explained by the rTPJ's role for perspective taking, future research should address the precise neuro-cognitive mechanisms underlying disadvantageous inequity in the rLPFC.

5.1.3. The causal role of medial prefrontal cortex for updating of mental model representations in competitive interactions

In the third research project we investigated whether the dorsomedial prefrontal cortex (dmPFC) plays a key role for belief updating and behavioural adaption following prediction errors in social cooperation in a healthy population. We used electroencephalography (EEG) and transcranial magnetic stimulation (TMS) for two independent experiments to test whether the dmPFC is involved in belief updating and behavioural adaption in response to unexpected defection in the prisoner's dilemma game. First, we used EEG to examine whether a medial frontal ERP component, the Medial-Frontal Negativity (MFN), reflects mismatched predictions about the others behavioural strategy in social cooperation. Even though the MFN has been interpreted as a neuronal marker of mismatched predictions when social expectancies are violated (Billeke, Zamorano, Cosmelli, & Aboitiz, 2013; Chen, Zhao, & Lai, 2019; Hu & Mai, 2021; Wang et al., 2022) it has never been directly tested whether the MFN reflects prediction errors in social decision making. Based on the limited spatial resolution of EEG data, we used cTBS stimulation over the dmPFC to investigated whether the dmPFC is causally involved in updating of mental model representations in response to unexpected defection when the other deviates from our social expectancies. While previous research proposes that the dmPFC is linked with building up mental model representations about the others behavioural strategy (Haroush & Williams, 2015; Nicolle et al., 2012), it's precise neuro-computational role in cooperative-competitive context still remains unknown.

The result of our EEG study demonstrate that the medial frontal negativity (MFN) is more pronounced in response to violated social expectancies, when the other unexpectedly defected. Our findings extend past research suggesting that the MFN component is reflected by mismatched predictions about the others social strategy rather than driven by outcomes in social

decision making paradigms (Billeke et al., 2014; Boksem & De Cremer, 2010; Campanha, Minati, Fregni, & Boggio, 2011). Thus, our results suggest that the mPFC might be linked with an early neuronal marker sensitive to unexpected defection, when the co-player choose a competitive over a cooperative strategy. This is in line with previous findings which suggest that the mPFC is sensitive to detect predictions errors when simulating the other's actions (Dungan et al., 2016; Lee & Seo, 2016). Overall, our findings provides new insights into the neurobiological response of the medial frontal ERP component (MFN) underlying unexpected defection.

Further, the results of our TMS experiment shed new light on our understanding of the crucial role of the dmPFC in cooperative-competitive social interactions: our results provide direct evidence that the dmPFC is causally involved in belief updating processes in response to negative prediction errors, when the other unexpectedly defected. We extend previous findings by demonstrating that the dmPFC is recruited more strongly when we update our predictions in response to social negative conflicts evoked by the others unexpected competitive strategy (Hertz et al., 2017). Moreover, our results suggest that effects on choice behaviour are driven by belief updating in response to unexpected defection why the disruption of the dmPFC with cTBS resulted in reduced belief updating leading to decreased behavioural adaption. Taken together, our results suggest that dmPFC cTBS impaired adjustments in predictions of others' behavior which affected one's own ability to flexibly adapt accordingly to the others unexpected defection.

Hereby, we extend previous findings demonstrating that the mPFC is sensitive to unexpected social norm violations, when the other defects more strongly than expected. Furthermore, we demonstrate causal evidence that the dmPFC is involved in belief updating following negative prediction errors in competitive- cooperative social interactions. Thus, we

provide insights into the neuro-cognitive mechanisms in the mPFC when our social expectancies are violated.

5.2. Theoretical Implications

In this dissertation I examined the crucial neuro-cognitive mechanisms in social decision making by implementing brain stimulation methods on brain regions typically involved in social cognition and decision making in the social context. Hereby, we provide direct evidence that the temporo-parietal cortex promotes pro-social choices explained by it's key role for social cognition enabling us to integrate the others perspective into our choice behaviour. Further, our findings demonstrate that the rLPFC is crucially relevant to implement norm enforcement by punishment of norm violators in reactive fairness contexts as well as to accept maximizing payoff options to overcome disadvantageous inequity, when we are worse off than others. Finally, our results show that the dmPFC is crucially relevant for our ability to represent and update beliefs about the others anticipated actions when our social expectancies are violated. Thus, our findings improve our understanding of the precise function of the temporo-parietal cortex and prefrontal cortex to guide one's own decision based on our fairness preferences while infering the others intentions, goals and prospective actions.

For many years it has been controversially debated whether the rDLPFC is involved in promoting pro-social behaviour rather than implementing selfish, more rational, payoff maximizing choices in social interactions (Buckholtz et al., 2008; Emonds, Declerck, Boone, Vandervliet, & Parizel, 2011; Fermin et al., 2016; Sanfey et al., 2003). Our findings provide support for this idea that the rDLPFC promotes norm enforcement (Buckholtz et al., 2015; Makwana & Hare, 2012), but extend previous assumptions by demonstrating that the rDLPFC resolves the conflict between our own interests and fairness concerns when violations of social norms can be sanctionized. Thus, our findings support theoretical accounts which claim that

the rDLPFC is crucially involved in punishing norm violators to enforce social norms (Baumgartner et al., 2011; Knoch et al., 2006; Spitzer et al., 2007). However, in contrast to previous findings (Muller-Leinss et al., 2018), our results provide no evidence that the rDLPFC is recruited to promote norm compliance. While past research examined the role of the rDLPFC for social decision making paradigms separately by analysing meta-analytic data across social contexts our findings shed new light on the understanding of the rDLPFC's more general role in social decision making. Thus, our findings provide important steppstones to conceptualize the social contextual factors which can explain the rDLPFC's precise function for norm enforcement.

Even though previous research demonstrates that the rTPJ and rLPFC are involved when we make decisions to reduce unequal outcomes between ourselves and others, our findings showed for the first time that theta oscillations causally implement inequity aversion in the rTPJ and promote the aversion to disadvantageous inequity in the rLPFC. Our results are in line with previous accounts which propose that dissociable brain regions crucially implement advantageous and disadvantageous inequity (Gao et al., 2018; Hutcherson et al., 2015; Morishima et al., 2012). Further, our results demonstrate that the rTPJ promotes the rejection of unequal outcomes which is in line with the hypothesized role of the rTPJ to increase more generous, altruistic choices by concluding the conflict between self- and other-regarding motives (Obeso et al., 2018). However, our findings shed new light on the controversy debate about the rTPJ's role in altruistic choices by providing new evidence that the rTPJ increases inequity aversion per se rather than specifically promoting social preferences for advantageous inequity (Morishima et al., 2012). Importantly, our findings provide direct evidence that the rTPJ's role for higher-order social cognition such as perspective taking explains our tendency to reduce inequity between ourselves and others. Even though past research suggested that our ability to infer the others mental state might moderate advantageous inequity aversion

(McAuliffe et al., 2017), our results demonstrate that the rTPJ crucially enhances the aversion to unequal outcomes per se mediated by the socio-cognitive ability to understand and differentiate our own from the others perspective (Courtney & Meyer, 2020; Quesque & Brass, 2019; Santiesteban et al., 2012). These insights extend our understanding of the precise neuro-cognitive mechanism underlying inequity aversion.

For many years it has been proposed that the MFN component is associated with violation of social expectancies (Alexopoulos, Pfabigan, Goschl, Bauer, & Fischmeister, 2013; Boksem & De Cremer, 2010; Campanha et al., 2011), but no previous account has ever directly tested whether the MFN is driven by prediction errors in the social context. Indeed, our findings demonstrate that the MFN reflects negative prediction errors, when the other defected more strongly than expected. Further, our findings reveal that the MFN predicts subsequent belief updating, thus suggesting that the more pronounced MFN response to negative prediction errors is linked with stronger belief updating that the other will defect. Thus, our results shed new insights into the neuro-cognitive mechanisms underlying our mismatched social expectations guiding our choices in social cooperation.

Although based on previous findings it has been hypothesized that the rTPJ and dmPFC both play key roles in social bargaining game, recent research has predominantly focused on the crucial role of the rTPJ in competitive-cooperative interactions (Hill et al., 2017). Here, we show for the first time that the dmPFC is causally relevant for belief updating in response to unexpected negative prediction errors, when the other defected more strongly than expected. Indeed, our findings provide support for the theory that the dmPFC is a key brain region to generate representation of the others intentions and strategies as well as update these beliefs in response to unexpected outcomes in the social context (Ferrari et al., 2016; Haroush & Williams, 2015; Jamali et al., 2021; Nicolle et al., 2012). Thus, our findings shed new light on our understanding of the neuro-psychological mechanism underlying the behavioural evidence:

Our results demonstrate that the dmPFC is sensitive to and reacts upon negative prediction errors specifically rather than prediction errors per se. Hence, our results could provide support for the theory that the dmPFC is involved when we are affected by social norm violations and perceive the others behaviour towards us as an unfair treatment (Civai, Miniussi, & Rumiati, 2015). Further, our results demonstrate that disruption of belief updating following dmPFC cTBS reduced our behavioural adaption to defect the co-player in response to unexpected defection. Thus, our results are similar to theoretical accounts which propose that our previous mismatched prediction are one of the key predictors to guide our own decision making to either choose a cooperative or competitive strategies (Pisauro, Fouragnan, Arabadzhiyska, Apps, & Philiastides, 2022).

5.3. Clinical implications for neuro-psychiatric disorders

Previous evidence shows that the main symptoms of neuro-psychiatric disorders are deficits in socio-cognitive abilities and social decision making which are linked with dysfunctions in the prefrontal cortex and temporo-parietal brain regions (Billeke et al., 2015; Bitsch, Berger, Nagels, Falkenberg, & Straube, 2019; Frascarelli et al., 2015; Horat et al., 2018; Hu & Mai, 2021). Previous findings suggest a link between positive psychotic symptoms in schizophrenia with temporo-parietal junction signal changes during deceptive repayments in the trust game (Gromann et al., 2013), while reduced activation ot the TPJ was linked with defective strategies in social cooperation in ASD in contrast to healthy controls (Edmiston, Merkle, & Corbett, 2014). Thus, by examining the causal roles of these brain regions with brain stimulation our research findings might help us to further understand the neuro-cognitive and psychological mechanisms underlying these neuro-psychiatric disorders. More specifically, our research findings provide a causal link between these brain regions and their underlying neuro-cognitive mechanisms which might extend our previous knowledge how to modulate social cognition and social decision making in clinical populations.

Previous evidence indicates that the rTPJ and mPFC, which are linked with socio-cognitive abilities such as mentalizing and representation of self and other related interests, are affected in autism spectrum disorder and schizophrenia (Lombardo, Chakrabarti, Bullmore, & Baron-Cohen, 2011; Porcelli et al., 2019; Schneider et al., 2013). In fact, previous findings demonstrate that unfair proposals f.e. in the ultimatum game are accepted more often in patients diagnosed with autism spectrum disorder and schizophrenia than in contrast to healthy controls which suggest that the clinical population is less inequity averse than the healthy controls (Hartley & Fisher, 2018; Tei et al., 2018). These social decision making deficits have been linked with reduced mentalizing abilities which reduce sensitivity to fairness-related social signals to guide one's own choices whether to accept or reject the others proposals (Yang et al., 2017). Our findings support this theoretical account by demonstrating that the tendency to avoid unequal splits in proactive fairness can be explained by differentiation between one's own from the others perspective in a healthy population. Thus, reduced aversion to unequal splits might be explained by socio-cognitive deficits which reduce the ability to self-other distinction and perspective taking in neuropsychiatric disorders. Although, previous research demonstrates that rTPJ dysfunction in neuro-psychiatric disorders is linked with behavioural deficits in social interactions (Eddy, 2016) our research findings provide new insights into the neuronal signature underlying social decision making and higher-order social cognition which might be affected in autism spectrum disorder or schizophrenia. Our findings extend previous research by hinting to a causal role of theta oscillations in the rTPJ enhancing the ability to represent one's own and the others perspective to implement pro-social behaviour in a healthy population. Thus, our recent findings can improve our understanding of the brain rythms underlying social decision making which might be disrupted in neuropsychiatric symptomatology.

During social cooperation reduced cooperativeness and higher rejection rates have been observed in patients diagnosed with schizophrenia (Hanssen et al., 2018). This could be

explained by past research which show that schizophrenia is linked with lower trustworthiness ratings of co-players in interactive economic games (Daan Baas et al., 2008), which is in line with the positive symptomatology in schizophrenia such as paranoia (Gromann et al., 2013). Nevertheless, the precise neuro-psychological processes promoting lower cooperativeness and higher defection rates are still debated. Previous neuroimaging research proposes that reduced activity in the dmPFC has been linked with deficits in higher-order social functioning deficits such as considering the others beliefs and intentions in economic games for schizophrenia (Kronbichler, Tschernegg, Martin, Schurz, & Kronbichler, 2017). As based on our recent findings we suggest that the dmPFC has a causal role in updating represention of the others mental state in repetitive social interactions in a healthy population why dmPFC dysfunction in schizophrenia might affect their ability to make predictions and update beliefs about the others future actions. Therefore, it is possible that schizophrenic patients overestimate the others defectiveness even when the co-player switches to cooperative strategies. Our findings give us great insights that disruption of the dmPFC results in lower belief updating and reduced behavioural adaption why dmPFC dysfunction in schizophrenia which might cause deficits in the ability to flexibly adapt one's own predictions about the social environment. This might promote stronger tendencies in schizophrenia to defect the other. Indeed, past research reveals that schizophrenic patients adapt their behaviour less often linked with reduced activation in the rTPJ-mPFC network (Bitsch et al., 2019). Furthermore, recent findings suggest that reduced neural responses such as decreased amplitudes for the MFN component in response to unfair proposals in schizophrenia can be explained by reduced sensitivity to social cues (Horat et al., 2018). More precisely, based on our EEG findings we suggest that reduced sensitivity to the others unexpected unfairness behaviour is reflected by the decreased amplitude of underlying neural markers such as the MFN component in schizophrenia.

Thus, by providing new insights into the causal link between the prefrontal cortex as well as temporo-parietal cortex in social decision making, we might improve our further understanding of the neural basis of reduced or altered pro-social behaviour in neuro-psychiatric disorders such as autism and schizophrenia. However, the results of the three research projects can only provide novel insights into the neural basis of social decision making in the healthy brain which might improve our understanding of neuro-cognitive mechanisms underlying clinical symptoms in neuro-psychiatric disorders. That's why future research focusing on similar paradigms in clinical populations is necessary for a further understanding of the precise neural mechanisms linked with behavioural deviations in psychiatric disorders. Furthermore, while non-invasive brain stimulation can modulate social behaviour such as social decision making, it might be a promising tool for potential therapeutic applications of neuro-psychiatric disorders (Levasseur-Moreau & Fecteau, 2012). Indeed, brain stimulation methods have been widely used for the treatment of psychiatric diseases such as schizophrenia while their therapeutic effects are still debated (Bhattacharya et al., 2022; George et al., 2009).

5.4. Methodological Considerations and Limitations of Neuroscientific Methods

In this dissertation for three research projects we combined behavioural assessments with neuroscientific methods such as EEG and NIBS (tACS, TMS) to advance our understanding about the neurobiological mechanisms in social decision making. Even though these neurosctientific methods provide information of neural markers underlying neuro-cognitive processes in social decision making linked with the prefrontal cortex and temporo-parietal cortex it is important to discuss their limitiations.

5.4.1. TMS

Further, in our first research project analyzed data across TMS studies which targeted the rDLPFC in social decision making. We included rTMS and cTBS stimulation protocols in

our meta-analysis which are assumed to interfere with ongoing neural activity in the targeted rDLPFC in line with past research (Huang, Edwards, Rounis, Bhatia, & Rothwell, 2005; Polania et al., 2018). Nevertheless, recent findings show inconsistent results whether cTBS can enhance rather than interfere ongoing neural activity depending on the stimulation intensity (Fitzgerald, Fountain, & Daskalakis, 2006; Huang et al., 2005). Overall, brain stimulation is an effective tool to modulate ongoing neural activity but it still needs to be further investigated how exactly parameters in TMS protocols modify neural activity in a specific brain region.

In our third research project we used TMS to investigate the causal role of the dmPFC in social cooperation. One of the major limitations of applying brain stimulation methods is that brain stimulation can cause effects on the targeted brain region as well as on other brain regions of the underlying neural network (Veniero et al., 2019). Thus, the stimulation effect might spread to other brain areas why it still has its limitations to link the observed behavioural effects with the modulated brain area. Previous findings proposed that the rTPJ and mPFC are functionally coupled with each other at the time of the feedback (Baumgartner, Gotte, Gugler, & Fehr, 2012; Burnett & Blakemore, 2009; Hill et al., 2017) why disruption of the rTPJ with rTMS decreased activity in the dmPFC (Hill et al., 2017). Thus, by targeting the dmPFC we primarily interfere with the ongoing neural activation in this focal brain region, however without combining brain stimulation with neuroimaging methods we cannot rule out the possibility that the dmPFC TMS stimulation reduced activity in network related areas such as the rTPJ. Moreover, for our study design we used an offline TMS design which limits our understanding of the precise temporal process underlying our behavioural effects at the feedback stage (Polania et al., 2018). Thus, based on our findings we can only make assumptions that the dmPFC is recruited to update our belief in response to negative prediction errors without any further specifications when the dmPFC might be recruited to update our beliefs at the feedback stage.

5.4.2. tACS

In our second research project we used tACS to entrain either theta, beta and sham stimulation in the rTPJ and rLPFC. One possible limitation of tACS methods is the wide-spreading stimulation effect depending on the electrode set up which shows reduced focality in contrast to TMS stimulation protocols (Liu et al., 2018; Polania et al., 2018; Thair, Holloway, Newport, & Smith, 2017). For our study design we used a similar electrode set up in line with previous tDCS/tACS studies either targeting the rTPJ (Santiesteban et al., 2012) or the rLPFC (Frings, Brinkmann, Friehs, & van Lipzig, 2018). However, simulating the stimulation effects with SIMNIBS Toolbox revealed that tACS over these brain regions targeted broad areas of the temporo-parietal cortex and the lateral prefrontal cortex, why we cannot make any strong claims about localization specificity of the brain stimulation effects.

5.4.3. EEG

In our third research project we used EEG to examine neuronal markers underlying unexpected norm violations in social cooperation. However, EEG recordings show a precise temporal, but low spatial resolution why it is difficult to make assumptions about the localization of neurobiological process such as ERP components or brain rythms, especially in contrast to neuroimaging data (Olejniczak, 2006). In our research findings the MFN component is thought to be generated at the mPFC based on which we assumed that these neural effects can be observed in the medial prefrontal cortex. This is further supported by the results of our TMS study which show that disruption of the dmPFC with cTBS lead to reduced belief updating for negative in contrast to positive prediction errors which is in line with our EEG results. Nevertheless, based on the limited spatial resolution of EEG data we are cautious with any assumptions about the specific localization of these effects and are cautious to claim in which part of the prefrontal cortex these effects were elicited.

Overall, the methodological considerations and limitations of neuroscientific methods should be carefully taken into account when interpreting our research findings. In particular, although brain stimulation methods show limitations on how much we can infer about the neurophysiological processes underlying social decision making, the assessment of brain stimulation offers new opportunities for possible therapeutic approaches. Nevertheless, past research showed inconsistent findings whether the application of brain stimulation in a clinical population has long-term effects on the psychiatric symptoms (Padberg et al., 2021). Thus future research projects on the precise stimulation parameters, protocols and their effects on the underlying brain function is highly important.

5.5. Conclusions

With the three research projects for my dissertation I tried to provide new insights into the neural basis of across social decision making paradigms. By applying electrophysiology and brain stimulation methods the research findings provide direct evidence for the crucial relevant neuro-cognitive processes of the prefrontal cortex and temporo-parietal cortex in proactive fairness, response to unfairness and social cooperation. Overall our findings improve our understanding of precise social contextual factors for which the rDLPFC implements norm enforcement. Further, we provide direct evidence of the dissociable roles of the rTPJ and rLPFC to promote inequity aversion which is mediated by the rTPJ's role for perspective taking and self other distinction. Further, our findings show direct evidence that the dmPFC updates our beliefs about others when our social expectancies are violated. Thus, our results provide new insights which brain regions are crucial to implement pro-social and cooperative behaviour. Taken together, we advance the field of the neural basis and neuro-psychological mechanisms implemented in these brain regions for such a complex human behaviour as social decision making.

References of General Introduction and Discussion

Alexopoulos, J., Pfabigan, D. M., Goschl, F., Bauer, H., & Fischmeister, F. P. (2013). Agency matters! Social preferences in the three-person ultimatum game. *Frontiers in Human Neuroscience, 7*, 312. doi:10.3389/fnhum.2013.00312

Axelrod, R., & Hamilton, W. D. (1981). The Evolution of Cooperation. *Science, 211*(4489), 1390-1396. doi:doi:10.1126/science.7466396

Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior, 27*(5), 325-344. doi:10.1016/j.evolhumbehav.2006.01.003

Baumgartner, T., Gotte, L., Gugler, R., & Fehr, E. (2012). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human Brain Mapping, 33*(6), 1452-1469. doi:10.1002/hbm.21298

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience, 14*(11), 1468-1474. doi:10.1038/nn.2933

Bechler, C., Green, L., & Myerson, J. (2015). Proportion offered in the Dictator and Ultimatum Games decreases with amount and social distance. *Behavioural Processes, 115*, 149-155. doi:10.1016/j.beproc.2015.04.003

Beeney, J. E., Hallquist, M. N., Clifton, A. D., Lazarus, S. A., & Pilkonis, P. A. (2018). Social disadvantage and borderline personality disorder: A study of social networks. *Personality Disorders: Theory, Research, and Treatment, 9*(1), 62–72. doi:10.1037/per0000234

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature, 442*(7105), 912-915. doi:10.1038/nature04981

Bhattacharya, A., Mrudula, K., Sreepada, S. S., Sathyaprabha, T. N., Pal, P. K., Chen, R., & Udupa, K. (2022). An Overview of Noninvasive Brain Stimulation: Basic Principles and Clinical Applications. *Canadian Journal of Neurological Sciences, 49*(4), 479-492. doi:10.1017/cjn.2021.158

Billeke, P., Armijo, A., Castillo, D., López, T., Zamorano, F., Cosmelli, D., & Aboitiz, F. (2015). Paradoxical Expectation: Oscillatory Brain Activity Reveals Social Interaction Impairment in Schizophrenia. *Biological Psychiatry, 78*(6), 421-431. doi:10.1016/j.biopsych.2015.02.012

Billeke, P., Zamorano, F., Cosmelli, D., & Aboitiz, F. (2013). Oscillatory brain activity correlates with risk perception and predicts social decisions. *Cerebral Cortex, 23*(12), 2872-2883. doi:10.1093/cercor/bhs269

Billeke, P., Zamorano, F., Lopez, T., Rodriguez, C., Cosmelli, D., & Aboitiz, F. (2014). Someone has to give in: theta oscillations correlate with adaptive behavior in social bargaining. *Social Cognitive and Affective Neuroscience, 9*(12), 2041-2048. doi:10.1093/scan/nsu012

Bitsch, F., Berger, P., Nagels, A., Falkenberg, I., & Straube, B. (2019). Impaired Right Temporoparietal Junction-Hippocampus Connectivity in Schizophrenia and Its Relevance for Generating Representations of Other Minds. *Schizophrenia Bulletin, 45*(4), 934-945. doi:10.1093/schbul/sby132

Boksem, M. A., & De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Social Neuroscience, 5*(1), 118-128. doi:10.1080/17470910903202666

Bolognini, N., & Ro, T. (2010). Transcranial magnetic stimulation: disrupting neural activity to alter and assess brain function. *Journal of Neuroscience, 30*(29), 9647-9650. doi:10.1523/jneurosci.1990-10.2010

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America, 100*(6), 3531-3535. doi:10.1073/pnas.0630443100

Brosnan, S. F. (2013). Justice- and fairness-related behaviors in nonhuman primates. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 10416-10423. doi:doi:10.1073/pnas.1301194110

Brown, E., & Brüne, M. (2012). The role of prediction in social neuroscience. *Frontiers in Human Neuroscience, 6*. doi:10.3389/fnhum.2012.00147

Brune, M., Scheele, D., Heinisch, C., Tas, C., Wischniewski, J., & Gunturkun, O. (2012). Empathy moderates the effect of repetitive transcranial magnetic stimulation of the right dorsolateral prefrontal cortex on costly punishment. *PLoS One, 7*(9), e44747. doi:10.1371/journal.pone.0044747

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron, 60*(5), 930-940. doi:10.1016/j.neuron.2008.10.016

Buckholtz, J. W., & Marois, R. (2012). The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience, 15*(5), 655-661. doi:10.1038/nn.3087

Buckholtz, J. W., Martin, J. W., Treadway, M. T., Jan, K., Zald, D. H., Jones, O., & Marois, R. (2015). From Blame to Punishment: Disrupting Prefrontal Cortex Activity Reveals Norm Enforcement Mechanisms. *Neuron, 87*(6), 1369-1380. doi:10.1016/j.neuron.2015.08.023

Burkart, J. M., Brügger, R. K., & van Schaik, C. P. (2018). Evolutionary Origins of Morality: Insights From Non-human Primates. *Frontiers in Sociology, 3*. doi:10.3389/fsoc.2018.00017

Burnett, S., & Blakemore, S. J. (2009). Functional connectivity during a social emotion task in adolescents and in adults. *European Journal of Neuroscience, 29*(6), 1294-1301. doi:10.1111/j.1460-9568.2009.06674.x

Camerer, C. F. (2003). *Behavioral Game Theory* Princeton, NJ: Princeton Univ.
Press.

Campanha, C., Minati, L., Fregni, F., & Boggio, P. S. (2011). Responding to unfair offers made by a friend: neuroelectrical activity changes in the anterior medial prefrontal cortex. *Journal of Neuroscience, 31*(43), 15569-15574. doi:10.1523/JNEUROSCI.1253-11.2011

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron, 70*(3), 560-572. doi:10.1016/j.neuron.2011.02.056

Charness, G., & Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics, 117*(3), 817-869. doi:10.1162/003355302760193904

Chen, M., Zhao, Z., & Lai, H. (2019). The time course of neural responses to social versus non-social unfairness in the ultimatum game. *Social Neuroscience, 14*(4), 409-419. doi:10.1080/17470919.2018.1486736

Cheng, X., Zheng, L., Liu, Z., Ling, X., Wang, X., Ouyang, H., . . . Guo, X. (2022). Punishment cost affects third-parties' behavioral and neural responses to unfairness. *International Journal of Psychophysiology, 177*, 27-33. doi:10.1016/j.ijpsycho.2022.04.003

Christov-Moore, L., Sugiyama, T., Grigaityte, K., & Iacoboni, M. (2017). Increasing generosity by disrupting prefrontal cortex. *Social Neuroscience, 12*(2), 174-181. doi:10.1080/17470919.2016.1154105

Civai, C., Miniussi, C., & Rumiati, R. I. (2015). Medial prefrontal cortex reacts to unfairness if this damages the self: a tDCS study. *Social Cognitive and Affective Neuroscience, 10*(8), 1054-1060. doi:10.1093/scan/nsu154

Courtney, A. L., & Meyer, M. L. (2020). Self-Other Representation in the Social Brain Reflects Social Connection. *Journal of Neuroscience, 40*(29), 5616-5627. doi:10.1523/JNEUROSCI.2826-19.2020

Cutler, J., & Campbell-Meiklejohn, D. (2019). A comparative fMRI meta-analysis of altruistic and strategic decisions to give. *Neuroimage, 184*, 227-241. doi:10.1016/j.neuroimage.2018.09.009

Daan Baas, André Aleman, Matthijs Vink, Nick F. Ramsey, Edward H.F. de Haan, & René S. Kahn. (2008). Evidence of altered cortical and amygdala activation during social decision-making in schizophrenia. *Neuroimage, 40*(2), 719-727. doi:10.1016/j.neuroimage.2007.12.039

Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature, 446*(7137), 794-796. doi:10.1038/nature05651

De Cremer, D., & Barker, M. (2003). Accountability and cooperation in social dilemmas: The influence of others' reputational concerns. *Current Psychology, 22*(2), 155-163. doi:10.1007/s12144-003-1006-6

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist, 13*(6), 580-593. doi:10.1177/1073858407304654

Dulleck, U., Schaffner, M., & Torgler, B. (2014). Heartbeat and economic decisions: observing mental stress among proposers and responders in the ultimatum bargaining game. *PLoS One, 9*(9), e108218. doi:10.1371/journal.pone.0108218

Dungan, J. A., Stepanovic, M., & Young, L. (2016). Theory of mind for processing unexpected events across contexts. *Social Cognitive and Affective Neuroscience, 11*(8), 1183-1192. doi:10.1093/scan/nsw032

Dunn, B. D., Evans, D., Makarova, D., White, J., & Clark, L. (2012). Gut feelings and the reaction to perceived inequity: the interplay between bodily responses, regulation, and perception shapes the rejection of unfair offers on the ultimatum game. *Cognitive, Affective, & Behavioral Neuroscience, 12*(3), 419-429. doi:10.3758/s13415-012-0092-z

Eddy, C. M. (2016). The junction between self and other? Temporo-parietal dysfunction in neuropsychiatry. *Neuropsychologia, 89*, 465-477. doi:10.1016/j.neuropsychologia.2016.07.030

Edmiston, E. K., Merkle, K., & Corbett, B. A. (2014). Neural and cortisol responses during play with human and computer partners in children with autism. *Social Cognitive and Affective Neuroscience, 10*(8), 1074-1083. doi:10.1093/scan/nsu159

Emonds, G., Declerck, C. H., Boone, C., Vandervliet, E. J. M., & Parizel, P. M. (2011). Comparing the neural basis of decision making in social dilemmas of people with different social value orientations, a fMRI study. *Journal of Neuroscience, Psychology, and Economics, 4*, 11-24. doi:10.1037/a0020151

Emonds, G., Declerck, C. H., Boone, C., Vandervliet, E. J. M., & Parizel, P. M. (2012). The cognitive demands on cooperation in social dilemmas: An fMRI study. *Social Neuroscience, 7*(5), 494-509. doi:10.1080/17470919.2012.655426

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences, 11*(10), 419-427. doi:10.1016/j.tics.2007.09.002

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*(6960), 785-791. doi:10.1038/nature02043

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63-87. doi:10.1016/s1090-5138(04)00005-4

Fehr, E., & Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *The American Economic Review, 90*(4), 980-994.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*, 817–868 doi:10.1162/ 003355399556151

Feinberg, M., Willer, R., & Schultz, M. (2014). Gossip and Ostracism Promote Cooperation in Groups. *Psychological Science, 25*(3), 656-664. doi:10.1177/0956797613510184

Feng, C., Eickhoff, S. B., Li, T., Wang, L., Becker, B., Camilleri, J. A., . . . Luo, Y. (2021). Common brain networks underlying human social interactions: Evidence from large-scale neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews, 126*, 289-303. doi:10.1016/j.neubiorev.2021.03.025

Fermin, A. S., Sakagami, M., Kiyonari, T., Li, Y., Matsumoto, Y., & Yamagishi, T. (2016). Representation of economic preferences in the structure and function of the amygdala and prefrontal cortex. *Scientific Reports, 6*, 20982. doi:10.1038/srep20982

Ferrari, C., Lega, C., Vernice, M., Tamietto, M., Mende-Siedlecki, P., Vecchi, T., . . . Cattaneo, Z. (2016). The Dorsomedial Prefrontal Cortex Plays a Causal Role in Integrating Social Impressions from Faces and Verbal Descriptions. *Cerebral Cortex, 26*(1), 156-165. doi:10.1093/cercor/bhu186

Fitzgerald, P. B., Fountain, S., & Daskalakis, Z. J. (2006). A comprehensive review of the effects of rTMS on motor cortical excitability and inhibition. *Clinical Neurophysiology, 117*(12), 2584-2596. doi:10.1016/j.clinph.2006.06.712

Fliessbach, K., Phillipps, C. B., Trautner, P., Schnabel, M., Elger, C. E., Falk, A., & Weber, B. (2012). Neural responses to advantageous and disadvantageous inequity. *Frontiers in Human Neuroscience, 6*, 165. doi:10.3389/fnhum.2012.00165

Fooken, J. (2017). Heart rate variability indicates emotional value during pro-social economic laboratory decisions with large external validity. *Science Reports, 7*, 44471. doi:10.1038/srep44471

Franzen, A., & Pointner, S. (2012). The external validity of giving in the dictator game. *Experimental Economics, 16*(2), 155-169. doi:10.1007/s10683-012-9337-5

Frascarelli, M., Tognin, S., Mirigliani, A., Parente, F., Buzzanca, A., Torti, M. C., . . . Fusar-Poli, P. (2015). Medial frontal gyrus alterations in schizophrenia: relationship with duration of illness and executive dysfunction. *Psychiatry Research, 231*(2), 103-110. doi:10.1016/j.pscychresns.2014.10.017

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186-204. doi:10.1016/j.cortex.2016.04.023

Frings, C., Brinkmann, T., Friehs, M. A., & van Lipzig, T. (2018). Single session tDCS over the left DLPFC disrupts interference processing. *Brain and Cognition, 120*, 1-7. doi:10.1016/j.bandc.2017.11.005

Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron, 50*(4), 531-534. doi:10.1016/j.neuron.2006.05.001

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology, 63*, 287-313. doi:10.1146/annurev-psych-120710-100449

Frith, C. D., & Singer, T. (2008). The role of social cognition in decision making. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363*(1511), 3875-3886. doi:10.1098/rstb.2008.0156

Gabay, A. S., Radua, J., Kempton, M. J., & Mehta, M. A. (2014). The Ultimatum Game and the brain: a meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews, 47*, 549-558. doi:10.1016/j.neubiorev.2014.10.014

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences, 7*(2), 77-83. doi:10.1016/s1364-6613(02)00025-6

Gangopadhyay, P., Chawla, M., Dal Monte, O., & Chang, S. W. C. (2021). Prefrontal-amygdala circuits in social decision-making. *Nature Neuroscience, 24*(1), 5-18. doi:10.1038/s41593-020-00738-9

Gao, X., Yu, H., Saez, I., Blue, P. R., Zhu, L., Hsu, M., & Zhou, X. (2018). Distinguishing neural correlates of context-dependent advantageous- and disadvantageous-inequity

aversion. *Proceedings of the National Academy of Sciences of the United States of America, 115*(33), E7680-E7689. doi:10.1073/pnas.1802523115

George, M. S., Padberg, F., Schlaepfer, T. E., O'Reardon, J. P., Fitzgerald, P. B., Nahas, Z. H., & Marcolin, M. A. (2009). Controversy: Repetitive transcranial magnetic stimulation or transcranial direct current stimulation shows efficacy in treating psychiatric diseases (depression, mania, schizophrenia, obsessive-complusive disorder, panic, posttraumatic stress disorder). *Brain Stimulation, 2*(1), 14-21. doi:10.1016/j.brs.2008.06.001

Gilam, G., Abend, R., Shani, H., Ben-Zion, Z., & Hendler, T. (2019). The anger-infused Ultimatum Game: A reliable and valid paradigm to induce and assess anger. *Emotion, 19*(1), 84-96. doi:10.1037/emo0000435

Gromann, P. M., Heslenfeld, D. J., Fett, A.-K., Joyce, D. W., Shergill, S. S., & Krabbendam, L. (2013). Trust versus paranoia: abnormal response to social reward in psychotic illness. *Brain, 136*(6), 1968-1975. doi:10.1093/brain/awt076

Guroglu, B., van den Bos, W., & Crone, E. A. (2014). Sharing and giving across adolescence: an experimental study examining the development of prosocial behavior. *Frontiers in Psychology, 5*, 291. doi:10.3389/fpsyg.2014.00291

Hallsson, B. G., Siebner, H. R., & Hulme, O. J. (2018). Fairness, fast and slow: A review of dual process models of fairness. *Neuroscience & Biobehavioural Reviews, 89*, 49-60. doi:10.1016/j.neubiorev.2018.02.016

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences of the United States of America, 105*(18), 6741-6746. doi:doi:10.1073/pnas.0711099105

Hanssen, E., Fett, A. K., White, T. P., Caddy, C., Reimers, S., & Shergill, S. S. (2018). Cooperation and sensitivity to social feedback during group interactions in schizophrenia. *Schizophrenia Research, 202*, 361-368. doi:10.1016/j.schres.2018.06.065

Hare, T. A., Camerer, C. F., Knoepfle, D. T., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience, 30*(2), 583-590. doi:10.1523/JNEUROSCI.4089-09.2010

Haroush, K., & Williams, Z. M. (2015). Neuronal prediction of opponent's behavior during cooperative social interchange in primates. *Cell, 160*(6), 1233-1245. doi:10.1016/j.cell.2015.01.045

Harth, N. S., & Regner, T. (2017). The spiral of distrust: (Non-)cooperation in a repeated trust game is predicted by anger and individual differences in negative reciprocity orientation. *International Journal of Psychology, 52 Suppl 1*, 18-25. doi:10.1002/ijop.12257

Hartley, C., & Fisher, S. (2018). Do Children with Autism Spectrum Disorder Share Fairly and Reciprocally? *Journal of Autism and Developmental Disorders, 48*(8), 2714-2726. doi:10.1007/s10803-018-3528-7

Hertz, U., Palminteri, S., Brunetti, S., Olesen, C., Frith, C. D., & Bahrami, B. (2017). Neural computations underpinning the strategic management of influence in advice giving. *Nature Communications, 8*(1), 2191. doi:10.1038/s41467-017-02314-5

Hill, C. A., Suzuki, S., Polania, R., Moisa, M., O'Doherty, J. P., & Ruff, C. C. (2017). A causal account of the brain network computations underlying strategic social behavior. *Nature Neuroscience, 20*(8), 1142-1149. doi:10.1038/nn.4602

Hobot, J., Klincewicz, M., Sandberg, K., & Wierzchoń, M. (2020). Causal Inferences in Repetitive Transcranial Magnetic Stimulation Research: Challenges and Perspectives. *Frontiers in Human Neuroscience, 14*, 586448. doi:10.3389/fnhum.2020.586448

Horat, S. K., Favre, G., Prevot, 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. doi:10.1016/j.schres.2017.05.037

Hu, X., & Mai, X. (2021). Social value orientation modulates fairness processing during social decision-making: evidence from behavior and brain potentials. *Social Cognitive and Affective Neuroscience, 16*(7), 670-682. doi:10.1093/scan/nsab032

Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron, 45*(2), 201-206. doi:10.1016/j.neuron.2004.12.033

Hutcherson, C. A., Bushong, B., & Rangel, A. (2015). A Neurocomputational Model of Altruistic Choice and Its Implications. *Neuron, 87*(2), 451-462. doi:10.1016/j.neuron.2015.06.031

Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., & Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: A meta-analytic review. *Developmental Psychology, 52*(8), 1192-1205. doi:10.1037/dev0000140

Izuma, K., Akula, S., Murayama, K., Wu, D.-A., Iacoboni, M., & Adolphs, R. (2015). A Causal Role for Posterior Medial Frontal Cortex in Choice-Induced Preference Change. *The Journal of Neuroscience, 35*(8), 3598-3606. doi:10.1523/jneurosci.4591-14.2015

Jamali, M., Grannan, B. L., Fedorenko, E., Saxe, R., Báez-Mendoza, R., & Williams, Z. M. (2021). Single-neuronal predictions of others' beliefs in humans. *Nature, 591*(7851), 610-614. doi:10.1038/s41586-021-03184-0

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature, 530*(7591), 473-476. doi:10.1038/nature16981

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange. *Science, 308*(5718), 78-83. doi:doi:10.1126/science.1108062

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex. *Science, 314*(5800), 829-832. doi:doi:10.1126/science.1129156

Kronbichler, L., Tschernegg, M., Martin, A. I., Schurz, M., & Kronbichler, M. (2017). Abnormal Brain Activation During Theory of Mind Tasks in Schizophrenia: A Meta-Analysis. *Schizophrenia Bulletin, 43*(6), 1240-1250. doi:10.1093/schbul/sbx073

Laury, S. K., & Taylor, L. O. (2008). Altruism spillovers: Are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good? *Journal of Economic Behavior & Organization, 65*(1), 9-29. doi:10.1016/j.jebo.2005.05.011

Lee, D. (2008). Game theory and neural basis of social decision making. *Nature Neuroscience, 11*(4), 404-409. doi:10.1038/nn2065

Lee, D., & Seo, H. (2016). Neural Basis of Strategic Decision Making. *Trends in Neurosciences, 39*(1), 40-48. doi:10.1016/j.tins.2015.11.002

Leimgruber, K. L., Rosati, A. G., & Santos, L. R. (2016). Capuchin monkeys punish those who have more. *Evolution and Human Behavior, 37*(3), 236-244. doi:10.1016/j.evolhumbehav.2015.12.002.

Levasseur-Moreau, J., & Fecteau, S. (2012). Translational application of neuromodulation of decision-making. *Brain Stimulation, 5*(2), 77-83. doi:10.1016/j.brs.2012.03.009

Levitt, S. D., & List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives, 21*(2), 153-174. doi:10.1257/jep.21.2.153

Liu, A., Vöröslakos, M., Kronberg, G., Henin, S., Krause, M. R., Huang, Y., . . . Buzsáki, G. (2018). Immediate neurophysiological effects of transcranial electrical stimulation. *Nature Communications, 9*(1), 5092. doi:10.1038/s41467-018-07233-7

Loewenstein, G. F., Bazerman, M. H., & Thompson, L. (1989). Social utility and decision-making in interpersonal contexts. *Journal of Personality and Social Psychology, 57*, 426–441.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., & Baron-Cohen, S. (2011). Specialization of right temporo-parietal junction for mentalizing and its relation to social impairments in autism. *Neuroimage, 56*(3), 1832-1838. doi:10.1016/j.neuroimage.2011.02.067

Luo, J. (2018). The Neural Basis of and a Common Neural Circuitry in Different Types of Pro-social Behavior. *Frontiers in Psychology, 9*, 859. doi:10.3389/fpsyg.2018.00859

Maier, M. J., Rosenbaum, D., Haeussinger, F. B., Brune, M., Enzi, B., Plewnia, C., . . . Ehlis, A. C. (2018). Forgiveness and cognitive control - Provoking revenge via theta-burst-stimulation of the DLPFC. *Neuroimage, 183*, 769-775. doi:10.1016/j.neuroimage.2018.08.065

Makwana, A., & Hare, T. (2012). Stop and be fair: DLPFC development contributes to social decision making. *Neuron, 73*(5), 859-861. doi:10.1016/j.neuron.2012.02.010

Marini, M., Banaji, M. R., & Pascual-Leone, A. (2018). Studying Implicit Social Cognition with Noninvasive Brain Stimulation. *Trends in Cognitive Sciences, 22*(11), 1050-1066. doi:10.1016/j.tics.2018.07.014

McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017). The developmental foundations of human fairness. Nature Human Behaviour, 1(2), 0042. doi:10.1038/s41562-016-0042. *Nature Human Behaviour, 1*, 0042. doi:doi:10.1038/s41562-016-0042

McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition, 134*, 1-10. doi:10.1016/j.cognition.2014.08.013

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review Neuroscience, 24*, 167-202. doi:10.1146/annurev.neuro.24.1.167

Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., & Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron, 75*(1), 73-79. doi:10.1016/j.neuron.2012.05.021

Muller-Leinss, J. M., Enzi, B., Flasbeck, V., & Brune, M. (2018). Retaliation or selfishness? An rTMS investigation of the role of the dorsolateral prefrontal cortex in prosocial motives. *Social Neuroscience, 13*(6), 701-709. doi:10.1080/17470919.2017.1411828

Nicolle, A., Klein-Flugge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron, 75*(6), 1114-1121. doi:10.1016/j.neuron.2012.07.023

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature, 437*(7063), 1291-1298. doi:10.1038/nature04131

Obeso, I., Moisa, M., Ruff, C. C., & Dreher, J.-C. (2018). A causal role for right temporo-parietal junction in signaling moral conflict. *Elife, 7*, e40671. doi:10.7554/eLife.40671

Olejniczak, P. (2006). Neurophysiologic basis of EEG. *Journal of Clinical Neurophysiology, 23*(3), 186-189. doi:10.1097/01.wnp.0000220079.61973.6c

Padberg, F., Bulubas, L., Mizutani-Tiebel, Y., Burkhardt, G., Kranz, G. S., Koutsouleris, N., . . . Brunoni, A. R. (2021). The intervention, the patient and the illness - Personalizing non-invasive brain stimulation in psychiatry. *Experimental Neurology, 341*, 113713. doi:10.1016/j.expneurol.2021.113713

Padmanabhan, A., Lynch, C. J., Schaer, M., & Menon, V. (2017). The Default Mode Network in Autism. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2*(6), 476-486. doi:10.1016/j.bpsc.2017.04.004

Passingham, D., & Sakai, K. (2004). The prefrontal cortex and working memory: physiology and brain imaging. *Current Opinion in Neurobiology, 14*(2), 163-168. doi:10.1016/j.conb.2004.03.003

Pisauro, M. A., Fouragnan, E. F., Arabadzhiyska, D. H., Apps, M. A. J., & Philiastides, M. G. (2022). Neural implementation of computational mechanisms underlying the continuous trade-off between cooperation and competition. *Nature Communications, 13*(1). doi:10.1038/s41467-022-34509-w

Polania, R., Nitsche, M. A., & Ruff, C. C. (2018). Studying and modifying brain function with non-invasive brain stimulation. *Nature Neuroscience, 21*(2), 174-187. doi:10.1038/s41593-017-0054-4

Porcelli, S., Van Der Wee, N., van der Werff, S., Aghajani, M., Glennon, J. C., van Heukelum, S., . . . Serretti, A. (2019). Social brain, social dysfunction and social withdrawal. *Neuroscience & Biobehavioral Reviews, 97*, 10-33. doi:10.1016/j.neubiorev.2018.09.012

Quesque, F., & Brass, M. (2019). The Role of the Temporoparietal Junction in Self-Other Distinction. *Brain Topography, 32*(6), 943-955. doi:10.1007/s10548-019-00737-5

Rahal, R.-M., Fiedler, S., & De Dreu, C. K. W. (2020). Prosocial Preferences Condition Decision Effort and Ingroup Biased Generosity in Intergroup Decision-Making. *Scientific Reports, 10*(1), 10132. doi:10.1038/s41598-020-64592-2

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences, 17*(8), 413-425. doi:10.1016/j.tics.2013.06.003

Reuben, E., & van Winden, F. (2008). Social ties and coordination on negative reciprocity: The role of affect. *Journal of Public Economics, 92*(1), 34-53. doi:10.1016/j.jpubeco.2007.04.012

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A Neural Basis for Social Cooperation. *Neuron, 35*, 395–405. doi:10.1016/s0896-6273(02)00755-9

Rilling, J. K., King-Casas, B., & Sanfey, A. G. (2008). The neurobiology of social decision-making. *Current Opinion in Neurobiology, 18*(2), 159-165. doi:10.1016/j.conb.2008.06.003

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology, 62*, 23-48. doi:10.1146/annurev.psych.121208.131647

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *Neuroimage, 22*(4), 1694-1703. doi:10.1016/j.neuroimage.2004.04.015

Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Review Neuroscience, 15*(8), 549-562. doi:10.1038/nrn3776

Sanfey, A. G. (2007). Social Decision-Making: Insights from
Game Theory and Neuroscience. *Science, 318*, 598 - 602. doi:10.1126/science.1142996

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science, 300*, 1755-1758. doi:10.1126/science.1082976

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012). Enhancing social ability by stimulating right temporoparietal junction. *Current Biology, 22*(23), 2274-2277. doi:10.1016/j.cub.2012.10.018

Santos, F. P., Santos, F. C., & Pacheco, J. M. (2018). Social norm complexity and past reputations in the evolution of cooperation. *Nature, 555*(7695), 242-245. doi:10.1038/nature25763

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage, 19*(4), 1835-1842. doi:10.1016/s1053-8119(03)00230-1

Schneider, K., Pauly, K. D., Gossen, A., Mevissen, L., Michel, T. M., Gur, R. C., . . . Habel, U. (2013). Neural correlates of moral reasoning in autism spectrum disorder. *Social Cognitive and Affective Neuroscience, 8*(6), 702-710. doi:10.1093/scan/nss051

Soutschek, A., Ruff, C. C., Strombach, T., Kalenscher, T., & Tobler, P. N. (2016). Brain stimulation reveals crucial role of overcoming self-centeredness in self-control. *Science Advances, 2*(10), e1600992. doi:doi:10.1126/sciadv.1600992

Soutschek, A., Sauter, M., & Schubert, T. (2015). The Importance of the Lateral Prefrontal Cortex for Strategic Decision Making in the Prisoner's Dilemma. *Cognitive, Affective, & Behavioral Neuroscience, 15*(4), 854-860. doi:10.3758/s13415-015-0372-5

Speer, S. P. H., & Boksem, M. A. S. (2019). Decoding fairness motivations from multivariate brain activity patterns. *Social Cognitive and Affective Neuroscience, 14*(11), 1197-1207. doi:10.1093/scan/nsz097

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., & Fehr, E. (2007). The Neural Signature of Social Norm Compliance. *Neuron, 56*(1), 185-196. doi:10.1016/j.neuron.2007.09.011

Stallen, M., & Sanfey, A. G. (2013). The cooperative brain. *Neuroscientist, 19*(3), 292-303. doi:10.1177/1073858412469728

Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B., & Sack, A. T. (2015). Be nice if you have to--the neurobiological roots of strategic fairness. *Social Cognitive and Affective Neuroscience, 10*(6), 790-796. doi:10.1093/scan/nsu114

Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S., & Kirsch, P. (2011). Beyond revenge: Neural and genetic bases of altruistic punishment. *Neuroimage, 54*(1), 671-680. doi:10.1016/j.neuroimage.2010.07.051

Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, II, Tobler, P. N., & Kalenscher, T. (2015). Social discounting involves modulation of neural value signals by temporoparietal junction. *Proceedings of the National Academy of Sciences of the United States of America, 112*(5), 1619-1624. doi:10.1073/pnas.1414715112

Suzuki, S., & O'Doherty, J. P. (2020). Breaking human social decision making into multiple components and then putting them together again. *Cortex, 127*, 221-230. doi:10.1016/j.cortex.2020.02.014

Tei, S., Fujino, J., Hashimoto, R. I., Itahashi, T., Ohta, H., Kanai, C., . . . Takahashi, H. (2018). Inflexible daily behaviour is associated with the ability to control an automatic reaction in autism spectrum disorder. *Science Reports, 8*(1), 8082. doi:10.1038/s41598-018-26465-7

Thair, H., Holloway, A. L., Newport, R., & Smith, A. D. (2017). Transcranial Direct Current Stimulation (tDCS): A Beginner's Guide for Design and Implementation. *Frontiers in Neuroscience, 11*. doi:10.3389/fnins.2017.00641

Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature, 463*(7284), 1089-1091. doi:10.1038/nature08785

Ulber, J., Hamann, K., & Tomasello, M. (2017). Young children, but not chimpanzees, are averse to disadvantageous and advantageous inequities. *Journal of Experimental Child Psychology, 155*, 48-66. doi:10.1016/j.jecp.2016.10.013

Underwood, B., & Moore, B. (1982). Perspective-taking and altruism. *Psychological Bulletin, 91*(1), 143–173. doi:10.1037/0033-2909.91.1.143

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage, 48*(3), 564-584. doi:10.1016/j.neuroimage.2009.06.009

Veniero, D., Strüber, D., Thut, G., & Herrmann, C. S. (2019). Noninvasive Brain Stimulation Techniques Can Modulate Cognitive Processing. *Organizational Research Methods, 22*(1), 116-147. doi:10.1177/1094428116658960

Wang, A., Zhu, L., Lyu, D., Cai, D., Ma, Q., & Jin, J. (2022). You are excusable! Neural correlates of economic neediness on empathic concern and fairness perception.

Cognitive, Affective, & Behavioral Neuroscience, 22(1), 99-111. doi:10.3758/s13415-021-00934-5

Wu, Y., Yu, H., Shen, B., Yu, R., Zhou, Z., Zhang, G., . . . Zhou, X. (2014). Neural basis of increased costly norm enforcement under adversity. *Social Cognitive and Affective Neuroscience, 9*(12), 1862-1871. doi:10.1093/scan/nst187

Yang, L., Li, P., Mao, H., Wang, H., Shu, C., Bliksted, V., & Zhou, Y. (2017). Theory of mind deficits partly mediate impaired social decision-making in schizophrenia. *BMC Psychiatry, 17*(1), 168. doi:10.1186/s12888-017-1313-3

Yang, Z., Zheng, Y., Yang, L., Li, Q., & Liu, X. (2019). Neural signatures of cooperation enforcement and violation: a coordinate-based meta-analysis. *Social Cognitive and Affective Neuroscience, 14*(9), 919-931. doi:10.1093/scan/nsz073

Yoshida, K., Saito, N., Iriki, A., & Isoda, M. (2011). Representation of others' action by neurons in monkey medial frontal cortex. *Current Biology, 21*(3), 249-253. doi:10.1016/j.cub.2011.01.004

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences of the United States of America, 104*(20), 8235-8240. doi:10.1073/pnas.0701408104

Zhu, L., Mathewson, K. E., & Hsu, M. (2012). Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proceedings of the National Academy of Sciences of the United States of America, 109*(5), 1419-1424. doi:10.1073/pnas.1116783109

List of Publications

Journal Articles

Christian, P., & Soutschek, A. (2022). Causal role of right dorsolateral prefrontal cortex for norm-guided social decision making: A meta-analysis of TMS studies. *Neuropsychologia, 176*, 108393. doi:10.1016/j.neuropsychologia.2022.108393

Christian, P., Kapetaniou, G. E. & Soutschek, A. (submitted). Causal roles of prefrontal and temporo-parietal theta oscillations for inequity aversion. *Social Cognitive and Affective Neuroscience.*

Christian, P., Kaiser, J., Taylor, P., George, M. Schütz-Bosbach, S. & Soutschek, A. (in preparation). The causal role of medial prefrontal cortex for updating of mental model representations in social interactions.

Kapetaniou, G. E., Reinhard, M. A., **Christian, P.,** Jobst, A., Tobler, P. N., Padberg, F., & Soutschek, A. (2021). The role of oxytocin in delay of gratification and flexibility in non-social decision making. *Elife, 10*. doi:10.7554/eLife.61844

Soutschek, A., Nadporozhskaia, L., & **Christian, P.** (2022). Brain stimulation over dorsomedial prefrontal cortex modulates effort-based decision making. *Cognitive Affective & Behavioural Neuroscience*. doi:10.3758/s13415-022-01021-z

Conference contributions

Christian, P., Kaiser, J., Taylor, P., George, M. Schütz-Bosbach, S. & Soutschek, A. The causal role of medial prefrontal cortex for updating of mental model representations in competitive interactions. Talk at the 6th Conference of the European Society for Cognitive and Affective Neuroscience (ESCAN), July 2022.

Christian, P., Kaiser, J., Taylor, P., George, M. Schütz-Bosbach, S. & Soutschek, A. The causal role of medial prefrontal cortex for updating of mental model representations in competitive interactions. Poster Presentation at the 6th Conference of the Social and Affective Neuroscience Conference (SANS), Mai 2022.

Christian, P., Kapetaniou, G. E. & Soutschek, A. Causal roles of prefrontal and temporo-parietal theta oscillations for inequity aversion. Poster presentation at the 5th Conference of the European Society for Cognitive and Affective Neuroscience (ESCAN), July 2021.

Christian, P., & Soutschek, A. Causal role of right dorsolateral prefrontal cortex for norm-guided social decision making: A meta-analysis of TMS studies. Poster presentation at the 18th Annual Meeting of the Society for Neuroeconomics (SNE), October 2020.

Author contributions

Chapter 1. Causal role of right dorsolateral prefrontal cortex for norm-guided social decision making: A meta-analysis of TMS studies

The author of this dissertation contributed to research design, programmed the tasks, collected the data, performed data analysis, interpreted the results, created plots and wrote the manuscript.

Alexander Soutschek contributed to research design, assisted with data collection and data analysis, contributed to the interpretation of results, supervised the experiment, and wrote the manuscript.

Chapter 2. Causal roles of prefrontal and temporo-parietal theta oscillations for inequity aversion.

The author of this dissertation contributed to research design, programmed the tasks, collected the data, performed data analysis, interpreted the results, created plots and wrote the manuscript.

Georgia Eleni Kapetaniou contributed to data analysis and provided comments on the manuscript.

Alexander Soutschek contributed to research design, assisted with data collection and data analysis, contributed to the interpretation of results, supervised the experiment, and wrote the manuscript.

Chapter 3. The causal role of medial prefrontal cortex for updating of mental model representations in social interactions.

The author of this dissertation contributed to research design, programmed the tasks, collected the data, performed data analysis, interpreted the results, created plots and wrote the manuscript.

Jakob Kaiser performed data analysis and contributed to the interpretation of results and supervised the experiment.

Paul Taylor contributed to research design, contributed to the interpretation of results and provided comments on the manuscript.

Michelle George assisted with data collection and provided comments on the manuscript.

Simone Schütz-Bosbach contributed to research design and provided comments on the manuscript.

Alexander Soutschek contributed to research design, assisted with data collection and data analysis, contributed to the interpretation of results, supervised the experiment, and wrote the manuscript.

Munich, 30.03.2023

Patricia Christian

Dr. Alexander Soutschek

Acknowledgements

First and foremost, I would like to thank my supervisor, Alexander Soutschek for his incredible help and support over the past years, no words can express how much I appreciate him being there for me on every step of the PhD process. I am extremely grateful for your unconditional supervision, guidance, patience and support during the course of this doctoral thesis. Thanks to your guidance and true mentorship and your openness and great support to give me the freedom to create my own ideas, you supported me to become the researcher I truly wanted to be. You created an incredible work environment with patience, positive attitude and kindness which truly motivated me on every step of the PhD and to pursue a career in research.

I also would like to thank my second supervisor Paul Sauseng who always supported me in my wish to pursue a research career starting from my master project to the end of the PhD thesis, thanks you so much for your help and support and your valuable comments on my PhD projects, I really appreciate it!

I am grateful for the support of Paul Taylor who supported me on how to use TMS methods, thereby helping me to realize one of my research projects and who provided valuable feedback and comments on the interpretation of my TMS results.

I want to thank Simone Schütz Bosbach for her expertise to support the research designs for my projects and being available for all kind of questions and problems concerning this technique.

I am grateful for the support of Jakob Kaiser with my EEG projects, who supported me how to analyze EEG Data and who provided valuable feedback and comments on the interpretation of the results and the manuscript.

I want to thank my family, who taught my the value of education and helped me to develop persistence and ambition to reach my goals. I want to thank my incredible sister who has been my rock, my inspiration and supported my on the whole way from the very beginning. I'm eternally grateful that she was always there for me, taught me how to choose the right path for me, truly believed in me and let me shine.

I want to thank my incredible partner Philip for his unconditional love, kindness and support, for his strength to always support me even in the most difficult times when I needed to work crazy working hours or complained about some testing difficulties. Thank you for being by my side and for not letting me lose sight of the important things in life!

Further, I want to thank all the people who love my unconditionally, always supported me and gave me the strength to pursue my goals and dreams, even when they seemed unreachable or just to much to keep up with: For my amazing friends Florentina, Gloria, Lanah, Miri, Thesi, Juli, Vanessa, Sarah, Nico und Svenja.

This work is dedicated to all my beloved friends, my amazing partner and my sister.