Journal of Neuroscience, Psychology, and Economics

ISSN: 1937-321X

2022, Vol. 15, No. 4, 222–240
https://doi.org/10.1037/npe0000163

Does Power Corrupt? An fMRI Study on the Effect of Power and Social Value Orientation on Inequity Aversion

Loren Pauwels, Carolyn H. Declerck, Christophe Boone, Paloma Diaz-Gutiérrez, and Bruno Lambert
Faculty of Business and Economics, University of Antwerp

Accumulating evidence corroborates that power asymmetries influence how people respond to violations of the equality norm. We investigate with functional magnetic resonance imaging how preferences for advantageous (receiving more than other) and disadvantageous (receiving less than other) inequality are affected by having social power and whether or not this differs according to an individual's social value orientation. Forty participants were primed with either a leader- or a teammate role (control) before conducting a task in the scanner during which they rated 36 monetary distributions which varied in degree of inequality. Behavioral data indicate that taking on a leader role generally increases aversion to disadvantageous inequality (DA-IE), but that it decreases aversion to larger advantageous inequality, especially for proself individuals. Consistently, the blood oxygen level–dependent (BOLD) data reveal that, as advantageous inequality mounts and the temptation to have more than others increases, leaders show reduced neural activation in regions associated with perspective taking and cognitive control (precuneus and frontal eye field). Proself leaders in particular show reduced activity in the left and right dorsolateral prefrontal cortex, which could be interpreted as a weaker restrain of self-interest when processing advantageous inequality. We found no evidence for an effect of power on processing DA-IE.

Keywords: inequity aversion, fairness, fMRI, social power, social value orientation

Supplemental materials: https://doi.org/10.1037/npe0000163.supp

Equality is a benchmark across human societies and violations of the equality norm are ubiquitously disliked and punished. Despite this well-documented universality (e.g., Henrich et al., 2006), there is considerable heterogeneity in the extent to which people are inequity averse and willing to set aside self-interest in support of equal outcomes (e.g., Fehr & Schmidt, 1999).

A well-studied factor with substantial influence on preferences for equality (and conversely, on inequity aversion) is power asymmetry. Gaining power in relationships has been shown to lead to reduced concern for others (Laurin et al., 2016; Righetti et al., 2015), lower trust (Schilke et al., 2015) and weaker reciprocation (Pauwels, 2021), while priming a position of power increases cheating during a subsequent experimental task (Cohn et al., 2014). Similarly, occupying a higher social class correlates negatively with compassion and generosity (Piff et al., 2010) and positively with unethical behaviors like breaking the law while driving or lying in negotiations, both of which can be attributed to a positive attitude toward greed (Piff et al., 2012). At the same time, not all those in power become greedy and corrupt; people also differ individually, which adds to the heterogeneity. Given the many societal problems associated with greed, social mobility, and growing inequality, understanding the cognitive mindset that makes a person in power more or less inequity

This article was published Online First September 8, 2022.
Carolyn H. Declerck https://orcid.org/0000-0001-9348-112X
This research was supported by a Bijzonder Onderzoeksfonds Grant from the University of Antwerp to Carolyn H. Declerck.
Correspondence concerning this article should be addressed to Carolyn H. Declerck, Faculty of Business and Economics, University of Antwerp, Prinsstraat 13, 2000 Antwerpen, Belgium. Email: carolyn.declerck@uantwerpen.be

averse is highly pertinent. Functional magnetic resonance imaging (fMRI) might help to reveal the motives behind actual behavior. With this technique, participants in the scanner are confronted with a dilemma-type situation and, through inferential statistics, it is then possible to identify those regions of the brain that are relatively more metabolically active at the time a decision is made. For example, fMRI can corroborate that people with different values solve the same problem by using different brain regions. As such, fairness-related decisions can be an expression of inequity aversion, but they can also be strategically motivated when tied to personal benefits (Emonds et al., 2011).

A vast number of neuroscience experiments have identified neural correlates of inequity aversion, but as far as we know there are no studies investigating how these are affected by the momentary experience of power. In the present study, we bring together what is known about the neuroscience of inequity aversion with the social psychology literature on the corrupting effects of power to uncover possible differences in the pattern of neural activation when people are put in powerful positions and then confronted with inequality. Power is here defined as a psychological state that comes with having control over others (Galinsky et al., 2003, 2006; Keltner et al., 2003). Because power may not affect everyone alike but, instead, may serve to strengthen one's inner disposition or core values (Guinote, 2017), we further consider how power interacts with an individual's social value orientation (SVO), a stable trait indicating how much weight a person attaches to the distributive outcome of others in relation to their own. While SVO is typically taken to reflect an individual's "inner compass" with much predictive value across situations (Declerck & Boone, 2018), most people do deviate to some extent from their baseline SVO in response to certain (sometimes subtle) environmental cues (Murphy & Ackermann, 2014). Being in a position of power may be one such cue. Following a long tradition in psychology asserting that trait- and state variables cannot be studied independently (Mischel & Shoda, 1998), we consider a person's *hic et nunc* revealed preference for equality to be the result of how an individual with a particular preference "type" responds to the social environment (Bruhin et al., 2019). That is, inequity aversion = $f(\text{SVO} \times \text{Power})$.

We assess stated preferences by means of an unincentivized judgment task which is unobtrusive, simple to implement, and allows us to elicit individuals' intuitive feelings about inequality without any influence of the anticipated consequences of one's choice. We develop hypotheses separately for resource distributions that are disadvantageous (i.e., receiving less than other) or advantageous (receiving more than others), as there is accumulating evidence that the resulting aversion to disadvantageous inequality (DA-IE) and advantageous inequality (A-IE) is substantially different. While a confrontation with both DA-IE and A-IE is expected to elicit emotions and activate social comparison (and hence mentalizing), aversion to DA-IE is naturally stronger than to A-IE (Fehr & Schmidt, 1999; Loewenstein et al., 1989), and the two correlate only weakly (Chapman et al., 2018). A breach of fairness in DA-IE unequivocally harms self-interest and automatically triggers feelings of disgust: noone enjoys being hurt. This universality is well-documented by the protesting behavior of many species of nonhuman primates confronted with DA-IE (Brosnan & de Waal, 2014). In contrast, A-IE creates a conflict between self-interest and the fairness principle, which is likely to elicit different emotions depending on the individual's intrinsic other-regarding values. Only those individuals who experience guilt when taking the other's perspective may be sufficiently motivated to restrain self-interest for the sake of a fair outcome. As a result, responses tend to be much more heterogeneous in A-IE than in DA-IE (Loewenstein et al., 1989). Furthermore, the neurocognitive processes associated with DA-IE and A-IE have also been shown to differ (Gao et al., 2018), a topic we turn to next.

The Effect of Power in DA-IE

The neural correlates of experiencing DA-IE have been well-studied and corroborate by and large the increased involvement of aversive emotional processing, notably in the *insula* (Haruno et al., 2014; Hsu et al., 2008; Sanfey et al., 2003; Tabibnia et al., 2008; Tanaka et al., 2017) and *amygdala* (Gao et al., 2018; Haruno & Frith, 2010; Haruno et al., 2014; Hu et al., 2016). More detailed descriptions of these regions can be found in the Supplemental Information. In addition, some studies indicate that the tolerance for inequality in DA-IE is reduced specifically when people feel entitled, for example, when they exerted more effort (Cappelen et al., 2014) or when they are of higher social status (Hu et al., 2014). The stronger the sense of entitlement, the more salient the inequality becomes, and the more we would expect concurrent amygdala and insula activation.

Social power as well has been argued to create entitlements, thereby increasing DA-IE aversion. First, according to social distance theory, the powerful have different beliefs, attitudes, and emotions compared to the powerless because their position gives them easy access to resources and a relatively greater independence. Being independent shifts the focus away from others in relationships, causing them to pay less attention to their needs and feel more entitled to the available resources (Magee & Smith, 2013; van Kleef et al., 2008). Second, identity theory in economics (Akerlof & Kranton, 2000) predicts that the powerful will behave conform the norms associated with the identity of being in power (Joshi & Fast, 2013). Having more than others in this case comes across as a privilege which may cause anger or disgust when broken. Consistently, power has been reported to heighten the sensitivity to breaches of the equality norm especially when the inequity negatively affects the self (Sawaoka et al., 2015).

Thus, we state Hypothesis 1:

> *Hypothesis 1:* In DA-IE, having power over others increases inequity aversion. At the neural level, processing inequality when having power is associated with increased activation in brain regions involved in emotional aversion, namely, the insula and amygdala.

The Effect of Power in A-IE

When confronted with A-IE, self-interest conflicts with the equality norm. To be able to express inequity aversion when inequity is in fact self-serving (assuming "more" is always better) requires higher order cognitive functions, including mentalizing abilities to take the perspective of others and prefrontal cortex (PFC)-associated cognitive control functions to overcome selfish impulses (Gao et al., 2018). Not surprisingly, because of the involvement of higher order cognition, A-IE is far less common among nonhuman primates (Brosnan & de Waal, 2014). Because not all humans experience the tension between self-interest and the equality norm to the same extent, the variance in inequity aversion is expected to be much greater in A-IE.

Because power increases entitlement through increased social distance and social norms that legitimize greed, A-IE is expected to be weaker for the powerful. The entitlements and higher earnings shift the focus to the self, thereby reducing a person's *hic et nunc* preferences for equality. This momentary change in inequity aversion reduces the need for perspective taking and/or caring for others, which we expect would engage mentalizing functions in the brain. This is consistent with a neuroimaging study revealing that, compared to individuals of high social status, those with a lower social status were more prone to mentalize, which corresponded with greater activity in the *precuneus* and *dorsomedial prefrontal cortex* (*dmPFC*, Muscatell et al., 2012, see Supplemental Information for description, as well as the meta-analysis by Molenberghs et al., 2016). Moreover, by feeling entitled, the powerful might not perceive A-IE as a norm violation, which would alleviate the need to resolve the conflict between self-interest and equality. In the brain, restraining self-interest to promote equality has been associated with regions of the *dorsolateral PFC* (dlPFC, Knoch et al., 2006; Spitzer et al., 2007) and the *temporoparietal junction* (*TPJ*, Morishima et al., 2012; see Supplemental Information for further descriptions of these regions). The TPJ is typically inferred as part of the mentalizing network (e.g., the meta-analysis by Molenberghs et al., 2016), and additional research suggests that its activity is especially related to mentalizing in difficult situations, as is the case when making decisions that go against self-interest (Morishima et al., 2012; Park et al., 2017).

Therefore,

> *Hypothesis 2a:* In A-IE, having power over others decreases inequity aversion. At the neural level, processing inequality when having power is associated with decreased activation in brain regions associated with higher order cognitive control (dlPFC) and mentalizing functions (TPJ, dmPFC, precuneus).

The conflict between self-interest and the equality norm is, however, not equally salient for all those in power. Individuals with a prosocial value orientation who intrinsically care about the fate of others and are motivated to seek equal outcomes will experience more discomfort and inequity aversion in A-IE compared to individuals with a proself value orientation (Bogaert et al., 2008; Van Lange, 1999). This difference is reflected in different neural correlates of prosocial decision making: first, while prosocials show more activation in the dlPFC and precuneus when they make cooperative decisions in mixed-motive games (Emonds et al., 2013), proselfs present increased activation in these same regions when they have to compute whether or not a decision will pay off (Emonds et al., 2011). Second, only prosocials continue to show higher amygdala activation for A-IE (Haruno & Frith, 2010).

Power could mitigate the internal social conflicts elicited in A-EI in two directions depending on one's chronically activated SVO (Chen et al., 2001; Côté et al., 2011; DeCelles et al., 2012; Guinote, 2017; Kraus et al., 2011). For proselfs, a position of power that legitimizes self-interest would relax the need for cognitive control as they no longer have to suppress the fact that they like receiving more than others. In contrast, for prosocials, the entitlements that come with the position of power create temptations that may interfere with their other-regarding nature. As the inequality becomes larger, this would exacerbate the conflict and increase the need for cognitive control in order to resist the personal benefits of A-IE and remain true to their intrinsic value system which advances equality.

This reasoning is consistent with the literature on the role of conflict to explain the relation between intuition and prosociality (e.g., Evans et al., 2015). Based on both reaction-time experiments (Yamagishi et al., 2017) as well as those with cognitive load (Cornelissen et al., 2011) it has been shown that prosocials and proselfs make decisions in a fast and intuitive manner when their choice is consistent with their own intrinsic preference, while they will be more deliberating and reflective when they experience conflict. Accordingly, A-IE aversion has already been shown to be intuitive for prosocials (Haruno & Frith, 2010), but not for proselfs. By legitimizing greed, power interferes with the intuitive A-IE aversion of prosocials because feeling entitled is no longer compatible with their intrinsic preference. Resolving this would require deliberation. In contrast, for proselfs, power reduces the gap between feeling entitled and their intrinsic preference, so that a preference for A-IE becomes intuitive with little need for deliberation.

Thus, we complement Hypothesis 2a:

> *Hypothesis 2b:* In A-IE, having power over others reduces inequity aversion especially for proselfs. At the neural level, for proselfs having power, processing inequality is associated with decreased activation in cognitive control regions of the brain (dlPFC). For prosocials having power, processing inequality is associated with increased activation in these regions.

Method

Forty healthy volunteers (university students from different departments, with average age = 22 years, *SD* = 2 years) participated in the scanning sessions of the experiment. Participants were recruited via university email and invited to participate in an MRI study on social judgment and decision-making with monetary compensation. All procedures were conform with the Declaration of Helsinki and were approved by the university's Commission for Medical Ethics. Based on questionnaire data obtained several weeks prior to the fMRI session, participants were selected to achieve a sample that was balanced for gender and SVO, categorized into a prosocial or proself type with the triple dominance measure (Van Lange, 1999). This measure consists of a series of nonstrategic "decomposed" games, in which the participant is asked to choose unilaterally between several options that affect both one's own payoff as well as that of another person. Because the other has no impact on payoffs, all strategic considerations are removed, making SVO a straightforward measure of a person's "baseline," or unconditional preference for equal outcomes. Prosocials put equal weight on payoffs for self and others, while proselfs maximize either payoffs for themselves or the relative difference between payoffs for self and others. By measuring SVO (the "baseline" preference) at least 1 week prior to the experiment, we dissociate it from the variable of interest (the revealed preference) which we postulate to be a function of both SVO and the power manipulation.

Each participant attended one experimental session which took place at the University hospital. Compensation ranged between 39.5€ and

49€, including a 30€ show up fee and a variable payment depending on the outcome of one of the tasks.[1] Debriefing occurred after conclusion of the study by sending all participants an email describing the intent, procedures, and results of the study. No deception was used.

The experimental session consisted of three parts which are described in three subsections below: (a) generating ideas for a creativity task, which was a part of the power manipulation prior to scanning, (b) evaluating reward pairs between self and other in the MRI scanner, (c) completing the creativity task and postscanning questionnaire.

Prescanning: Generating Creative Ideas (Power Manipulation Task)

Participants were informed in truth that they were matched with another person in the study (who remained anonymous) to create a team of two in which each member was to come up with ideas for the creative use of 1,000 pipe cleaners. These creative ideas were generated independently prior to scanning, and without consulting the other person in the team. To incentivize the task, participants were told that one idea per team would be submitted to a jury and that the members of the winning team would be awarded a bonus of 15€.[2]

Within each team, participants to be scanned were assigned one of two possible roles which remained constant throughout the experiment: (a) they could be the leader of the team (n = 20, high power condition), in which case they were matched with a follower, or (b) a teammate (n = 20, control condition) in which case they were matched with another teammate. Roles were balanced over SVO, yielding an equal number of prosocial/proself leaders, and prosocial/proself teammates. Role descriptions were adapted from published work by Inesi et al. (2012).[3]

Participants labeled as "leader" were told that they would be able to read the ideas of their "follower" after the scan session was completed. At that time they were to (a) evaluate the creative ideas provided by their follower, (b) decide on the best creative idea (either their own, or that of their follower) which would be submitted to a jury, and (c) determine how to divide a 19€ compensation between themselves and their follower. Followers (n = 20) did not participate in the scan session, but they submitted their creative ideas 2 weeks prior to the experimental session so that they could be matched with leaders. Followers had no control over the outcome of the creativity task, but they were compensated in truth based on leaders' decisions.

In the control condition, the role description was such that power over the task outcome was symmetrically distributed; both teammates evaluated their own work, had a vote in deciding the best idea, and both received a fixed 9.5€ compensation for the task. The creative ideas of teammates were randomly matched with each other after the scan session.

During Scanning: Evaluating Reward Pair Distributions

In the scanner, participants observed and rated a series of 36 hypothetical distributions of a sum of money split between themselves and the anonymous person with whom they were matched in the creativity task. Half of the distributions yielded an advantage to the participant, while the other half yielded a disadvantage. Hence DA-IE and A-IE was a within factor in the experiment, while the assumed power position (leader or teammate) was a between factor. The task instructions explaining how to evaluate the money distributions emphasized again the power position assumed by the participants and the interaction partners with whom they were matched. Immediately before conducting this task in the scanner, participants were asked a series of questions regarding their power role,

[1] This was the reward obtained for partaking in a creativity task which served to assign each participant a leader- or teammate role (explained later in text). All teammates received a reward of 9,5€, while leaders could decide how much of a 19€ to allocate to themselves and how much to reward their "follower".

[2] This was implemented as stated in the participant instructions. After data collection of all participants was complete, the best ideas of each team were gathered and submitted to an independent jury of three colleagues from the research group. The jury was asked to judge these ideas on innovation, feasibility and usefulness. They deliberated in group and came up with one clear winner.

[3] These authors used this method successfully to show that leaders distance themselves from their subordinates by discounting their generous acts. In addition Pauwels (2021) successfully manipulated power with this method and showed that it significantly reduced the amount of money high power participants were willing to pay in a costly punishment experiment.

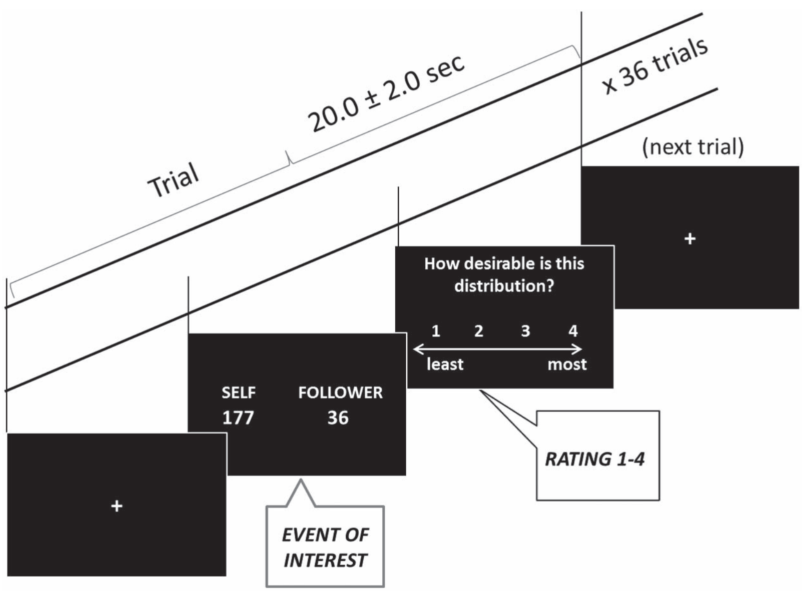
which they had to answer with a button press: (a) "What is your role in the creativity task? Press left for leader–press right for subordinate"; (b) "Do I get to see the creative ideas of my subordinate? Press right for yes–press left for no"; (c) "Do I have to evaluate my subordinate on their ideas? Press left for yes–press right for no"; (d) "Do I get to decide on the compensation for this task? Press right for yes–press left for no." Finally, during the actual evaluation task, participants were again reminded of their roles in each of the 36 trials through the labels above the money allocations ("me" and "subordinate" or "teammate").

Figure 1 illustrates the sequence of frames in one trial. First, a fixation cross appeared for 2 s plus a jitter of 0–2 s. Next, during the stimulus presentation phase (the event of interest), a unique combination of monetary rewards for the participant and his follower/teammate was shown for a fixed time of 5.5 s. This is the time interval during which we want to know how the blood oxygen level–dependent (BOLD) signal is affected by condition (DA-IE vs. A-IE) and by the power manipulation. Because the goal of the present study is to test if there are differences in how people with power react intuitively when they are confronted with inequality, dissociating the stimulus presentation phase from the actual rating seems more appropriate to capture participants' *gut response*. An additional advantage of not including the participant's decision in this event of interest is to be able to compute the BOLD signal across fixed time intervals for all participants (see Gao et al., 2018, for similar reasoning).

The 36 reward distributions varied in their level of inequality (with 18 distributions resulting in DA-IE and 18 in A-IE) and were taken integrally from Haruno and Frith (2010, see Supplemental Table 1). In the last frame (the response phase), participants were asked to rate each distribution on a scale from 1 to 4 (least to most desirable), yielding a "liking score" (the inverse of inequity aversion), the dependent variable in the behavioral analyses. Ratings were entered via button presses using Lumina LS-PAIR response pads with two buttons in each hand (coupled to the Lumina LSC-400 controller,

Figure 1

Schematic Representation of One Trial of the Evaluation Task Conducted in the Scanner



Note. fMRI = functional magnetic resonance imaging. The wording on the screens were in Dutch (native language). Dependent variables are the blood oxygen level–dependent (BOLD) signal during reward pair presentation (event of interest) for the fMRI analyses, and the liking scores that were assigned after reward pair presentation (rating 1–4) for the behavioral analyses.

Cedrus, CA, USA). No correction for left/right hand bias was made.[4] The Presentation software (Neurobehavioral Systems, Inc., Albany, California, USA) was used for stimulus presentation and for recording the button press responses and trial-related timing variables.

Images were gathered with a Siemens MR Prisma scanner and an eight-channel head coil (30 slices per image, TR = 2 s). The image analysis was conducted with the statistical parametric mapping package (SPM12; Wellcome Department of Cognitive Neurology, London). During preprocessing, images were (a) corrected for slice timing, (b) realigned, (c) normalized against the Montreal Neurologic Institute (MNI) template and (d) spatially smoothed (full width half maximum = 7 mm) and (e) temporally filtered with a 128 s high-pass filter.

First-level analyses were conducted for each participant, following a general linear model (GLM). The estimated model included two regressors for the monetary distributions (advantageous condition, A-IE; disadvantageous condition, DA-IE) and one for the response event (rating). All regressors were convolved with the hemodynamic response function. Regressors for the conditions of interest (A-IE and DA-IE distributions) were modeled with the duration of the event (5.5 s), whereas the rating regressor was modeled with the duration of the response. In addition, we included the magnitude of the inequality (i.e., the absolute value of the difference between the share for self and the other, denoted as "Difference") as a parametric modulator for both A-IE and DA-IE events. Six movement parameters were also included in the model as nuisance regressors.

Prior to testing the hypotheses, we carried out whole brain analyses with cluster-wise inference for A-IE > DA-IE and DA-IE > A-IE to assess if the brain processes advantageous and DA-IE differently. We also computed how the level of inequality affects processing in A-IE and DA-IE separately by contrasting [A-IE × Difference > baseline] and [DA-IE × Difference > baseline], whereby "baseline" refers to the built-in feature of SPM 12 that computes the average activity over the entire experiment excluding the events of interest. Results of these first-level contrasts (reported in the Supplemental Table 2A–B) reveal significant activations in somatosensory and (pre)motor areas of the brain, which we interpret to indicate individuals' preparedness in anticipation of assigning a liking score through pushing a button during the next phase.

To address Hypotheses 1 and 2a, contrasts were calculated according to power role, namely, [$\text{Teammate}_{\text{DA-IE}\times\text{Difference}>\text{baseline}} > \text{Leader}_{\text{DA-IE}\times\text{Difference}>\text{baseline}}$], [$\text{Leader}_{\text{DA-IE}\times\text{Difference}>\text{baseline}} > \text{Teammate}_{\text{DA-IE}\times\text{Difference}>\text{baseline}}$], and the same contrasts substituting DA-IE with A-IE. These contrasts reveal the interaction effect of role (teammate/leader) and the magnitude of inequality on the BOLD signal when the processed inequality is respectively disadvantageous or advantageous.

Finally, we tested the effect of the power role on the BOLD signal given a participant's SVO (i.e., Hypothesis 2b). We repeat the second level contrast separately for proselfs and prosocials: [$\text{Proself Teammate}_{\text{AIE}\times\text{Difference}>\text{baseline}} > \text{Proself Leader}_{\text{AIE}\times\text{Difference}>\text{baseline}}$] and [$\text{Prosocial Teammate}_{\text{AIE}\times\text{Difference}>\text{baseline}} > \text{Prosocial Leader}_{\text{AIE}\times\text{Difference}>\text{baseline}}$].

Postscanning: Concluding the Creativity Task and Postexperimental Questionnaire

After scanning, participants received a final questionnaire to complete the remaining part of the creativity task (i.e., evaluating ideas, selecting and submitting the best one, and, for leaders, deciding on the compensation). To test the effectiveness of the leader/teammate prime, six questions queried participants' feelings of being dependent, powerful, important, powerless, subordinate, and influential. As in previous research (Pauwels, 2021), these questions were inserted randomly into the 20 items of the positive and negative affect schedule in order to minimize experimenter demand effects.

[4] In the experimental design we opted for consistency between the stimulus/response options and the button press (i.e., left button press always corresponded to liking Scores 1 and 2 and right button press to Scores 3 and 4). This makes it easier for participants to respond to stimuli and therefore reduces the likelihood that they make mistakes. We acknowledge this may have introduced a bias in the motor- and somatosensory cortex when computing the contrasts between A-IE and DA-IE as subjects like A-IE (more use of right button) more than DA-IE (more use of left button; see Supplemental Table 2). However, it is unlikely that this affected the findings and conclusions regarding our hypotheses in the A-IE processing phase because a systematic left/right bias, if any, would affect subjects in the high- and low power conditions in the same way, and, therefore, would not affect our focal contrasts.

Results

To obtain a reliable assessment of the effectiveness of the power manipulation, we aggregated the responses on the six relevant items inserted in the postexperimental questionnaire. A *t* test corroborates that leaders felt more powerful than teammates ($t = 2.7$, $p < .01$, Cohen's $d = 0.854$). Such a value of Cohen's *d* can be considered to be relatively large (Sawilowski, 2009). Histograms comparing the mean scores of each item for leaders and teammates are shown in Supplemental Information, Figure 1A.[5]

Behavioral Data

To visualize the influence of power on inequity aversion, we plot the mean liking scores (+/− standard error) of all individuals against the difference in reward to self and other (See Figure 2). This shows that leaders dislike DA-IE more than followers. For A-IE, however, we observe the opposite (leaders like it more than teammates), although the difference is less pronounced, and the variance is greater (as shown by the larger error bars in A-IE compared to DA-IE). Figure 2B shows the same plots decomposed according to SVO. Here a three-way interaction between inequality, power, and SVO manifests itself in A-IE: as inequality increases, proselfs leaders do not become more inequity averse (they retain high liking scores). For large differences in rewards (>100), proself leaders become less inequity averse, showing an upward sloping curve which deviates substantially from the other three groups. Prosocial leaders, in contrast, show a downward sloping curve (disliking inequality), much like teammates. No such interactive effect is seen in DA-IE.

To analyze the data statistically, we conduct general least square (GLS) regressions on panel data using the "xtreg" function in STATA 15 StataCorp, 2017), an estimation technique that accounts for the fact that we collected 18 liking scores per individual in the DA-IE condition and 18 in the A-IE condition. Because these scores cannot be assumed independent, we use a random effects model that accounts for clustering of observations per individual and report robust standard errors. To avoid interpreting four-way interactions (between condition, power, SVO, and the magnitude of the inequality), we estimate four models to estimate the liking scores in DA-IE and A-IE separately and report model parameters (unstandardized B) with robust standard errors (see Table 1).

$$\text{(Models 1 \& 5)}\quad \text{Liking} = B1(\text{Difference}) + B2(\text{Power}) + B3(\text{SVO}) + \text{Constant},$$

$$\text{(Models 2 \& 6)}\quad \text{Liking} = B1(\text{Difference}) + B2(\text{Power}) + B3(\text{SVO}) + B4(\text{Power} \times \text{SVO}) + \text{Constant},$$

$$\text{(Model 3 \& 7)}\quad \text{Liking} = B1(\text{Difference}) + B2(\text{Power}) + B3(\text{SVO}) + B5(\text{Difference} \times \text{Power}) + \text{Constant},$$

$$\text{(Model 4 \& 8)}\quad \text{Liking} = B1(\text{Difference}) + B2(\text{Power}) + B3(\text{SVO}) + B4(\text{Power} \times \text{SVO}) + B5(\text{Difference} \times \text{Power}) + B6(\text{Difference} \times \text{SVO}) + B7(\text{Difference} \times \text{Power} \times \text{SVO}) + C.$$

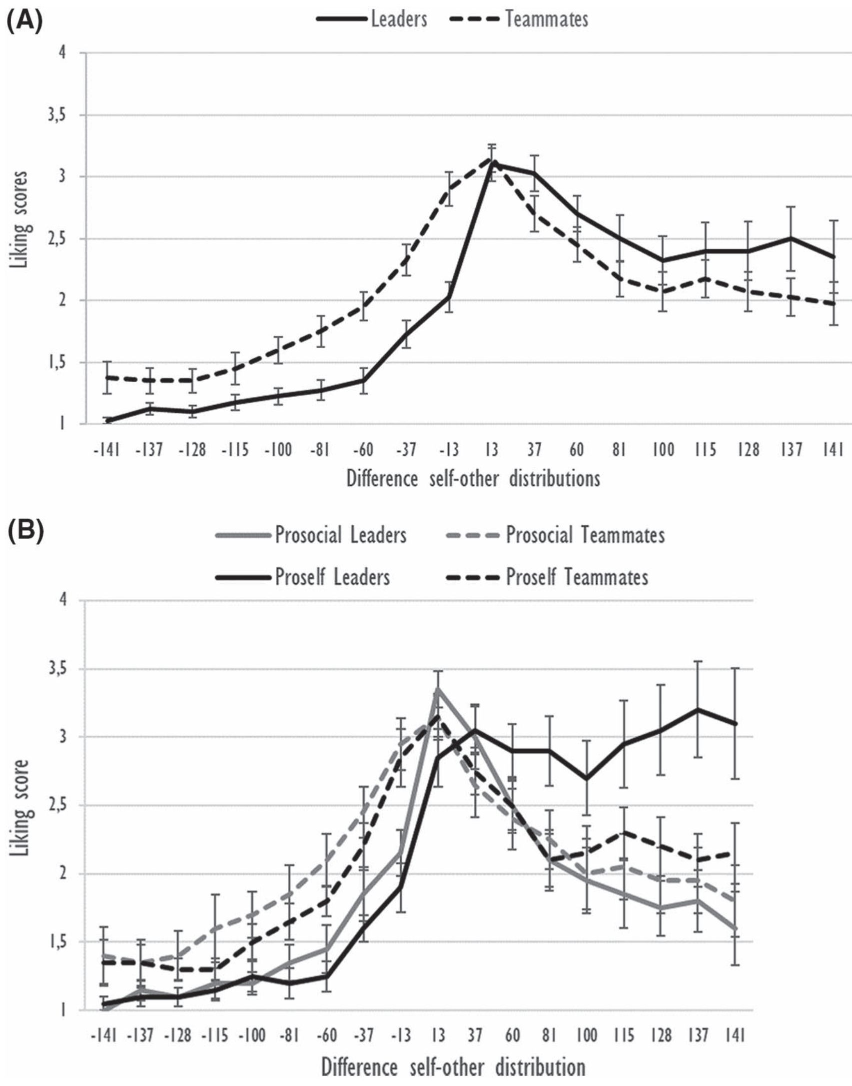
Liking Scores in DA-IE

Models 1–4 relate to Hypothesis 1, testing the effect of power on DA-IE aversion. The main effects of power (coded 1 for high power, 0 for control) and SVO (coded 1 for prosocial, 0 for proself) are tested in model 1, controlling for the effect of varying sizes of inequality (denoted "Difference"). We note that, the larger the difference in rewards between self and other, the more negative the liking scores ($B1 = -0.01$, $p < .001$). Individuals in a powerful position

[5] These results are similar to the manipulation checks of two other studies. In Pauwels (2021), participants ($N = 144$) primed with a leader role felt significantly more powerful, more influential, more important, less dependent, and less subordinate than low power-participants (see Supplemental Material 1B). In another (as of yet unpublished) behavioral study ($N = 208$) this procedure led leaders to feel more influential, more powerful, more in control of the situation, and less subordinate than teammates (see Supplemental Figure 1C). Proself leaders especially kept more of the money awarded for the creativity task to themselves compared to prosocial leaders.

Figure 2

Plots Depicting the Relationship Between the Liking Scores (Y-Axis) and the Difference in Rewards Between Self and Others (X-Axis)



Note. Each individual rated 18 different reward distributions that resulted in nine different levels of disadvantageous inequality (DA-IE), and 18 reward distributions that yielded nine levels of advantageous inequality (A-IE). A: Plot illustrating the difference between teammates (n = 20) and leaders (n = 20); B: plot decomposed for proself teammates (n = 10), prosocial teammates (n = 10), proself leaders (n = 10) and prosocial leaders (n = 10).

indicated overall lower liking scores ($B2 = -0.45$, $p < .001$). This effect is robust when controlling for self-reported social status, age and gender (Supplemental Table 3) and is consistent with Hypothesis 1. Model 1 furthermore indicates no effect of SVO.

Models 2–4 (DA-IE) report all interaction effects. We had no hypotheses regarding any interaction effects for DA-IE. We do note, however, that the negative effect of power on liking scores wanes as inequality increases (Model 3, $B5 = 0.004$, $p < .001$).

Table 1

GLS-Regression Estimates (Nonstandardized B) of the Impact of the Absolute Values of the Reward Differences for Self and Other (Denoted "Difference"), Power and SVO on Liking Scores in the Disadvantageous Inequality (DA-IE) Condition and in the Advantageous (A-IE) Condition

In DA-IE condition	Model 1	Model 2	Model 3	Model 4
Difference (*B*1)	−0.01*** (0.001)	−0.01*** (0.001)	−0.01*** (0.001)	−0.01*** (0.002)
Power (*B*2)	−0.45*** (0.11)	−0.41*** (0.12)	−0.85*** (0.19)	−0.87*** (0.23)
SVO (*B*3)	0.13 (0.11)	0.17 (0.19)	0.13 (0.11)	0.25 (0.27)
Power × SVO (*B*4)		−0.07 (0.22)		0.05 (0.37)
Difference × Power (*B*5)			0.004*** (0.001)	0.01** (0.002)
Difference × SVO (*B*6)				−0.001 (0.002)
Difference × Power × SVO (*B*7)				−0.001 (0.003)
Constant	2.53*** (0.11)	2.51*** (0.13)	2.73*** (0.14)	2.67*** (0.20)
Wald chi-square[a]	177.41***	180.65***	225.21***	259.72***
In A-IE condition	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Difference (*B*1)	−0.01*** (0.001)	−0.01*** (0.001)	−0.01*** (0.001)	−0.01*** (0.002)
Power (*B*2)	0.28 (0.20)	0.59* (0.27)	0.07 (0.21)	−0.17 (0.25)
SVO (*B*3)	−0.44* (0.20)	−0.13 (0.26)	−0.44* (0.20)	0.06 (0.23)
Power × SVO (*B*4)		−0.62 (0.39)		0.50 (0.34)
Difference × Power (*B*5)			0.002 (0.003)	0.01* (0.004)
Difference × SVO (*B*6)				−0.001 (0.002)
Difference × Power × SVO (*B*7)				−0.01** (0.004)
Constant	3.15*** (0.16)	3.00*** (0.18)	3.25*** (0.15)	3.00*** (0.15)
Wald chi-square	46.53***	48.78***	67.69***	256.67***

Note. GLS = general least square; SVO = social value orientation. Power is scored 1 for leaders, 0 for teammates and SVO is scored 1 for prosocials, 0 for proselfs.
[a] For all models, $N = 720$ observations clustered on 40 individuals.
* $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed), robust standard errors in parenthesis.

Liking Scores in A-IE

Models 5–8 relate to Hypothesis 2a and Hypothesis 2b, testing the effect of power on AI-E aversion and its interaction with SVO. Main effects of power and SVO are tested with Model 5, again controlling for varying sizes of inequality which significantly affect liking scores ($B1 = -0.01$, $p < .001$). In this model, power does not affect liking scores on top and beyond the effect of reward differences, contradicting Hypothesis 2a. SVO in Model 5 exerts some effect: prosocials, relative to proselfs, indicate overall lower liking scores ($B3 = -0.44$, $p = .027$, Bonferroni corrected $p = .081$). Although we did not specify a hypothesis on the main effect of SVO, the greater A-IE aversion for prosocials is consistent with the SVO construct.

Models 6–8 report on the interaction effects that relate to Hypothesis 2b. Of particular interest is Model 8, which tests the postulated three-way interaction between power, SVO and reward Difference. Corroborating Hypothesis 2b, power appears to influence liking scores for increasingly A-IE only for proselfs ($B7 = -0.01$, $p = .003$, Bonferroni corrected $p < .021$).

As a robustness check, we compare the average liking scores for distributions that are nearly equal (with absolute differences between rewards for self and others less than 60) and distribution that are very unequal (absolute difference between 81 and 141, see Supplemental Table 4). This substantiates that proselfs with high power do not become inequity averse, even when A-IE becomes very unequal.

Neural Correlates of Inequity Aversion in DA-IE

Referring to Hypothesis 1, contrasting the parametrically modulated DA-IE (>baseline) of leaders and teammates (FWE corrected $p < .05$, see Table 2) does not show the hypothesized insula- or amygdala activation as a function of

Table 2
Clusters of Voxels Significantly Activated in the Whole Brain During Reward Pair Presentation for the Parametrically Modulated Effect of the Size of Inequality

Contrasts	Anatomical label (functional region)	BA	Side	*x*	*y*	*z*	Cluster size	*t*	*p*
A-IE × Difference > baseline									
Teammate > Leader	Precentral gyrus (FEF, adjacent to dlPFC)	6	R	26	−6	50	123	5.71	.008
	Precuneus	5	R	14	−58	60	117	4.71	.010
	Precuneus		L	−10	−56	52	132	4.59	.005
Proself Teammate > Proself Leader	Superior frontal gyrus (dlPFC)	6	R	22	−4	50	134	5.73	<.001
	Middle frontal gyrus (dlPFC)	46	L	−42	38	32	80	5.11	.012
DA-IE × Difference > baseline									
Teammate > Leader	Superior Parietal Lobule	40	L	−42	−42	48	149	4.37	.029
	Superior Parietal Lobule	2	R	36	−44	68	169	4.19	.017

Note. Peak voxel-level uncorrected $p < .001$, cluster-level FWE-corrected $p < .05$. Anatomical labels for the peak activation coordinates were determined using the maximum probability maps of the Harvard-Oxford cortical and subcortical structural atlases. Coordinates in MNI space. Voxel size of 2.0 × 2.0 × 2.0 mm. BA = Brodmann area. L = left; R = right; DA-IE = disadvantageous inequality condition; A-IE = advantageous inequality condition. Difference = absolute value of the difference in reward size for self and other; MNI = Montreal Neurologic Institute; FEF = frontal eye field; dlPFC = dorsolateral PFC; FWE = family-wise error; PFC = prefrontal cortex.

increasing inequality in high power participants. Instead, teammates (relative to leaders) show more activation in regions labeled as left and right superior parietal lobule (SPL) when confronted with increasing DA-IE, for which we had no a priori hypothesis.

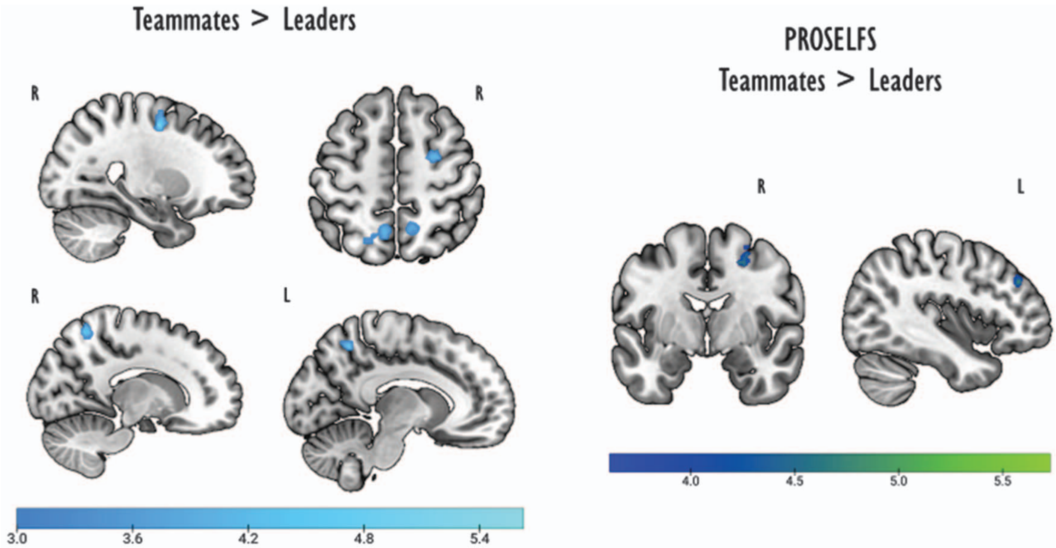
Unfortunately, the whole brain contrast maps suffered from signal disturbance around the ventricles, resulting in loss of data for functional amygdala activation. Hence, we cannot infer to what extent the amygdala was involved in processing disadvantageous inequity and whether or not it was affected by power. To further explore the possible involvement of the insula, we conducted region of interest analyses using the Marsbar tool in SPM (Release 0.44). Neither for the anterior nor posterior parts of the insula did significant differences between groups emerge.

Neural Correlates of Inequity Aversion in A-IE

Referring to Hypothesis 2a, we observe significant clusters of activation when contrasting A-IE (>baseline) between teammates and leaders (see Figure 3A). Consistent with the behavioral data indicating that leaders like A-IE better than teammates, the neural data show that, relative to teammates, leaders show attenuated activation in three areas that are known from previous research to be involved in higher level cognition and perspective taking: (a) the right precentral gyrus (mni coordinates: 26 –6 50) with the peak coordinates overlapping with the frontal eye field (FEF), an area bordering on the posterior end of the dlPFC (Vernet et al., 2014); (b) the left and right precuneus (mni coordinates: −10 –56 52 and 14 –58 60) an area of the parietal cortex which is nearly always activated in tasks that require social interaction, especially if asked to take the perspective of others (Cavanna & Trimble, 2006). None of the contrasts show differences in activation within the hypothesized regions of the TPJ or dmPFC.

To test Hypothesis 2b, we repeat the same (whole brain) analysis separately for prosocials ($n = 20$) and proselfs ($n = 20$). We find no difference in the activation pattern when contrasting prosocial leaders and teammates. As for proself leaders, relative to proself teammates, increasing reward differences in A-IE correlate

Figure 3
Group Results for the A-IE (>Baseline) Contrast comparing Teammates > Leaders



Note. A-IE = advantageous inequality; FWE = family-wise error. Left: the results from the whole sample ($N = 40$) shows lower activation in the right precentral gyrus and bilateral precuneus for leaders. Right: results from the proself participants ($N = 20$) reveal decreased activation for leaders in the right superior frontal gyrus and left middle frontal gyrus. Scales reflect peaks of significant t-values ($p < .05$, FWE-corrected for multiple comparisons).

negatively with the BOLD signal in two regions that are typically included in the dlPFC (Figure 3B): (a) the lateral and posterior part of the right superior frontal gyrus (mni coordinates: 22 −4 50); (b) the anterior, dorsolateral part of the left middle frontal gyrus (mni coordinates: −42 38 32). Thus Hypothesis 2B is partially corroborated: the neural data are consistent with a reduction of cognitive control for proselfs, but show no evidence for increased cognitive control in prosocials.

Discussion

Both the academic literature and popular writings abound with claims that power changes people's attitudes toward greed and wealth accumulation (e.g., Piff et al., 2012). Leadership positions supposedly create social distance (Magee & Smith, 2013) and legitimize entitlements (Akerlof & Kranton, 2000) which in turn may alter how violations of the equality norm are perceived: tolerance for disadvantageous inequity weakens, while advantageous inequity becomes taken for granted. We set out to test whether this is unambiguously so or whether it depends on a person's value system which serves to manage interpersonal interactions. With fMRI, we pinpointed the regions in the brain that show increased activation when prosocial and proself individuals are given leadership positions and asked to express how much they (dis)like hypothetical unequal divisions of resources between themselves and an anonymous subordinate partner.

Summary of Main Results

Statistical analyses were conducted to test three a priori hypotheses regarding behavioral- and neural responses to inequity. First, for DA-IE, the behavioral data corroborate Hypothesis 1: power significantly lowered participants' liking scores, thereby increasing DA-IE aversion, and this was true irrespective of individual differences in values. From the corresponding neural data, however, no conclusions can be drawn. We neither find evidence that DA-IE is associated with increased activation in the amygdala (due to signal loss), nor the insula.

Second, for A-IE, the behavioral data show a trend (Figure 2A) supporting the hypothesis that power decreases A-IE aversion (Hypothesis 2a), but this is not statistically significant. The corresponding neural data partially corroborates the hypothesized effects of power: consistent with Hypothesis 2a, leaders (compared to teammates) show decreased activity in one of the hypothesized mentalizing regions (bilateral precuneus). We do not find any difference in activity within the other two hypothesized mentalizing regions (dmPFC and TPJ). Neither do we find activity in the postulated dlPFC. However, high power correlates with decreased activity in the frontal eye field (FEF), which lies adjacent to the dlPFC and is anatomically and functionally related to it (Vernet et al., 2014).

Third, we further hypothesized that in A-IE the effect of power would be more pronounced for proself leaders than for prosocial leaders. (Hypothesis 2b). This is corroborated by the behavioral data, but only for larger inequalities. The corresponding neural data support the hypothesis that cognitive control is reduced for proself leaders (compared to proself teammates), as shown by the attenuated signal in the dlPFC. However, unlike expected, increased dlPFC is not observed for prosocial leaders (compared to prosocial teammates).

Taken together, these results contribute to the existing debate on the corrupting effect of power in two important ways. First, DA-IE aversion—a universally observed phenomenon also among nonhuman primates—is aggravated for leaders (compared to teammates), irrespective of individual differences in SVO. Second, power affects both the behavioral and neural response to A-IE. Only for proself leaders (compared to proself teammates) is growing inequality not matched by growing inequity aversion. Their response indicates that they like it, which matches their intrinsic self-regarding social preference. The neural data points to reduced dlPFC activation.

Relation to Previous Literature

The neural activation pattern corresponding to decreased A-IE aversion in leaders (Hypothesis 2a), and only in proself leaders (Hypothesis 2b) accords with several brain-function relationships that have previously been described in the literature.

Consistent with Hypothesis 2a, the data reported here indicate lower activity in the FEF and precuneus for leaders evaluating A-IE. The role of the FEF beyond controlling eye movements is becoming increasingly noted in cognitive tasks such as attentional orienting, consciousness, perceptual performance, inhibitory control, and decision-making (Vernet et al., 2014). fMRI experiments show FEF involvement during tasks that require participants to mentally take on the frame of reference of another person when viewing a scene (Wallentin et al., 2008), or during tasks that require inhibitory control (Muggleton et al., 2010), such as resisting a potentially deleterious reward (Padmala & Pessoa, 2010). Viewing a reward distribution that puts oneself at a monetary advantage may be desirable from a self-centered point of view, but deleterious for one's relationships. Thus, by taking on the viewpoint of another person with whom a reward is shared, and by inhibiting the desire for personal advantage, the FEF might play a role in assessing the outcome for others. For leaders this propensity seems to be more easily compromised than for teammates.

This interpretation is strengthened by finding coactivation in bilateral precuneus, a region associated with mental imagery, attention and mentalizing. More specifically, in the left hemisphere, the coordinates of the peak activity (−10 −56 52, Table 2) are very similar to those reported in the study by Muscatell et al. (2012) linking precuneus activation to differences in mentalizing between high- and low social class college students. In general, this part of the precuneus is believed to function in differentiating, attributing, and matching first- and third person perspectives (Cavanna & Trimble, 2006; Farrer & Frith, 2002; Vogeley et al., 2001). Some authors have proposed that the precuneus not only becomes activated when matching expectations of the self- and other, but also when matching one's own (moral) standards with expected outcomes (Emonds et al., 2013; Ruz & Tudela, 2011). Again, this appears to be less important for leaders than for teammates.

Consistent with Hypothesis 2b, the data also indicated attenuated dlPFC activity for proselfs leaders confronted with A-IE, which fit with the proposed hypothesis that they have a reduced need for cognitive control to restrain selfish impulses. In particular, the region in the middle frontal gyrus with peak coordinates [−42 38 32] overlaps with the neural correlates of norm compliance reported in Buckholtz and Marois (2012) and Spitzer et al. (2007). This particular region appears to be more activated when fair decision-making is driven by a punishment threat. The finding that it also shows reduced activity when proself (but not prosocial) leaders are confronted with A-IE could indicate that proself leaders do not consider inequality a violation of the norm. Instead, they might regard inequality as more or less taken-for-granted, reducing potential punishment and reputation concerns.

In contrast, the data show that powerful prosocials do not become less inequity averse with growing A-IE. As they do experience a conflict between their inner compass and their feelings of entitlement, we would expect their A-IE aversion to be no longer intuitive, and they would need to exert more cognitive control to curb the temptation associated with the entitlements of being a leader. We did not, however, observe the hypothesized different neural activation pattern between prosocial leaders and teammates. This could either mean that their values protect prosocials from the effect of power, but the null finding may also very well be the result of other limitations, which we address below.

Limitations and Concluding Remarks

In addition to the more typical constraints of experimental research and neuroimaging (such as the artificial environment and the need to collect repeated measures to obtain sufficient power), there are a number of other concerns worthy of further discussion. First, the small sample size ($n = 40$) implies low statistical power and high proneness to Type II errors, meaning that the null findings we report should be interpreted as inconclusive. Low power becomes especially problematic when comparing the neural data of prosocial leaders versus teammates, and proself leaders versus teammates ($n = 10$ in each of the four categories), which might be one of the reasons why we did not find the hypothesized difference in cognitive control activation between prosocial leaders and teammates.

Second, the power manipulation task as well may have contributed to Type II errors and inconclusive results. While the manipulation appeared to have been successful as indicated by the effect size, assessing the strength of the manipulation separately for prosocials and proselfs, reveals that the power induction was more effective for proselfs ($t = 2.13$, $p = .047$, Cohen's $d = 0.952$) than for prosocials ($t = 1.45$, $p = .16$, Cohen's $d = 0.65$). If assuming a leader position was more salient for proselfs, while prosocials remained mostly unaffected, it would explain why both the behavioral and the neural data for A-IE reveal differences between leaders and teammates only for proselfs. If prosocials did not sufficiently respond to being primed with power, the sample size of successfully primed individuals may have been too small to reveal, for example, the hypothesized insula response to DA-IE.

To avoid priming, many behavioral- and neuroeconomic experiments have manipulated structural power by relying, for example, on economic games (dictator- and ultimatum games) that affect both the participant's and her interaction partner's payoffs. Social preferences, derived from actual decisions with real monetary consequences, have been shown to have much predictive value in the context of A-IE (e.g., Bruhin et al., 2019; Hedegaard et al., 2021). Compared to the emotional, or "hot" response of being treated unfairly in an ultimatum game, rating the desirability of an unequal monetary split would be more of a "cold" response, eliciting less disgust. This is a plausible reason why we did not observe the hypothesized insula activation in the first-level contrast of DA-IE > A-IE (e.g., Gao et al., 2018; Güroğlu et al., 2014; Hu et al., 2016; Yang et al., 2019 for meta-analysis), nor any deactivation in the brain's reward system (in the reverse contrast) associated with reduced pleasure resulting from A-IE aversion (ventral striatum and ventromedial PFC, see Tricomi et al., 2010).

Third, because desirability ratings have no pecuniary consequences for the participant, there has been a general concern that subjectively stated preferences (rating hypothetical reward distributions) are "cheap" and that this may not reflect actual decision utilities (Beshears et al., 2008). However, in the present study, the aim was not to estimate utility (the subjective value of equality), but rather to find out how the subjective preference for equality changes in response to being in a position of power. For this purpose, normative judgments may be just as valid (Carlsson, 2010). For example, in a natural field experiment, the influence of normative information on donations to a public good remained the same both for actual- and hypothetical decisions (Alpizar et al., 2008). While more research comparing revealed preferences in judgment- versus choice tasks is needed, relying on hypothetical vignettes is common practice in studies investigating the neural correlates of moral judgments (e.g., Yoder & Decety, 2014) or justice sensitivity (e.g., Buckholtz & Marois, 2012; Haruno & Frith, 2010).

Finally, the correlational nature of fMRI data call for caution to not overinterpret or overgeneralize these results. By testing hypotheses based on existing theory and previously published reports, we tried to minimize the reverse inference problem. Nevertheless, the picture is likely more complicated, as the brain regions we identified are all parts of interdependent networks that respond to multiple inputs, leaving room to further explore with connectivity analyses the intricacies of judging and responding to inequity in different contexts.

To conclude, by experimentally inducing power asymmetry when people are evaluating resource distributions between themselves and an anonymous partner, this study contributes in a novel way to the literature on inequity aversion. At the same time, it provides insights into the possible reasons why and when power corrupts. Based on the results reported here, we conclude that power legitimizes advantageous inequity and that this is associated with weakened activity in a number of brain regions supporting cognitive control and mentalizing, namely, the FEF and precuneus. A proself value orientation exacerbates the preference for A-IE associated with power, which goes hand in hand with a reduced need for cognitive control and decreased activation of the dlPFC. The proposal that a prosocial value orientation combined with power requires more cognitive control to remain inequity averse to growing A-IE could not be corroborated but remains an intriguing possibility which invites further investigations.

References

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715–753. https://doi.org/10.1162/003355300554881

Alpizar, F., Carlsson, F., & Johansson-Stenman, O. (2008). Does context matter more for hypothetical than for actual contributions? Evidence from a natural field experiment. *Experimental Economics*, *11*, 299–314. https://doi.org/10.1007/s10683-007-9194-9

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, *411*(6835), 305–309. https://doi.org/10.1038/35077083

Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2008). How are preferences revealed? *Journal of Public Economics*, *92*(8–9), 1787–1794. https://doi.org/10.1016/j.jpubeco.2008.04.010

Bogaert, S., Boone, C., & Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology*, *47*(3), 453–480. https://doi.org/10.1348/014466607X244970

Brosnan, S. F., & de Waal, F. B. (2014). Evolution of responses to (un)fairness. *Science*, *346*(6207), Article 1251776. https://doi.org/10.1126/science.1251776

Bruhin, A., Fehr, E., & Schunk, D. (2019). The many faces of human sociality. Uncovering the distribution and stability of social preferences. *Journal of the European Economic Association*, *17*(4), 1025–1069. https://doi.org/10.1093/jeea/jvy018

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. https://doi.org/10.1038/nn.3087

Cappelen, A. W., Eichele, T., Hugdahl, K., Specht, K., Sørensen, E. O., & Tungodden, B. (2014). Equity theory and fair inequality: A neuroeconomic study. *Proceedings of the National Academy of Sciences*, *111*(43), 15368–15372. https://doi.org/10.1073/pnas.1414602111

Carlsson, F. (2010). Design of stated preference surveys: Is there more to learn from behavioral economics? *Environmental and Resource Economics*, *46*, 167–177. https://doi.org/10.1007/s10640-010-9359-4

Carter, R. M., & Huettel, S. A. (2013). A nexus model of the temporal-parietal junction. *Trends in Cognitive Sciences*, *17*(7), 328–336. https://doi.org/10.1016/j.tics.2013.05.007

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain: A Journal of Neurology*, *129*(3), 564–583. https://doi.org/10.1093/brain/awl004

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. https://doi.org/10.1016/j.neuron.2011.02.056

Chapman, J., Dean, M., Ortoleva, P., Snowberg, E., & Camerer, C. (2018) *Econographics*. CESifo Working Paper Series 7202. CESifo.

Chen, S., Lee-Chai, A. Y., & Bargh, J. A. (2001). Relationship orientation as a moderator of the effects of social power. *Journal of Personality and Social Psychology*, *80*(2), 173–187. https://doi.org/10.1037/0022-3514.80.2.173

Civai, C., Crescentini, C., Rustichini, A., & Rumiati, R. I. (2012). Equality versus self-interest in the brain: Differential roles of anterior insula and medial prefrontal cortex. *NeuroImage*, *62*(1), 102–112. https://doi.org/10.1016/j.neuroimage.2012.04.037

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. https://doi.org/10.1093/scan/nsu154

Cohn, A., Fehr, E., & Maréchal, M. A. (2014). Business culture and dishonesty in the banking industry. *Nature*, *516*(7529), 86–89. https://doi.org/10.1038/nature13977

Cornelissen, G., Dewitte, S., & Warlop, L. (2011). Are social value orientations expressed automatically? Decision making in the dictator game. *Personality and Social Psychology Bulletin*, *37*(8), 1080–1090. https://doi.org/10.1177/0146167211405996

Corradi-Dell'Acqua, C., Civai, C., Rumiati, R. I., & Fink, G. R. (2013). Disentangling self- and fairness-related neural mechanisms involved in the ultimatum game: An fMRI study. *Social Cognitive and Affective Neuroscience*, *8*(4), 424–431. https://doi.org/10.1093/scan/nss014

Côté, S., Kraus, M. W., Cheng, B. H., Oveis, C., van der Löwe, I., Lian, H., & Keltner, D. (2011). Social power facilitates the effect of prosocial orientation on empathic accuracy. *Journal of Personality and Social Psychology*, *101*(2), 217–232. https://doi.org/10.1037/a0023171

DeCelles, K. A., DeRue, D. S., Margolis, J. D., & Ceranic, T. L. (2012). Does power corrupt or enable? When and why power facilitates self-interested behavior. *Journal of Applied Psychology*, *97*(3), 681–689. https://doi.org/10.1037/a0026811

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

Declerck, C. H., & Boone, C. (2018). The neuroeconomics of cooperation. *Nature Human Behaviour*, *2*(7), 438–440. https://doi.org/10.1038/s41562-018-0387-3

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179. https://doi.org/10.1016/j.tics.2010.01.004

Emonds, G., Declerck, C. H., Boone, C., Seurinck, R., & Achten, R. (2013). Establishing cooperation in a mixed-motive social dilemma. An fMRI study investigating the role of social value orientation and dispositional trust. *Social Neuroscience*, *9*(1), 10–22. https://doi.org/10.1080/17470919.2013.858080

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

Evans, A. M., Dillon, K. D., & Rand, D. G. (2015). Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas. *Journal of Experimental Psychology: General*, *144*(5), 951–966. https://doi.org/10.1037/xge0000107

Farrer, C., & Frith, C. D. (2002). Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency. *NeuroImage*, *15*(3), 596–603. https://doi.org/10.1006/nimg.2001.1009

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. https://doi.org/10.1162/003355399556151

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Galinsky, A. D., Gruenfeld, D. H., & Magee, J. C. (2003). From power to action. *Journal of Personality and Social Psychology*, *85*(3), 453–466. https://doi.org/10.1037/0022-3514.85.3.453

Galinsky, A. D., Magee, J. C., Inesi, M. E., Gruenfeld, D. H., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychological Science*, *17*(12), 1068–1074. https://doi.org/10.1111/j.1467-9280.2006.01824.x

Gao, X., Yu, H., Sáez, 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. https://doi.org/10.1073/pnas.1802523115

Guinote, A. (2017). How power affects people: Activating, wanting, and goal seeking. *Annual Review of Psychology*, *68*(1), 353–381. https://doi.org/10.1146/annurev-psych-010416-044153

Güroğlu, 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*, Article 291. https://doi.org/10.3389/fpsyg.2014.00291

Haruno, M., & Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nature Neuroscience*, *13*(2), 160–161. https://doi.org/10.1038/nn.2468

Haruno, M., Kimura, M., & Frith, C. D. (2014). Activity in the nucleus accumbens and amygdala underlies individual differences in prosocial and individualistic economic choices. *Journal of Cognitive Neuroscience*, *26*(8), 1861–1870. https://doi.org/10.1162/jocn_a_00589

Hedegaard M., Kerschbamer R., Müller D., & Tyran J. (2021). Distributional preferences explain individual behavior across games and time. *Games and*

Economic Behavior, *128*, 231–255. https://doi.org/10.1016/j.geb.2021.05.003

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767–1770. https://doi.org/10.1126/science.1127333

Hsu, M., Anen, C., & Quartz, S. R. (2008). The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science*, *320*(5879), 1092–1095. https://doi.org/10.1126/science.1153651

Hu, J., Blue, P. R., Yu, H., Gong, X., Xiang, Y., Jiang, C., & Zhou, X. (2016). Social status modulates the neural response to unfairness. *Social Cognitive and Affective Neuroscience*, *11*(1), 1–10. https://doi.org/10.1093/scan/nsv086

Hu, J., Cao, Y., Blue, P. R., & Zhou, X. (2014). Low social status decreases the neural salience of unfairness. *Frontiers in Behavioral Neuroscience*, *8*, Article 402. https://doi.org/10.3389/fnbeh.2014.00402

Inesi, M. E., Gruenfeld, D. H., & Galinsky, A. D. (2012). How power corrupts relationships: Cynical attributions for others' generous acts. *Journal of Experimental Social Psychology*, *48*(4), 795–803. https://doi.org/10.1016/j.jesp.2012.01.008

Joshi, P. D., & Fast, N. J. (2013). I am my (high-power) role: Power and role identification. *Personality and Social Psychology Bulletin*, *39*(7), 898–910. https://doi.org/10.1177/0146167213485443

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review*, *110*(2), 265–284. https://doi.org/10.1037/0033-295X.110.2.265

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. https://doi.org/10.1126/science.1129156

Kraus, M. W., Chen, S., & Keltner, D. (2011). The power to be me: Power elevates self-concept consistency and authenticity. *Journal of Experimental Social Psychology*, *47*(5), 974–980. https://doi.org/10.1016/j.jesp.2011.03.017

Laurin, K., Fitzsimons, G. M., Finkel, E. J., Carswell, K. L., vanDellen, M. R., Hofmann, W., Lambert, N. M., Eastwick, P. W., Fincham, F. D., & Brown, P. C. (2016). Power and the pursuit of a partner's goals. *Journal of Personality and Social Psychology*, *110*(6), 840–868. https://doi.org/10.1037/pspi0000048

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, *58*(1), 259–289. https://doi.org/10.1146/annurev.psych.58.110405.085654

Liljeholm, M., Dunne, S., & O'Doherty, J. P. (2014). Anterior insula activity reflects the effects of intentionality on the anticipation of aversive stimulation. *The Journal of Neuroscience*, *34*(34), 11339–11348. https://doi.org/10.1523/JNEUROSCI.1126-14.2014

Loewenstein, G. F., Thompson, L., & Bazerman, M. H. (1989). Social Utility and Decision Making in Interpersonal Contexts. *Journal of Personality and Social Psychology*, *57*(3), 426–441. https://doi.org/10.1037/0022-3514.57.3.426

Magee, J. C., & Smith, P. K. (2013). The social distance theory of power. *Personality and Social Psychology Review*, *17*(2), 158–186. https://doi.org/10.1177/1088868312472732

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*(5695), 503–507. https://doi.org/10.1126/science.1100907

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure & Function*, *214*(5–6), 655–667. https://doi.org/10.1007/s00429-010-0262-0

Mischel, W., & Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annual Review of Psychology*, *49*(1), 229–258. https://doi.org/10.1146/annurev.psych.49.1.229

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience and Biobehavioral Reviews*, *65*, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020

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. https://doi.org/10.1016/j.neuron.2012.05.021

Muggleton, N. G., Chen, C. Y., Tzeng, O. J. L., Hung, D. L., & Juan, C. H. (2010). Inhibitory control and the frontal eye fields. *Journal of Cognitive Neuroscience*, *22*(12), 2804–2812. https://doi.org/10.1162/jocn.2010.21416

Murphy, R. O., & Ackermann, K. A. (2014). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, *18*(1), 13–41. https://doi.org/10.1177/1088868313501745

Muscatell, K. A., Morelli, S. A., Falk, E. B., Way, B. M., Pfeifer, J. H., Galinsky, A. D., Lieberman, M. D., Dapretto, M., & Eisenberger, N. I. (2012). Social status modulates neural activity in the mentalizing network. *NeuroImage*, *60*(3), 1771–1777. https://doi.org/10.1016/j.neuroimage.2012.01.080

Padmala, S., & Pessoa, L. (2010). Interactions between cognition and motivation during response inhibition. *Neuropsychologia*, *48*(2), 558–565. https://doi.org/10.1016/j.neuropsychologia.2009.10.017

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), Article 15964. https://doi.org/10.1038/ncomms15964

Pauwels, L. (2021) *Power, social values and their role directing 'the orchestra of heuristics' in social dilemmas* [PhD dissertation]. University of Antwerp.

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron*, *48*(2), 175–187. https://doi.org/10.1016/j.neuron.2005.09.025

Piff, P. K., Kraus, M. W., Côté, S., Cheng, B. H., & Keltner, D. (2010). Having less, giving more: The influence of social class on prosocial behavior. *Journal of Personality and Social Psychology*, *99*(5), 771–784. https://doi.org/10.1037/a0020092

Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., & Keltner, D. (2012). Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(11), 4086–4091. https://doi.org/10.1073/pnas.1118373109

Righetti, F., Luchies, L. B., van Gils, S., Slotter, E. B., Witcher, B., & Kumashiro, M. (2015). The prosocial versus proself power holder: How power influences sacrifice in romantic relationships. *Personality and Social Psychology Bulletin*, *41*(6), 779–790. https://doi.org/10.1177/0146167215579054

Ruz, M., & Tudela, P. (2011). Emotional conflict in interpersonal interactions. *NeuroImage*, *54*(2), 1685–1691. https://doi.org/10.1016/j.neuroimage.2010.08.039

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

Sawaoka, T., Hughes, B. L., & Ambady, N. (2015). Power Heightens Sensitivity to Unfairness Against the Self. *Personality and Social Psychology Bulletin*, *41*(8), 1023–1035. https://doi.org/10.1177/0146167215588755

Sawilowski, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*, Article 26. https://doi.org/10.22237/jmasm/1257035100

Schilke, O., Reimann, M., & Cook, K. S. (2015). Power decreases trust in social exchange. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(42), 12950–12955. https://doi.org/10.1073/pnas.1517057112

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, 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

StataCorp. (2017). *Stata statistical software: Release 15*.

Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science*, *19*(4), 339–347. https://doi.org/10.1111/j.1467-9280.2008.02091.x

Tanaka, T., Yamamoto, T., & Haruno, M. (2017). Brain response patterns to economic inequity predict present and future depression indices. *Nature Human Behaviour*, *1*(10), 748–756. https://doi.org/10.1038/s41562-017-0207-1

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

van Kleef, G. A., Oveis, C., van der Löwe, I., LuoKogan, A., Goetz, J., & Keltner, D. (2008). Power, distress, and compassion: Turning a blind eye to the suffering of others. *Psychological Science*, *19*(12), 1315–1322. https://doi.org/10.1111/j.1467-9280.2008.02241.x

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientations. *Journal of Personality and Social Psychology*, *77*(2), 337–349. https://doi.org/10.1037/0022-3514.77.2.337

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*(3), 829–858. https://doi.org/10.1002/hbm.20547

Vernet, M., Quentin, R., Chanes, L., Mitsumasu, A., & Valero-Cabré, A. (2014) Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. *Frontiers in Integrative Neuroscience*, *8*, Article 66. https://doi.org/10.3389/fnint.2014.00066

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., Maier, W., Shah, N. J., Fink, G. R., & Zilles, K. (2001). Mind reading: Neural mechanisms of theory of mind and self-perspective. *NeuroImage*, *14*(1), 170–181. https://doi.org/10.1006/nimg.2001.0789

Wallentin, M., Roepstorff, A., & Burgess, N. (2008). Frontal eye fields involved in shifting frame of reference within working memory for scenes. *Neuropsychologia*, *46*(2), 399–408. https://doi.org/10.1016/j.neuropsychologia.2007.08.014

Yamagishi, T., Matsumoto, Y., Kiyonari, T., Takagishi, H., Li, Y., Kanai, R., & Sakagami, M. (2017). Response time in economic games reflects different types of decision conflict for prosocial and proself individuals. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(24), 6394–6399. https://doi.org/10.1073/pnas.1608877114

Yang, Z., Zheng, Y., Yang, G. C., Li, Q., & Liu, X. (2019). Signatures of cooperation enforcement and violation: A coordinate-based meta-analysis. *Social Cognitive and Affective Neuroscience*, *14*(9), 919–931. https://doi.org/10.1093/scan/nsz073

Yoder, K. J., & Decety, J. (2014). The good, the bad, and the just: Justice sensitivity predicts neural response during moral evaluation of actions. *Journal of Neuroscience*, *34*(12), 4161–4166. https://doi.org/10.1523/JNEUROSCI.4648-13.2014

Zhang, S., & Li, C. S. (2012). Functional connectivity mapping of the human precuneus by resting state fMRI. *NeuroImage*, *59*(4), 3548–3562. https://doi.org/10.1016/j.neuroimage.2011.11.023

Received November 15, 2021
Revision received July 15, 2022
Accepted July 18, 2022 ■