Civilization Through the Suppression of Selfish-Instincts? Helping (Or Hurting) Strangers Through Upregulation of the Other, Not Suppression of the Self

Anthony Romyn[1] and William A. Cunningham[1]

[1]Department of Psychology, University of Toronto

Psy2002: Design of Experiments II

Dr. Liz Page-Gould

April 23rd, 2021

Abstract

Models of human prosociality and antisociality have been rooted in the inhibition of self-interest for centuries. A plethora of contemporary experimental work continues to observe self-inhibiton as central to social behaviour, yet the orthogonalization of self-related and other-related value in experimental contexts remains a difficult challenge limiting the ability to clearly differentiate self-inhibiton from upregulation of the other. The current work isolated self- and other-value and related neural processes with an fMRI implementation of the Dual Gamble task (Allidina, Arbuckle, & Cunningham, 2019; Arbuckle & Cunningham, 2012). Prosocial and antisocial behaviour towards a stranger was critically dependent on the upregulation of the other rather than suppression of the self. Increased other-related striatal reward representations, produced through engagement of the VLPFC, lead to differences in choice. Individual differences in empathy moderated the effect of VLPFC engagement on striatal signals, thus representing the first evidence of how differences in motivation may modulate value-guided choice processes.

Keywords: social decision-making, prosociality, value-guided choice, reward, ventral striatum, multi-level modelling (MLM), structural equation modelling (SEM)

Methods

Participants

51 right-handed participants were recruited for participation from The Ohio State University and surrounding community. All participants had no current or previous history as a person surviving with mental health challenges, no reported neurological history, and normal or corrected-to-normal vision. Two participants were excluded from analysis due to technical issues, leaving a final count of 49 participants used during analysis (26 female; age_{mean}= 22.80, age_{range}= 18-43). The entire session lasted an average of 1.5 hours and participants were paid $20 USD for their time, plus any self-related gamble winnings and any winnings from the gambling decisions of the participant with which they were paired. The fixed compensation and self-related gamble winnings were paid out immediately at the end of the study, while participants were then contacted within the following two weeks about any additional money they were due from the decisions of the participant with which they were paired and arrangements were made for participants to pick up the winnings. No deception was used in the study.

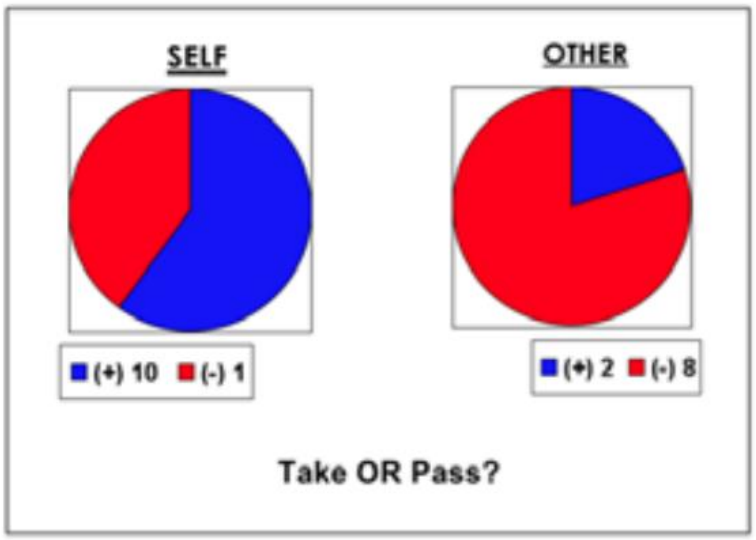
The Dual Gamble Task

In the Dual Gamble task, participants made gamble decisions simultaneously for themselves and an unknown other participant who they were paired with. On each trial, participants saw a screen depicting two gambles, with one gamble for the self and a second gamble for the other participant. The gambles were always presented on the same side (counterbalanced across participants), such that if the gamble for the self was on the left side of the screen in the first trial, it was on the left side of the screen for all further trials.

Each gamble contained three pieces of information: the probability of winning the gamble, the amount that would be won, and the amount that would be lost. For example, a participant may encounter a trial where the gamble for the self has a 60% chance of winning 10 points and a 40% chance of losing one point. They may then see that the gamble for the other has a 20% chance of winning 2 points and an 80% chance of losing 8 points. The details of each trial were communicated through pie chart illustrations and numbers (see Figure 1.), which participants mastered through practice before scanning.

Figure 1

Exemplar trial from the Dual Gamble task



Note. The area within a chart indicated the probability of each outcome while numbers below the pie chart communicated the value of each outcome. In this trial, the self-gamble had a 60% chance of winning 10 or a 40% chance of losing 1, while the other-gamble had a 20% chance of winning 2 and an 80% chance of losing 8. Participants had to make one decision, they could not choose between gambles but had to either take or pass both gambles together.

The gambles were presented on-screen until participants made a button-response, with a timeout of five seconds. A fixation cross then appeared for four seconds before the outcome of each gamble was presented after an inter-trial fixation cross presented randomly for four, six, or eight seconds (ITI_{mean} = six seconds). Outcomes consisted of two numbers, with one indicating the outcome of the gamble for the self and one indicating the outcome of the gamble for the other (see Figure 2.).

Figure 2

Structure of the Dual Gamble task for fMRI scanning

The outcomes were presented onscreen for a total of four seconds, with no button-press response required or accepted from participants. Placement of self and other outcomes on the left or right side of the screen was always the same across trials, and matched the side which the gambles appeared on.

When evaluating gambles, participants had to make a single decision, they could not choose between gambles but had to either take or pass both gambles together. The outcomes of gambles that participants chose to take were added to their total points. The outcomes of gambles

that participants chose to pass were not added to their point totals. Gamble outcomes were presented regardless of whether the gamble was taken or passed.

Making a single decision to accept or reject both gambles may be relatively easy on some trials, such as when both gambles are attractive (accept) or when both gambles are unattractive (reject). However, the relative weighting of self-interest and concern for the other can be explored when the decision is not obvious. When conflicting information was present, participants needed to weigh the relative benefit or cost to self and other to make their decision. Participants had the opportunity to forego opportunities for themselves if the costs were too high for the other, or to take on risk to allow an opportunity for another player.

Orthogonality of Self- and Other-Value

Critically, probabilities and values for self and other were selected randomly and independently on each trial, allowing for the independent estimation of the influence of self- and other-related value on decision. On each trial, gamble win and loss values for each of the self and other were selected randomly and independently from eight possible values: +/-10, +/-7, +/- 4, and +/-1. Win/loss probabilities were similarly selected randomly and independently on each trial for each of the self and other from four possibilities: 80%/20%, 60%/40%, 40%/60%, and 20%/80%.

Practice and Study Sessions

To ensure participants understood the design, they completed a practice session before entering the scanner (see practice session section below). Participants first did 20 practice trials before going into the scanner. The first ten practice trials did not have a time limit, and the second ten practice trials required that participants respond within five seconds, making it equivalent to the study in the scanner. Participants were informed that their outcomes in the

practice trials would not affect their final outcomes in any way. The orientation of the practice trials (i.e. self gamble on the left side or right side of the screen) matched the orientation of the real decision task. Practice trials were completed immediately prior to scanning. During fMRI scanning, participants were then presented with 18 gambles per fMRI run, and completed six fMRI runs, for a total of 108 gambles per participant.

Questionnaires

Personality and demographics questionnaires were randomly given either before scanning (before the practice trials) or after scanning. Questionnaires included the Toronto Empathy Questionnaire (TEQ; Spreng, McKinnon, Mar, & Levine, 2009). On the TEQ, participants respond to 16 items on a five-point scale (ranging from 0 – 4) asking how frequently they feel or behave in the way described. The scale response options are "never", "rarely", "sometimes", "often", and "always". Sample items include "I enjoy making other people feel better", and "I am not really interested in how other people feel" (reverse coded). Higher scores on the TEQ indicate more frequently feeling or behaving in the way described and thus higher levels of empathy.

fMRI Acquisition

Scanning was conducted using a Siemens 3T Trio functional magnetic resonance imaging (TIM) system at The Center for Cognitive and Behavioral Brain Imaging (CCBBI) at The Ohio State University. Functional images were acquired in 34 axial slices parallel to the to the AC-PC line, and nearly isotropic functional images were acquired from inferior to superior using a single-shot gradient echo planar pulse sequence (3.33 mm thick; TE = 25 ms; TR = 2000 ms; in-plane resolution = 3 mm x 3 mm; matrix size = 64 x 64; FOV = 260 mm). The first five volumes of each run were discarded to allow for T1 equilibration effects. Following functional imaging, a

high resolution T1-weighted anatomical image (MPRAGE; 60 sagittal slices; TE = 4.73 ms, TR = 1900 ms; resolution = 0.9×0.9×1.2 mm) was collected for normalization.

fMRI Preprocessing

fMRI Expert Analysis Tool (FEAT) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl) was used to perform brain extraction using BET (Smith, 2002), motion correction estimations with FSL's MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002), spatial smoothing with a Gaussian kernel with full-width at half-maximum (FWHM) of 6mm, and non-linear warping estimations with FMRIB's non-linear counterpart (FNIRT) using brain-based registration (Greve & Fischl, 2009), 12 degrees-of-freedom, and a non-linear warp resolution of 10mm. This FEAT output folder was fed into ICA-AROMA (Pruim, Mennes, Buitelaar, & Beckmann, 2015) for motion correction. The non-aggressive denoised functional outputs from ICA-AROMA for each run were warped to standard space and scaled to a common mean.

Modelling of trial-level fMRI Activity

To extract decision-related BOLD estimates for multi-level modelling of the decision process, the preprocessed functional data was taken to AFNI's 3dDeconvolve function (Cox, 1996). We obtained unique β estimates for the BOLD response for each of the decision and outcome phases of each trial by modelling fMRI time series with individual trial regressors (-stim_times_IM and the least-squares all "LSA" method; see Mumford, Turner, Ashby, & Poldrack, 2012, for a similar analysis strategy). The regressors were convolved with AFNI's "dmBLOCK(1)" function, with decision-phase regressors duration-modulated by decision response time, while all outcome-phase regressors were duration-modulated by a constant length

of two seconds. In both deconvolutions, 3dDeconvolve's polynomial noise removal (POLORT) was set to automatic for removal of trends in the data.

Trials on which no response was made before a five second timeout were modelled as junk trials with a duration of five seconds (mean number of junk trials is 1.8% of trials, or 96 out of 5292. Minimum number of such trials across participants is 1/96 trials, maximum is 12/96 trials).

Multilevel Logistic and Linear Regression Modelling

The β weights for each trial's decision period were then taken to R for the generation of multilevel logistic or linear regression models. All models which predicted participant decisions (take or pass) were performed through the use of multilevel logistic regressions. All models which predicted trial-level fMRI BOLD were performed through the use of multilevel linear regressions. Models which predicted decisions were fit with Laplace approximation maximum likelihood while models which predicted fMRI BOLD were fit with residualized maximum likelihood (REML). Model fitting was performed in R 3.6.3 with the glmer and lmer functions and the nloptwrap optimizer from the lme4 package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2021) while degrees of freedom were estimated with the Satterthwaite method from the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). Assumptions of skewness and kurtosis were examined for all predictor variables. All variables displayed distributions which fell within an acceptable skewness range of -1 to +1 and kurtosis range of +2 to +4. In models with cross-level interaction effects, random slopes were estimated for the level one variable (Aguinis, Gottfredson, & Culpepper, 2013).

In models which contained full-brain voxel-wise analyses, the definition of significant effects was determined with AFNI's 3dClustSim (version 18.0.5), with voxel-wise threshold $p < 0.001$ and cluster threshold $p < 0.05$, the acf option, first-nearest neighbor clustering, and two-sided thresholding. This corresponded to a cluster extent threshold of greater than 117 voxels at $p < 0.001$. 3dFWHMx was used to determine average participant-level noise in the BOLD signal.

Structural Equation Modelling

Structural Equation Modelling (SEM) was performed with the 'sem' function from the lavaan package (version 0.6-7; Rosseel, 2012) in R 3.6.3 (R Core Team, 2020). Interactions between latent variables were protected from bias in estimation through the use of the double-mean-centering technique (Lin, Wen, Marsh, & Lin, 2010). Estimation of model *p*-values was completed through bootstrapping with 1000 iteration.

Results

The Influence of Self and Other Value on Decisions

To examine the degree to which value for the self and value for the other influenced decisions to take or pass on the pair of gambles, self- and other-related value were computed as "expected value" (EV) variables for the self and other gambles as follows:

Expected Value (EV) = (probability of winning) x (amount of possible points won) - (probability of losing) x (amount of possible points lost)

As seen in (1), decisions (take or pass) were then modelled with multilevel logistic regression as a function of the expected value for the Self gamble (SelfEV) and the expected value for the

Other gamble (OtherEV), with random intercepts estimated within participants and runs to capture between- and within-participant variation in general rates of taking or passing.

$$\text{Level 1:} \quad Choice(take/pass)_{ijk} = \beta_{0jk} + \beta_{1ijk}SelfEV_{ijk} + \beta_{2ijk}OtherEV_{ijk} + \varepsilon_{ijk} \quad (1)$$

$$\text{Level 2:} \quad B_{0jk} = \gamma_{00k} + \mu_{0jk}$$

$$\text{Level 3:} \quad \gamma_{00k} = \delta_{000} + V_{0k}$$

Replicating previous work (Allidina, Arbuckle & Cunningham, 2019; Arbuckle & Cunningham, 2012), positive main effects of SelfEV ($\beta = 0.497$, $p < 0.001$) and OtherEV ($\beta = 0.076$, $p < 0.001$) were observed on the decision to take a gamble, indicating that on average, people choose to take gambles that increased point totals for both themselves and others. The effect of SelfEV was larger than that of OtherEV, $\chi^2(1) = 186.79$, $p < 0.001$, indicating that, on average across all participants, self-related value has a greater influence on decision, relative to other-related value.

Variability in the Influence of Other Value, but not Self Value, Produced Differences in Prosocial Behaviour

Having replicated previous work establishing that gamble decisions are influenced by both self- and other-related value, a key first inquiry was to examine whether differing levels of empathy moderated either of these relationships. Individual differences in empathy have previously been related to variability in prosocial decision-making (Zaki & Ochsner, 2012). Thus, modelling of the decision-making process while allowing for moderation by individual differences in empathy can reveal whether individual differences in prosocial behaviour are

produced through moderation of the use of self-related information, other-related information, or both.

As seen in (2), decisions were modelled as a function of SelfEV, OtherEV, self-report empathy, the interaction of SelfEV with empathy, and the interaction of OtherEV with empathy. Random intercepts estimated within participants and runs to capture between- and within-participant variation in general rates of taking or passing. Random slopes for SelfEV and OtherEV were modelled to allow for the correct estimating of the cross-level interaction between the participant-level factor of self-report empathy and the trial-level factors of SelfEV and OtherEV (Aguinis, Gottfredson, & Culpepper, 2013).

Level 1:
$$Choice(take/pass)_{ijk} = \beta_{0jk} + \beta_{1ijk}SelfEV_{ijk} + \beta_{2ijk}OtherEV_{ijk} + B_{3k}Empathy_k + \beta_{4ijk}SelfEVxEmpathy_{ijk} + \beta_{5ijk}OtherEVxEmpathy_{ijk} + \varepsilon_{ijk} \quad (2)$$

Level 2:
$$B_{0jk} = \gamma_{00k} + \mu_{0jk}$$
$$B_{1jk} = \gamma_{10k} + \mu_{1jk}$$
$$B_{2jk} = \gamma_{20k} + \mu_{2jk}$$

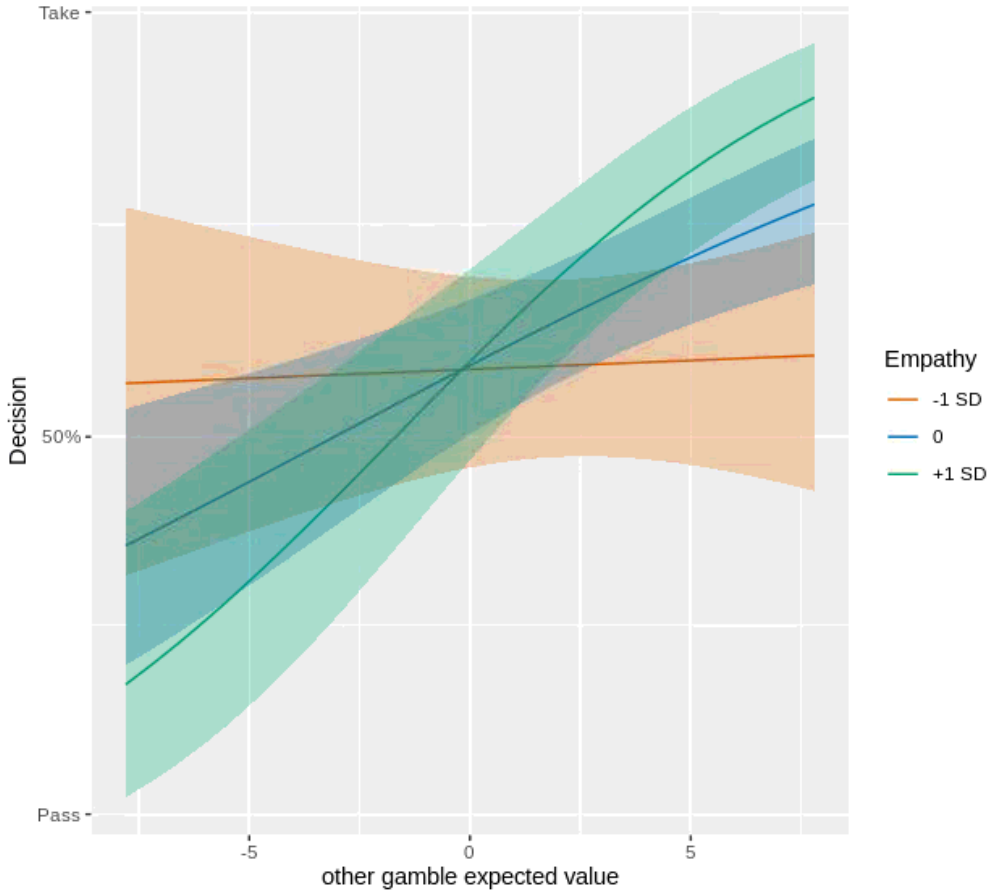
Level 3:
$$\gamma_{00k} = \delta_{000} + V_{0k}$$
$$\gamma_{10k} = \delta_{100} + V_{1k}$$
$$\gamma_{20k} = \delta_{200} + V_{2k}$$

Intriguingly, self-report empathy displayed no moderating effect on the relationship between decision and SelfEV ($p > .05$), but self-report empathy did significantly moderate the relationship between decision and OtherEV ($\beta = 0.09$, $t = 3.32$, $p < .001$; see Figure 3),

Specifically, other-related value had a greater influence on decisions in those with higher self-report empathy, while other-related value had little influence on decisions in those with lower self-report empathy.

Figure 3

Self-report empathy moderated the influence of other value (EV) on decision



Neural Functions Supporting the Use of Other Value

With the results of behavioural modelling suggesting that empathy influences social decision-making through moderating the use of other-related information without influencing the use of self-related information, we next wished to examine whether any neural activity could be

located as potential functional sources for the moderating influence empathy had on the use of other-related information.

As seen in (3), potential neural functions supporting the use of other-related value was examined through the modelling of decision behaviour as a function of self and other EV, self-report empathy, voxel-wise neural activity, and higher-order interactions of these variables. Random slopes for SelfEV, OtherEV, and fMRI were modelled to allow for the correct estimating of the cross-level interaction between the participant-level factor of self-report empathy and the trial-level factors of SelfEV, OtherEV, and fMRI.

Level 1:
$$Choice(take/pass)_{ijk} = \beta_{0jk} + \beta_{1ijk}SelfEV_{ijk} + \beta_{2ijk}OtherEV_{ijk} + B_{3k}SelfValuation_k + \beta_{4k}OtherValuation_k + \beta_{5ijk}SelfEVxEmpathy_{ijk} + \beta_{6ijk}OtherEVxEmpathy_{ijk} + \beta_{7ijk}fMRI_{ijk} + \beta_{8ijk}SelfEVxfMRI_{ijk} + \beta_{9ijk}OtherEVxfMRI_{ijk} + \beta_{10ijk}EmpathyxfMRI_{ijk} + \beta_{11ijk}SelfEVxEmpathyxfMRI_{ijk} + \beta_{12ijk}OtherEVxEmpathyxfMRI_{ijk} + \varepsilon_{ijk} \quad (3)$$

Level 2:
$$B_{0jk} = \gamma_{00k} + \mu_{0jk}$$
$$B_{1jk} = \gamma_{10k} + \mu_{1jk}$$
$$B_{2jk} = \gamma_{20k} + \mu_{2jk}$$
$$B_{7jk} = \gamma_{70k} + \mu_{7jk}$$

Level 3:
$$\gamma_{00k} = \delta_{000} + V_{0k}$$
$$\gamma_{10k} = \delta_{100} + V_{1k}$$
$$\gamma_{20k} = \delta_{200} + V_{2k}$$
$$\gamma_{70k} = \delta_{700} + V_{7k}$$

A three-way interaction of other value by empathy by fMRI activity was observed in the ventrolateral prefrontal cortex (VLPFC) (see Figure 4). Higher activity in this region led to greater influence of other-related value on decision, with self-report empathy moderating how the other-related value shaped. Specifically, when activity was high in those with high empathy, decision-making became more altruistic. In contrast, when activity was high in those with low empathy, decision-making became less altruistic and more competitive or counter-empathic to the rewards of the other (e.g. VLPFC: $\beta = 0.15$, $t = 5.28$, $p < .001$; see Figure 5). Activity in this region did not influence the use of self-related value ($p > .05$).

Figure 4

Ventrolateral prefrontal (VLPFC) showing three-way interactions of other value by empathy by fMRI activity predicting decision

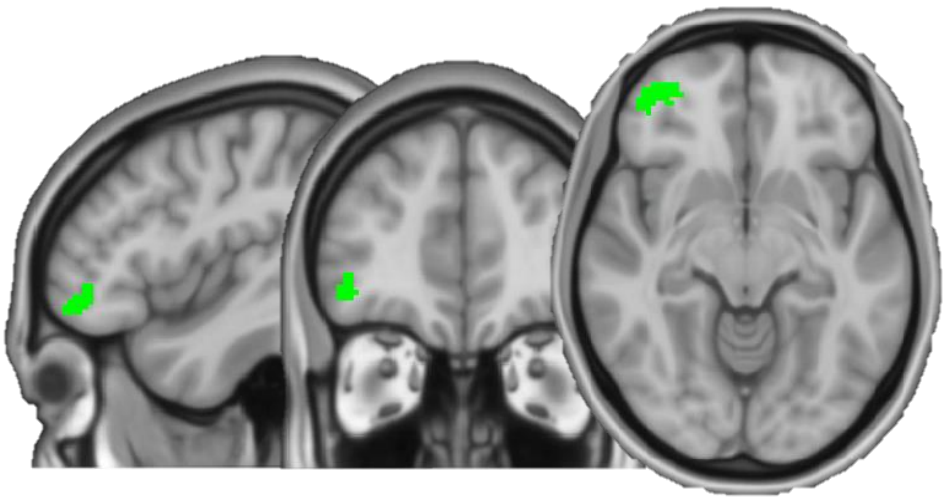
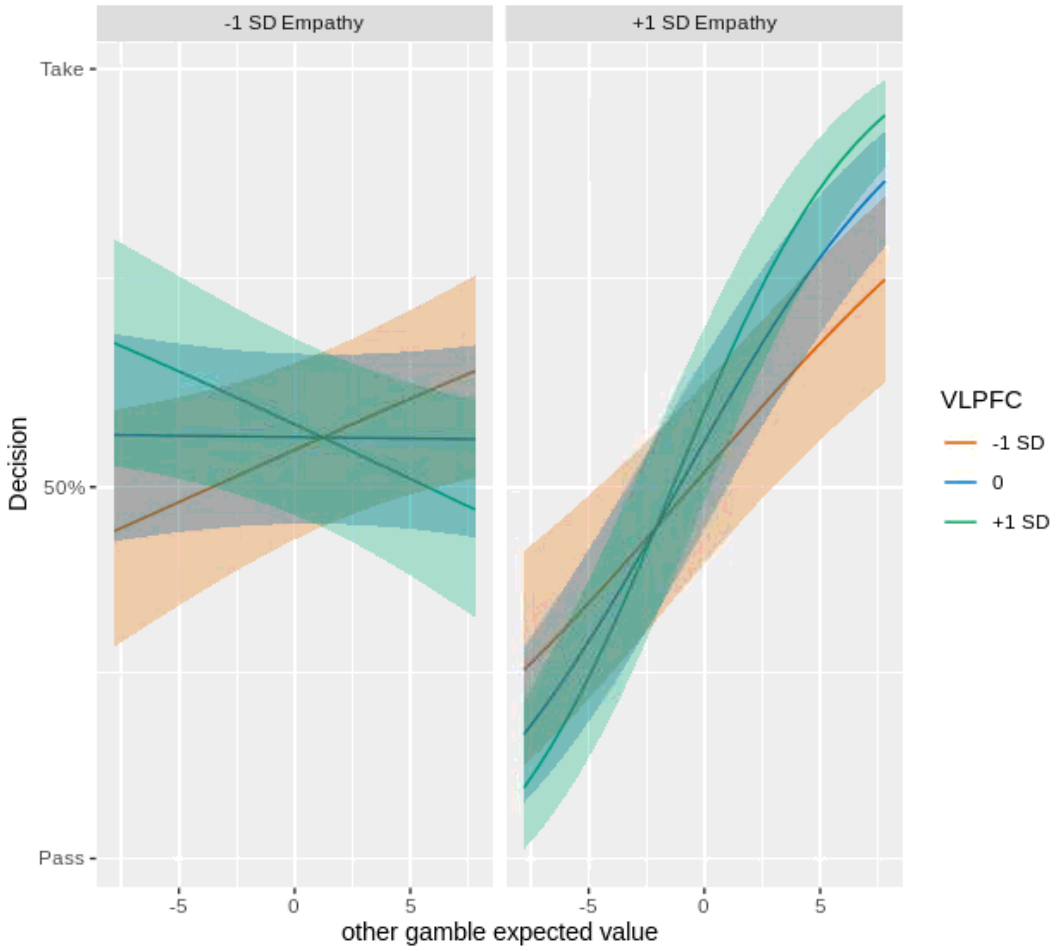


Figure 5

VLPFC activity led to greater influence of other-related value on decision, with self-report empathy moderating how the other-related value shaped decision



VLPFC Shapes Other-Value Representations within the Ventral Striatum

With our prior findings suggesting that the our VLPFC clusters provide a neural basis for the moderating role of empathy on the use of other-related information during decision, a key next question is how this VLPFC functioning is translated into behavioural differences. One possibility is that VLPFC activity may be serving to influence the degree to which other-related information is being represented within the ventral striatum (vStr) during choice. The ventral

striatum is believed to be critical to action selection (e.g. O'Doherty et al., 2004), with striatal processes believed to produce action selection based on value expectations (O'Reilly, Munakata, Frank, & Hazy, 2012).

To examine whether our VLPFC functioning is influencing other-related value representations in the ventral striatum, we first wished to isolate the ventral striatum. Given the previously mentioned role of the vStr in choice, vStr activity is frequently seen predicting choice behaviour (e.g. Kable & Glimcher, 2007). Leveraging this relationship between vStr and choice for the isolation of the vStr, a simple multilevel logistic regression was used predicting choice (take/pass) from full-brain voxel-wise fMRI, with random intercepts modelled at the level of participant and runs within participant, as seen in (4). Clusters in both the left and right vStr were observed.

Level 1: $$Choice(take/pass)_{ijk} = \beta_{0jk} + \beta_{1ijk} fMRI_{ijk} + \varepsilon_{ijk} \quad (4)$$

Level 2: $$B_{0jk} = \gamma_{00k} + \mu_{0jk}$$

Level 3: $$\gamma_{00k} = \delta_{000} + V_{0k}$$

The vStr clusters isolated from (4) were then used as masks for a subsequent analysis where VStr activity was modelled a function of self-related value, other-related value, self-report empathy, VLPFC cluster activity, and higher-order interactions of these variables, as seen in (5). Random slopes for SelfEV, OtherEV, and fMRI were modelled to allow for the correct estimating of the cross-level interaction between the participant-level factor of self-report empathy and the trial-level factors of SelfEV, OtherEV, and fMRI.

Level 1: $$\textit{Ventral Striatum Voxel Activity}_{ijk} = \beta_{0jk} + \beta_{1ijk}SelfEV_{ijk} + \beta_{2ijk}OtherEV_{ijk} \quad (5)$$

$$+ B_{3k}SelfValuation_k + \beta_{4k}OtherValuation_k + \beta_{5ijk}SelfEVxEmpathy_{ijk}$$

$$+ \beta_{6ijk}OtherEVxEmpathy_{ijk} + \beta_{7ijk}VMPFCActivity_{ijk} + \beta_{8ijk}SelfEVxVMPFCActivity_{ijk}$$

$$+ \beta_{9ijk}OtherEVxVMPFCActivity_{ijk} + \beta_{10ijk}EmpathyxVMPFCActivity_{ijk}$$

$$+ \beta_{11ijk}SelfEVxEmpathyxVMPFCActivity_{ijk}$$

$$+ \beta_{12ijk}OtherEVxEmpathyxVMPFCActivity_{ijk} + \varepsilon_{ijk}$$

Level 2: $$B_{0jk} = \gamma_{00k} + \mu_{0jk}$$

$$B_{1jk} = \gamma_{10k} + \mu_{1jk}$$

$$B_{2jk} = \gamma_{20k} + \mu_{2jk}$$

$$B_{7jk} = \gamma_{70k} + \mu_{7jk}$$

Level 3: $$\gamma_{00k} = \delta_{000} + V_{0k}$$

$$\gamma_{10k} = \delta_{100} + V_{1k}$$

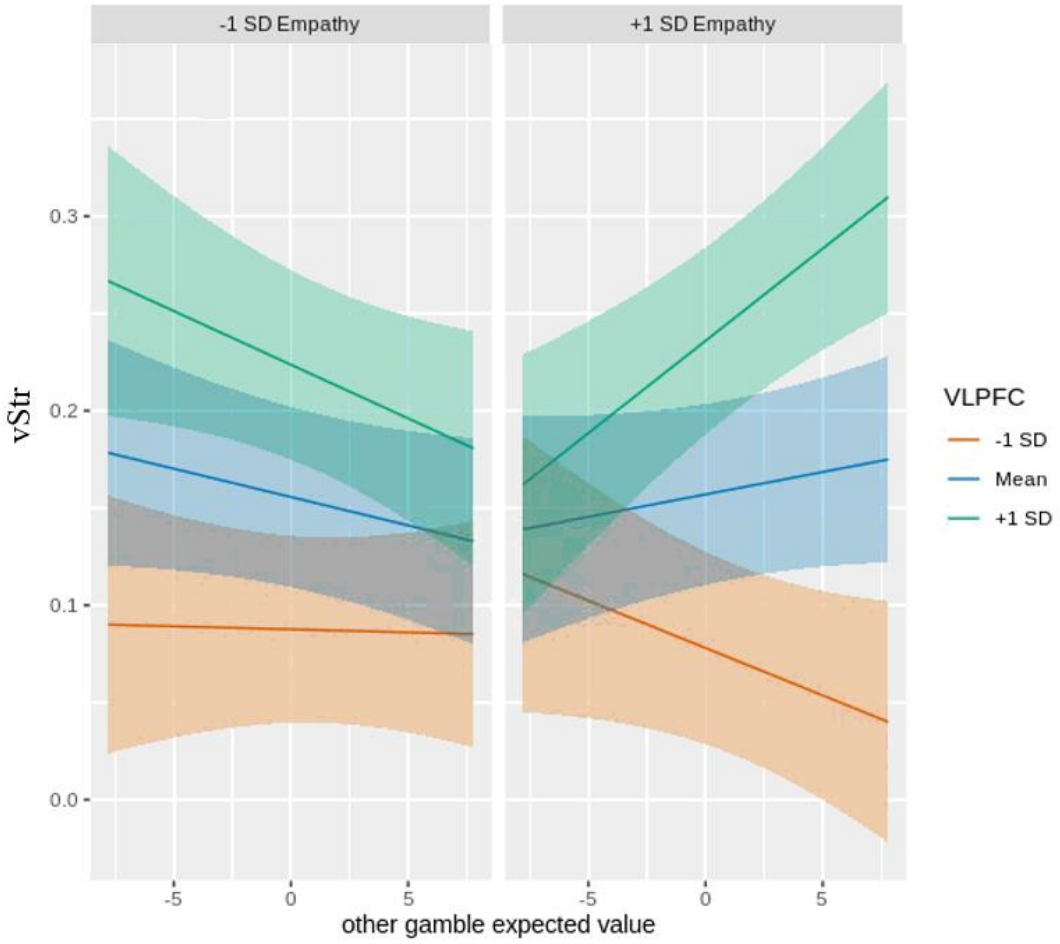
$$\gamma_{20k} = \delta_{200} + V_{2k}$$

$$\gamma_{70k} = \delta_{700} + V_{7k}$$

As seen in Figure 6, the three-way interaction of other value by empathy by VLPFC activity predicted a cluster within the vStr mask (FDR-corrected cluster size = 44 voxels). Higher levels of VLPFC activity increasingly strengthened the relationship between ventral striatum activity and other-related value, in an empathy-dependent manner. Increasing VLPFC activity in those with higher empathy resulted in an increasingly strong positive relationship between ventral striatum activity and other-related value. In contrast, increasing VLPFC activity in those with low empathy resulted in an increasingly strong negative relationship between ventral stiratum activity and other-related value ($\beta = 1.61$, $t = 3.65$, $p < .001$).

Figure 6

Ventral Striatum activity (VStr) predicted by the three-way interaction of other-related value, self-report empathy, and trial-level VLPFC activity



VLPFC-based changes in vStr Leads to Motivation-Consistent Behavioural Change

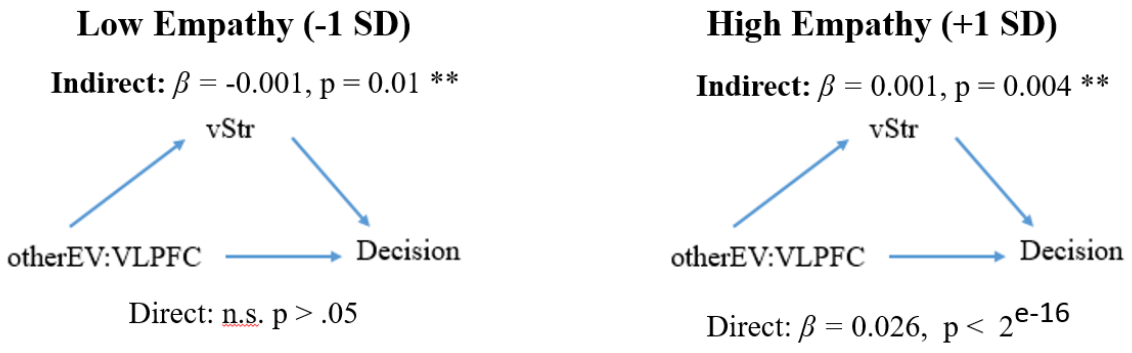
Having observed that vStr representations of other-related value were moderated by VLPFC activity, the key final question is whether these fluctuations in vStr representations are leading to differences in choice. The identified VLPFC mechanism can be the mechanism through which variability in empathy leads to changes in the use of other-related information, but

only if the VLPFC-induced changes in vStr representations lead to differences in choice. To test this question, a moderated-mediation path analysis was performed with Structural Equation Modelling (SEM) and the use of double-mean-centering (Lin et al., 2010).

Examination of model fit metrics revealed that the model fit the data well, SRMR = .013, RMSEA = .04, 90% CI [.036, .049], CFI = .941 (e.g. see Hu & Bentler, 1999; MacCallum, Browne, & Sugawara, 1996). As seen in Figure 7, moderated-mediation was observed, with decision behaviour predicted by the interaction of other-related value (OtherEV) with the VLPFC through the ventral striatum (vStr). This mediation relationship was moderated by levels of empathy, with the mediation effect through the vStr displaying opposing effects on decision for those with lower levels of empathy (β = -0.001, Z = -2.57, p = 0.01) relative to those with higher levels of empathy (β = 0.001, Z = 2.88, p = 0.004).

Figure 7

Moderated-mediation revealing that VLPFC-based changes in vStr Other-Reward Representations Lead to Changes in Behaviour



Discussion

Models of human prosociality and antisociality have been rooted in the inhibition of self-interest for centuries (for contemporary iterations, see: Capraro & Cococcioni, 2016; DeWall, Baumeister, Gailliot, & Maner, 2008; Haidt, 2008; Hofmann, Meindl, Mooijman, & Graham, 2018; Rand, Greene, & Nowak, 2012; for reviews, see: Mansbridge, 1990; Wallach & Wallach, 1983). Yet the orthogonalization of self-related and other-related value in experimental contexts has remained a difficult challenge limiting the ability to clearly differentiate self-inhibiton from upregulation of the other. Isolated self- and other-value and related neural processes with an fMRI implementation of the Dual Gamble task, the current work suggests that prosocial or antisocial behaviour towards a stranger was critically dependent on the upregulation of other-related value representations rather than any form of suppression of the self.

Variability in Prosocial Motivation within Value-Guided Decision-Making

Decisions for the self and another person were shaped by the degree to which value related to the other was represented within the ventral striatum during choice. The ventral striatum is believed to be critical to action selection (e.g. O'Doherty et al., 2004), with striatal processes believed to produce action selection based on value expectations (O'Reilly, Munakata, Frank, & Hazy, 2012). The nature of the other-related value representations within the striatum were dependent on the degree to which participants were motivated to behave prosocially towards the stranger, as measured through self-report empathy. Thus, variability in motivation produced variability in behaviour through change to specific other-related reward representations in the striatum. Previous work has found evidence of individual differences in the degree to which other-value was pursued as a reward (Hutcherson et al., 2015), yet the mechanism through which these individual differences shifted value-guided choice have been unclear. The

modulation of striatal reward signals illustrates one pathway through which variability in motivation can manifest within value-guided decision processes

The Role of the VLPFC

The representation of other-related value within the ventral striatum was supported by the engagement of the ventrolateral prefrontal cortex (VLPFC), such that increased engagement of the VLPFC region related to increased representations of other-related value within the ventral striatum. VLPFC producing up-regulation of other-related value representations within the striatum during choice may be due to functions previously associated with the VLPFC such as the active maintenance of contextual information like one's current goals or information pertaining to those goals (e.g. Reynolds, O'Reilly, Cohen, & Braver, 2012), or increasing the attention allocated to other-related stimuli on screen (e.g. Hunt et al., 2018).

Motivation and Social Identity

The moderating effects of individual differences in empathy on striatal representations of the other highlights the potential critical role of social identity in shaping value representations, leading to changes in social behaviour (Tajfel & Turner, 1979; Turner, Oakes, Haslam, & McGarty, 1994). In non-clinical populations, such as the sample used in the current work, prosocial and antisocial behaviour can vary by how participants contextualize themselves within the social context (e.g. Arbuckle & Cunningham, 2012). Thus, it may be that variability in prosocial and antisocial behaviour in the current task stems from how participants contextualized themselves within the experiment.

However, the modulation observed with our self-report empathy variable may not necessarily represent variability in participant social identity. For instance, it may be serving as a proxy for attention to the task, such that those with lowest attention to the task may have

misinterpreted the instructions, thereby acting antisocially by mistake. Such an account could also explain why other-related striatal representations were inverted in these participants. Thus, our inferential ability is limited with the currently used individual difference variables.

To further test the motivation and social identity hypothesis, experimental manipulations of motivation would be of particular value, such as through the use of a minimal groups paradigm (e.g. Van Bavel & Cunningham, 2012). With such designs, manipulation of target grouping as either ingroup or outgroup has successfully produced shifts in motivation to behave prosocially. Thus, minimal groups design would allow for the greater isolation of social-identity processes from alternative individual differences which limit the inferential abilities of the current work.

Conclusion

Models of human prosociality and antisociality have been rooted in the inhibition of self-interest for centuries. Yet the current work suggests that prosocial or antisocial behaviour towards a stranger was critically dependent on the upregulation of the other rather than suppression of the self. Upregulation of the other took the form of increased other-related striatal reward representations, produced through engagement of the VLPFC, leading to differences in choice. The effect of the VLPFC on striatal representations of the other was moderated by individual differences in motivation to behave prosocially or antisocially, thus representing the first evidence of how differences in motivation may modulate value-guided choice processes.

References

Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management, 39*(6), 1490-1528.

Allidina, S., Arbuckle, N. L., & Cunningham, W. A. (2019). Considerations of Mutual Exchange in Prosocial Decision-Making. *Frontiers in Psychology, 10,* 1216.

Arbuckle, N. L., & Cunningham, W. A. (2012). Understanding everyday psychopathy: Shared group identity leads to increased concern for others among undergraduates higher in psychopathy. *Social Cognition, 30*(5), 564-583.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29*(3), 162-173.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage, 48*(1), 63-72.

Haidt, J. (2008). Morality. *Perspectives on Psychological Science, 3*, 65–72

Hofmann, W., Meindl, P., Mooijman, M., & Graham, J. (2018). Morality and self-control: How they are intertwined and where they differ. *Current Directions in Psychological Science, 27*(4), 286-291.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55.

Hunt, L. T., Malalasekera, W. N., de Berker, A. O., Miranda, B., Farmer, S. F., Behrens, T. E., & Kennerley, S. W. (2018). Triple dissociation of attention and decision computations across prefrontal cortex. *Nature neuroscience, 21*(10), 1471-1481.

Lin, G. C., Wen, Z., Marsh, H. W., & Lin, H. S. (2010). Structural equation models of latent interactions: Clarification of orthogonalizing and double-mean-centering strategies. *Structural Equation Modeling, 17*(3), 374-391.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304*(5669), 452-454.

O'Reilly, R. C., Munakata, Y., Frank, M. J., and Hazy, T. E. (2012). *Computational Cognitive Neuroscience*, 1st Edn. Wiki Book. Available online at: http://ccnbook.colorado.edu

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature neuroscience, 10*(12), 1625-1633.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software, 82*(13), 1-26.

Mansbridge, J. J. (Ed.). (1990). *Beyond self-interest.* University of Chicago Press.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods, 1*(2), 130.

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage, 59*(3), 2636-2643.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2021). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-152, https://CRAN.R-project.org/package=nlme.

Pruim, R. H., Mennes, M., Buitelaar, J. K., & Beckmann, C. F. (2015). Evaluation of

ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI. *Neuroimage, 112*, 278-287.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rand, D. G., Greene, J. D. & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature 489*, 427–430.

Reynolds, J. R., O'Reilly, R. C., Cohen, J. D., & Braver, T. S. (2012). The function and organization of lateral prefrontal cortex: a test of competing hypotheses. *PloS one, 7*(2), e30284.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1 - 36. doi: http://dx.doi.org/10.18637/jss.v048.i02

Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping, 17*(3), 143-155.

Spreng*, R. N., McKinnon*, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment, 91*(1), 62-71.

Tajfel, H., & Turner, J. C. (1979). *An integrative theory of intergroup conflict*. In W. G. Austin & S. Worchel (Eds.), The social psychology of intergroup relations (pp. 94-109). Monterey, CA: Brooks-Cole.

Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin, 20,* 454-463

Van Bavel, J. J., & Cunningham, W. A. (2012). A social identity approach to person memory:

Group membership, collective identification, and social role shape attention and memory. *Personality and Social Psychology Bulletin, 38*(12), 1566-1578.

Wallach, M. A. and Wallach, L. (1983) *Psychology's Sanction for Selfishness: The Error of Egoism in Theory and Therapy.* San Francisco, CA: W. H. Freeman and Company.

Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nature neuroscience, 15*(5), 675-680.