Neural representations of vicarious rewards are linked to interoception and prosocial behaviour

Contreras-Huerta, Luis Sebastian; Coll, Michel-Pierre; Bird, Geoffrey; Yu, Hongbo; Prosser, Annayah; Lockwood, Patricia L.; Murphy, Jennifer; Crockett, M.J.; Apps, Matthew A.J.

DOI:
10.1016/j.neuroimage.2023.119881

NeuroImage 269 (2023) 119881

Contents lists available at ScienceDirect

NeuroImage

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

Neural representations of vicarious rewards are linked to interoception and prosocial behaviour

Luis Sebastian Contreras-Huerta[a,b,c,d,l,*], Michel-Pierre Coll[e], Geoffrey Bird[a,m], Hongbo Yu[f], Annayah Prosser[g], Patricia L. Lockwood[a,b,c,d,h], Jennifer Murphy[i], M.J. Crockett[a,j,k,1], Matthew A.J. Apps[a,b,c,d,h,1,*]

[a] Department of Experimental Psychology, University of Oxford, Oxford OX1 3PH, UK
[b] Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford OX3 9DU, UK
[c] Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham B15 2TT, UK
[d] Institute for Mental Health, School of Psychology, University of Birmingham, Birmingham B15 2TT, UK
[e] School of Psychology and CIRRIS research center, Laval University, Quebec City QC G1V 0A6, Canada
[f] Department of Psychological and Brain Sciences, University of California Santa Barbara, Santa Barbara, CA 93106, USA
[g] Department of Psychology, University of Bath, BA2 7AY, United Kingdom
[h] Christ Church, University of Oxford, Oxford OX1 1DP, UK
[i] Department of Psychology, Royal Holloway, University of London, London TW20 0EY, UK
[j] Department of Psychology, Yale University, New Haven, CT 06511, USA
[k] Department of Psychology and University Center for Human Values, Princeton University, Princeton, USA
[l] Center for Social and Cognitive Neuroscience (CSCN), School of Psychology, Universidad Adolfo Ibáñez, Viña del Mar, Chile
[m] School of Psychology, University of Birmingham, Birmingham B15 2TT, UK

ARTICLE INFO

Keywords:
Anterior cingulate cortex
Anterior insula
Vicarious reward
Interoception
Prosocial behaviour

ABSTRACT

Every day we constantly observe other people receiving rewards. Theoretical accounts posit that vicarious reward processing might be linked to people's sensitivity to internal body states (interoception) and facilitates a tendency to act prosocially. However, the neural processes underlying the links between vicarious reward processing, interoception, and prosocial behaviour are poorly understood. Previous research has linked vicarious reward processing to the anterior cingulate gyrus (ACCg) and the anterior insula (AI). Can we predict someone's propensity to be prosocial or to be aware of interoceptive signals from variability in how the ACCg and AI process rewards? Here, participants monitored rewards being delivered to themselves or a stranger during functional magnetic resonance imaging. Later, they performed a task measuring their willingness to exert effort to obtain rewards for others, and a task measuring their propensity to be aware and use interoceptive respiratory signals. Using multivariate similarity analysis, we show that people's willingness to be prosocial is predicted by greater similarity between self and other representations in the ACCg. Moreover, greater dissimilarity in self-other representations in the AI is linked to interoceptive propensity. These findings highlight that vicarious reward is linked to bodily signals in AI, and foster prosocial tendencies through the ACCg.

1. Introduction

From seeing strangers enjoy delicious-looking meals in restaurants, to observing likes on social media posts, witnessing other people receive rewards is a fundamental feature of our social lives. Theories suggest that vicarious reward processing, which occurs when passively observing others getting reward outcomes, is a key component of social behaviour (Ruff and Fehr, 2014; Lockwood, 2016). More strongly representing others' rewards may allow people to share others' positive experiences and lead to stronger visceral responses (Lockwood, 2016; Adolfi et al., 2017; Critchley and Garfinkel, 2017; Grynberg and Pollatos, 2015; Lischke et al., 2020; Shah et al., 2017), as well as it may lead us to choose to exert effort into prosocial acts aimed at obtaining positive outcomes for them (Lockwood et al., 2017; Rilling et al., 2002; Contreras-Huerta et al., 2020a, 2020b; De Waal, 2008). As such, representing other people's outcomes may be linked to the propensity to

* Corresponding authors at: Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham B15 2TT, UK.
E-mail addresses: l.contrerashuerta@bham.ac.uk (L.S. Contreras-Huerta), m.a.j.apps@bham.ac.uk (M.A.J. Apps).
[1] Shared senior co-authorship.

https://doi.org/10.1016/j.neuroimage.2023.119881.
Received 31 August 2022; Received in revised form 12 December 2022; Accepted 14 January 2023
Available online 23 January 2023.

be aware of internal bodily signals that influence our social behaviour. However, such accounts have largely not been empirically tested. Although some previous studies have linked vicarious reward processing to prosocial behaviour (Fukuda et al., 2019; Greening et al., 2014; Harbaugh et al., 2007), and variability in interoception to the processing of social information (Grynberg and Pollatos, 2015; Lischke et al., 2020; Shah et al., 2017; Piech et al., 2017; Terasawa et al., 2013), it is unclear whether it is the vicarious representation of others' rewards that is linked to interoception and prosocial behaviour, or some other aspect of the tasks that are used to probe them (e.g. the desire to help). As a result, the neural mechanisms that relate either interoception or prosocial behaviour to vicarious reward signals remain poorly understood.

Although a number of areas have been linked to vicariously processing other's states, two regions of the brain have been consistently linked to the processing of other's rewards - the gyral portion of the anterior cingulate cortex (ACCg) and the anterior insula (AI) (Lockwood, 2016; Apps and Ramnani, 2014; Morelli et al., 2015; Lockwood et al., 2015; Apps et al., 2013a). Comprehensive reviews have suggested these regions may have at least partially specialised roles in processing others' rewards, contrasting them with other neighbouring sub-regions of the ACC or insula (e.g. ACC sulcus, posterior insula) more focused on the processing of self events (Lockwood, 2016; Apps and Ramnani, 2014; Apps et al., 2013a; Fan et al., 2011; Lindquist et al., 2016; Contreras-Huerta et al., 2013; Hill et al., 2016; Monfardini, 2013; Behrens et al., 2008; Apps et al., 2016; Lamm et al., 2011; Hadland et al., 2003; Rudebeck et al., 2006). The AI and ACCg are engaged when we see cues indicating other people will receive a reward, respond differently when people receive a reward themselves, and do not respond to foregone rewards delivered neither to self or other (Lockwood, 2016; Apps et al., 2013a). Strikingly, separate lines of research also implicate these same regions in multimodal interoceptive processes and to prosocial behaviour (Harbaugh et al., 2007; Apps et al., 2016; Critchley et al., 2004; Critchley and Nagai, 2012; Craig, 2009; Pollatos et al., 2016; Harrison et al., 2021). The vast majority of this research uses cardiac-based interoceptive tasks, showing that variability in signalling in these regions is linked to variability in people's cardiac sensibility or sensitivity. However, other studies have also shown their involvement across other interoceptive domains, including respiration (Pollatos et al., 2016; Harrison et al., 2021). This anatomical overlap might suggest that reward processing, interoception and motivation may be intimately linked (Naqvi and Bechara, 2009; Paulus, 2007; Paulus and Stewart, 2014). However, it is unclear whether vicarious reward processing in the ACCg and AI are linked separately to these different functions.

Indeed, research has suggested that despite both the ACCg and AI being engaged when seeing others' rewards, the processing in each region may serve different functional roles. Broadly speaking, the AI is engaged in processing the links between information about rewards and bodily states, while ACC may be more strongly linked to driving the motivation to obtain rewarding outcomes (Paulus, 2007; Paulus and Stewart, 2014; Craig, 2002; Dillon et al., 2008; Holroyd and Yeung, 2012; Rolls, 2016; Williams et al., 2004; Mega and Cohenour, 1997; Vogt, 2009; Le Heron et al., 2018). Indeed the AI is part of the association cortex, receiving input from multiple sensory regions that in turn receive input from a range of bodily systems (Mesulam and Mufson, 1982). Thus, AI is well placed to integrate interoceptive signals from a variety of different modalities (e.g. cardiac or respiratory (Nord and Garfinkel, 2022)), with information about one's affective state in different social contexts. All of this might be crucial to disentangle vicarious from self rewards (Craig, 2009; Craig, 2002; Critchley, 2005). On the other hand, ACCg is putatively involved in prosocial motivation and vicarious processes, with some studies reporting specific responses for social stimuli, while others showing general activity for self and other (Lockwood, 2016; Apps and Ramnani, 2014; Apps et al., 2013a; Hill et al., 2016; Monfardini, 2013; Behrens et al., 2008; Apps et al., 2016; Lamm et al., 2011; Lockwood et al., 2020, 2022). However, as most tasks examining the neural processes underlying prosocial behaviour involve self and other rewards, and also may evoke interoceptive processes, it has been difficult to disentangle whether vicarious processing in the ACCg and AI is associated with distinct functions, and how specific or general these roles are for self and vicarious processes. We propose that one solution to this problem is to examine whether neural responses to others' rewards in the ACCg and AI when simply passively observing them are linked to either individual differences in people's willingness to be prosocial, or their propensity to be aware of interoceptive signals, in separate tasks.

How can we interpret neural responses to others' rewards? While a number of studies have examined vicarious reward processing with functional magnetic resonance imaging (fMRI), there are contrasting viewpoints on how vicarious reward BOLD signals should be interpreted (Ruff and Fehr, 2014; Lockwood, 2016; Apps et al., 2016; Lockwood et al., 2020). On the one hand, the overlap between clusters that respond to self and other rewards in the ACC and AI is often interpreted as being indicative of a 'common currency' that represents information about both ourselves and other people in the same manner (Ruff and Fehr, 2014; Saxe and Haushofer, 2008). Thus, it is assumed that this neural overlap could be indicative of a similar positive affective response at another receiving rewards to rewards being received by oneself. In contrast, other studies have shown that greater specialisation – more distinct processing between self and other rewards – is linked to higher levels of empathy and higher learning speed for prosocial actions (Lockwood et al., 2015, 2016). Such greater specialisation, rather than a common currency, might therefore underlie stronger prosocial tendencies (Lockwood et al., 2015, 2020, 2016). However, only a few studies have examined variability in vicarious reward processing between people, with the majority of them using univariate fMRI analysis, where the strength of activity in a certain brain location is correlated between self and other conditions. Even though these studies have provided valuable evidence to support either perspective on this dispute, there is an inherent limitation of this approach - similar strength in univariate BOLD signal does not necessarily imply similar multivariate patterns of responses for self and other rewards. Therefore, even in the presence of neural overlap between self and other rewards in some individuals, there could be a different pattern of response for each condition within a brain area, suggesting distinctive processing. Crucially, analysis techniques that can formally test the similarity between self and other reward representations, such as multivariate pattern analysis, have rarely been employed in previous studies, and typically without paying attention to individual differences.

If the degree in which people are aware of their internal signals is involved in vicarious reward processing, then it is also expected that it might influence motivated behaviour to obtain rewards for others. Indeed, previous research has suggested that people's propensity to rely on interoceptive signals is linked to their levels of altruism (Piech et al., 2017), although this evidence has not been consistent (Lenggenhager et al., 2013). One explanation for these contradictory results is that the tasks used to measure prosocial or interoceptive tendencies may confound self and other processing together (Contreras-Huerta et al., 2020b). For instance, studies have used economic games for measuring prosociality in which the rewards delivered to the other person directly impact on the magnitude of rewards obtained by oneself. Although interesting for quantifying variability in one's desire to benefit others, they cannot distinguish between two different motives - sensitivity to one's own rewards, or an increasing desire to give rewards to other people (Contreras-Huerta et al., 2020b; Lockwood et al., 2020). Therefore, using tasks that can address this potential confound, and separate self from other processing, might shed light on the link between interoception and prosocial behaviour, as well as the links to vicarious reward representations.

Here, we use a multivariate approach to test whether similarity in neural responses to self and other reward in the AI and ACCg is linked to people's propensity to be aware and use respiratory interoceptive signals and their willingness to exert effort into prosocial behaviours. We

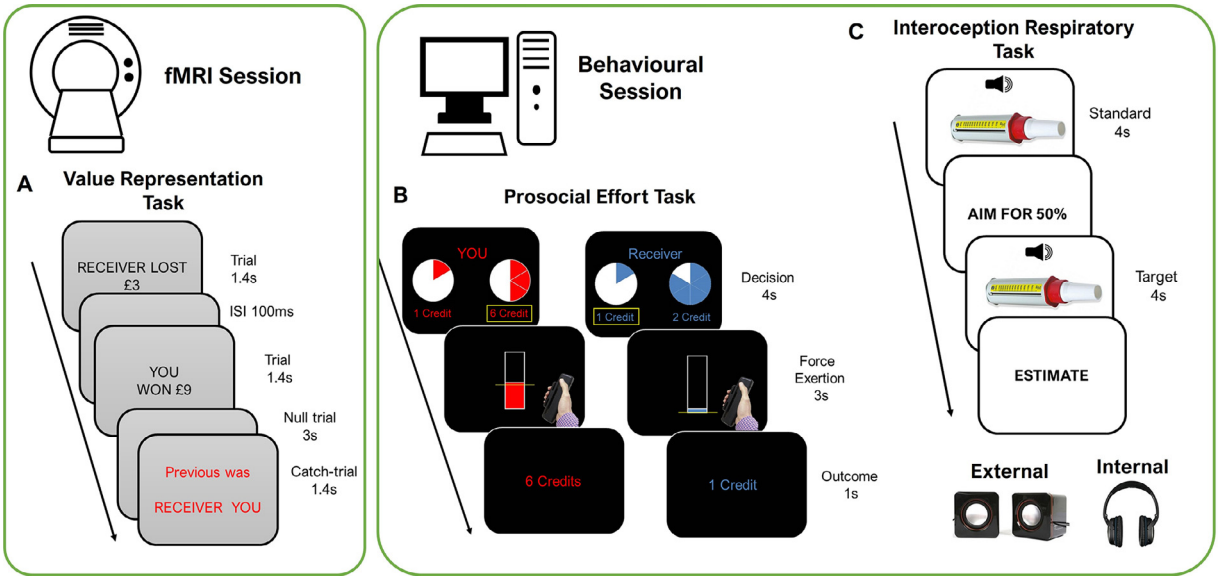

Fig. 1. fMRI and Behavioural tasks. Participants completed three tasks across two sessions separated by at least one week A. *fMRI session: The value representation task.* This task measured neural similarity between representation of rewards and losses for the self and others**.** Participants passively witnessed financial gains (6 levels) and losses (6 levels) for themselves and for another unknown person, the Receiver**.** Catch trials asked either who was the recipient of the last outcome, or the magnitude of the reward on the previous trial, to encourage attention to the task. Similarity was calculated as the Pearson correlation between the spatial representation of self- and other-related neural activity. *B Prosocial Effort task.* Participants made choices about whether to exert different amounts of effort (30–70% of their own maximum grip strength) on a handheld dynamometer for variable amounts of reward (2–10 credits). Participants worked either to benefit themselves or an anonymous other. *C. Interoceptive Respiratory Task.* On each trial participants blew into a peak-flow metre twice, the first blow setting a standard for that trial. In the second blow, participants were required to achieve a percentage of the first, with participants estimating their actual percentage performance at the end of the trial. This was conducted in an internal condition with white noise played through headphones to prevent the use of external auditory cues, and an external condition where external auditory cues were available. The difference in estimation accuracy between internal and external conditions was taken as a measure of how much participants rely on internal vs external signals. Note that the sequence of screens is an illustration of the test, as participants were blindfolded.

define reward as either winning or losing a financial reward, as studies have shown neural responses in the ACC and AI for winning and losses for self (Sallet et al., 2007; Liu et al., 2011) and others (Apps et al., 2013b). Participants performed three tasks (Fig. 1), the first while undergoing fMRI and the second two in a separate behavioural session. Participants first underwent fMRI during a value representation task, in which they observed cues indicating that rewards were being accrued – or being taken away – from oneself or from an anonymous other person. In addition, they performed two behavioural tasks without fMRI. First, an interoceptive respiratory task, which measured people's propensity to be aware and use internal rather than external signals to judge their respiratory output (Murphy et al., 2018). Second, a prosocial effort task (Lockwood et al., 2017), which measured willingness to exert physical effort to obtain rewards for oneself or for an anonymous stranger (Lockwood et al., 2017; Contreras-Huerta et al., 2020a; Lockwood et al., 2022, 2021). These two tasks enable measurement of individual differences in reliance on interoceptive respiratory cues and prosocial motivation in ways that could control for some of the confounds described above. Thus, while the interoceptive respiratory task measures specifically the propensity of people to use internal vs external signals, and not merely interoceptive accuracy, the prosocial effort task allows us to measure sensitivity to one's own and other's rewards, as well as separately sensitivity to exerting effort for oneself and others. Finally, to examine whether the distinctiveness of self and other reward representations in ACCg and AI is linked to prosocial motivation and interoception respectively, offline task behaviour was correlated with pattern similarity in each region from the value representation task (Kriegeskorte et al., 2008).

2. Materials and methods

2.1. Participants

The study was approved by the University of Oxford ethics committee (R50262/RE001) and all procedures were in line with the declaration of Helsinki. 80 healthy volunteers aged 18–38 years, recruited from the University and the city of Oxford, participated in the study, gave informed consent and were compensated for their time (£15/h for the fMRI session, £8/h for behavioural session, plus extra potential bonuses earned through the reward-related tasks). Sample size was based on previous work using the behavioural tasks included in this study (Lockwood et al., 2017; Lockwood et al., 2022; Murphy et al., 2018; Lockwood et al., 2021). Participants were excluded if they had a history of neurological or neuropsychiatric disorders, psychoactive medication or drug use. Pregnant participants were also excluded from the study, as well as those who had participated in studies involving social interaction and who had studied psychology, addressing concerns that prior experience could influence psychological and brain processes.

61 participants (age $M = 22.6$, SD = 3.8, 32 females) were included in the behavioural analysis – four participants failed to show-up for the behavioural session, two participants expressed doubts about the social manipulation and the existence of the receiver, nine did not complete either the interoception respiratory task or the prosocial effort task, and four were excluded from analysis because of poor quality data in either of the two tasks (see below). From these 61 participants, 56 were included for the fMRI analysis (age $M = 22.3$, SD = 3.3, 30 females). One participant's neuroimaging data was not registered due to technical is-

sues, one participant showed excessive head motion, and three participants were excluded from analysis due to poor quality data in the value representation task (see below).

2.2. General procedure

Participants took part in an fMRI and a separate behavioural session. The value representation task was performed while in the fMRI scanner, whereas the interoception respiratory task and the prosocial effort task were tested as part of a behavioural testing session, performed at least one week after the fMRI session.

2.2.1. fMRI session: value representation task

Before entering the scanner, participants were mock-randomly assigned the role of the Decider, which they kept during the whole study, while another unknown person, a confederate, was assigned the role of the Receiver, following a well-established social manipulation protocol (Lockwood et al., 2017; Crockett et al., 2014) (see **Supplementary Information** for details). Participants were instructed that, on each trial, a computer would randomly select a value to add or to subtract to their total earnings (self trials) or to the earnings of the receiver (other trials). Thus, participants passively observed screens indicating that themselves or the anonymous receiver had gained or lost different amounts of reward. From this set of trials, participants were told that one of them would be randomly selected and added or subtracted to the receivers' or to their own earnings at the end of the study. However, to avoid reducing participants' compensation, a trial with a gain of £3 for themselves was always selected.

There were 12 different types of trials in the task according to a 2 (target: self, other) x 6 (value: -£15, -£9, -£3, £3, £9, £15) design. Each trial consisted of a screen indicating "You *OR* Receiver / Won *OR* Lost / Value" presented on a light grey background for 1400 ms followed by a 100 ms interstimulus interval. Furthermore, catch trials with the same duration were also presented to ensure attention to the task, as well as null trials presented for 3000 ms. The trial presentation order was pseudo-randomised using a "type 1 index 1″ sequence which ensures that each stimulus is preceded and followed by all other stimuli (Drucker and Aguirre, 2009; Aguirre, 2007). For the purpose of this experiment, the four most efficient "type 1 index 1″ sequences were chosen amongst 1000 randomly generated sequences using the Python code available online (https://cfn.upenn.edu/aguirre/wiki/public:t1i1_sequences/). Each of the four chosen sequences included 14 presentations of each stimulus – 14 repetitions x 14 trials (12 reward trials + one catch trial + one null/blank trial) - for a total of 196 trials. Since the full experiment included four sequences, 56 presentations of each stimulus were included with a scanning time of approximately 21 min across the four runs. The presentation order of the four sequences was counterbalanced across participants.

Catch trials were included to ensure that participants maintained attention across all trials without influencing the processing of the magnitude and the target of the gain or loss. During these catch trials, participants were asked to indicate as quickly as possible if recipient of the previous trial was the self or the receiver (i.e. "You or Receiver?") or to identify the amount that was awarded in the previous trial compared to a randomly generated amount (e.g. "Won £15 or Lost £9″) using a response box. Accuracy at the catch trial task was computed by counting the number of correct answers and dividing it by the total number of catch trials. Participants who had an accuracy below 50% of the catch trials were excluded from analysis.

2.2.2. Behavioural session: prosocial effort task

The prosocial effort task measures how motivated participants are to obtain rewards for self and others (Lockwood et al., 2017, 2022, 2021). Participants in this task were instructed that, as Deciders, they would be paired with one of the Receivers of the experiment, but not the same one they had in the fMRI session. The task consisted of participants making decisions between options with different magnitudes of financial rewards (represented by number of credits) in exchange for different levels of physical effort (grip force). For each trial, participants chose between two options: a rest baseline option associated with no effort and low reward (1 credit); and a work offer option, which results in higher monetary gain (2, 4, 6, 8 and 10 credits) for higher effort that varies across 30, 40, 50, 60 or 70% of each participant's maximum voluntary contraction (MVC). In half of the trials the reward was obtained by participants themselves (self trials), while in the other half rewards were received by the Receiver (other trials). If the work option was chosen, participants needed to make the effort required to obtain the credits on offer - they needed to squeeze a handle with the required force with their dominant hand for 1 s out of a 3 s window. They received zero credit if they failed to do so. If the rest option was chosen, participants rested for the same amount of time and obtained the single credit in exchange. The effort and reward levels in the work option varied independently over trials, and each effort–reward combination was repeated three times per beneficiary condition, giving a total of 150 trials: 75 self trials and 75 other trials. The rest option with one credit was used to make sure that there was a conscious and motivated decision to choose the alternative work option. If a choice was not selected, zero credits were given. All trials had the same duration, controlling for potential temporal discounting effects. Participants who did not actively choose any of the options for more than 10% of the trials were excluded from analysis.

Before participants made decisions in the prosocial effort task, they were asked to grip a handheld dynamometer with as much force as they could to determine their MVC. Thus, the task measures similar effort levels across participants regardless of their variability in strength. After MVC estimation and prior to the main decision task, participants completed 18 trials where they experienced each effort level three times, and also learned to associate each level of effort with the elements in the pie chart (e.g. one element of the pie chart corresponded to 0% force, i.e. the rest option).

2.2.3. Behavioural session: interoception respiratory task

After the prosocial effort task, participants completed the interoception respiratory task. The goal of this task (Murphy et al., 2018) was to assess how much participants rely on their internal signals relative to exteroceptive cues when assessing the force of their exhalation. To assess participants' force of exhalation, a standard peak flow metre was used. Participants who had a history of breathing difficulties did not undertake this task. On each trial, participants were first required to make a large, fast exhalation into the peak flow metre. This first exhalation was taken as the 'standard' for that trial, i.e., 100% performance. Participants were then required to perform a second exhalation. In this second, 'target' exhalation, participants were told to aim to perform a percentage of the standard exhalation (e.g., 30%) for that trial. There were four possible targets for each trial - 30, 50 70 or 90% of the standard. Once the participant had performed the target, they were asked by the experimenter to estimate their performance as a percentage of the standard. The accuracy of their estimate (difference between their estimate of the force of their second blow as a percentage of the standard and the objectively-measured force of their second blow as a percentage of the standard, see Eq. (1)) served as the dependant variable for each condition.

Trials in the interoception respiratory task were completed under two conditions. In the *internal condition*, participants performed the standard and the target exhalations while hearing background white noise through headphones connected to a laptop (~79 dB), ensuring they were unable to hear auditory cues produced by their exhalation, and thus could rely only on their internal signals when judging exhalation force. In the *external condition*, participants performed each exhalation with a background white noise coming from a laptop loudspeaker (~79 dB)

located approximately one metre from their right ear. Thus, in the external condition, although participants had an equally-distracting auditory input as in the internal condition, this auditory input did not masque the noise of their exhalation. As a result, participant could rely on auditory external cues when judging the force of their exhalation. For both conditions, the white noise started approximately one second before the exhalation, lasting for four seconds. Contrasting the performance in each of these conditions allows to measure the extent to which participants rely on interoceptive (relative to exteroceptive) signals. If in the external condition, where both internal and external signals are available to judge the force of one's exhalation, a participant only uses external cues, then their performance is likely to differ markedly in the internal condition where external cues are unavailable. Conversely, if a participant relies entirely on internal signals to judge the force of their exhalation in the external condition – even though external cues are also available – then their performance should be relatively unaffected in the internal condition as internal signals are still available to be used. In the latter case, the external condition could even be distracting for an internally focused person and therefore showing a lower performance compared with the internal condition.

During the task, and across conditions, participants were blindfolded, to prevent them using visual information to aid performance. For both internal and external conditions, participants completed six repetitions of the four targets presented in a random order, with a total of 24 trials per condition. The order in which the internal and the external conditions were presented was counterbalanced across participants. To ensure that the 30% target trial could be measured, participants were required to surpass a threshold of 200 L/min in their standard exhalation. If they did not accomplish this, the standard blow was repeated until the threshold was surpassed. No feedback about participants' performance was provided across the experiment.

The peak-flow metre into which participants performed their exhalations was gently secured in a horizontal position using a vice clamp and elevated in line with each participant's mouth using a stand. Participants were instructed to keep their hands resting at the bottom of the stand during exhalations, using their hands only to locate the gauge prior to performing exhalations, and to be still in the chair and sat upright, without pushing the mouthpiece forward while exhaling. Trials in which these conditions were not accomplished were repeated or removed from analysis. Participants with more than 10% of missed trials in either of the two conditions were excluded.

2.3. Analysis of behavioural data

2.3.1. Interoception score: reliance on internal vs external signals

To calculate participants' reliance on their interoceptive signals, the absolute error scores were computed for each trial of the interoception respiratory task, such that:

$$AE_{ij} = \left[\frac{T_{ij} - E_{ij}}{T_{ij}}\right]$$
$$\overline{AE_j}\begin{cases}\overline{AE}_{int} \; for \; internal \; trials \\ \overline{AE}_{ext} \; for \; external \; trial\end{cases} \quad (1)$$

Where the absolute error AE in the trial i for the participant j is the absolute difference between the actual performance of the participant in target T as a percentage of the standard, and their estimation E, divided by their actual performance T. The mean of the AEs is computed separately for the internal and the external conditions for each participant. The interoception score was then the difference in performance between the internal and the external condition, such that:

$$Interoception_j = \overline{AE}_{ext[j]} - \overline{AE}_{int[j]} \quad (2)$$

Where the interoception score for the participant j is the difference between their mean AEs in the external condition and in the internal condition. Scores below zero indicate that a participant made more mistakes in the internal than the external conditions, suggesting that they are more externally-focused. The interoception score therefore reflects how much participants rely on internal versus external signals (**Supplementary Fig. S1A**).

2.3.2. Association between interoception and motivation

For the prosocial effort task, choices to work relative to rest were taken as an index of motivation to obtain self and other rewards. Mixed effects models were used to predict trial-by-trial decisions using the *glmer* function in R. Thus, two models were built to test whether interoception was linked to participants decisions, and to which variable (effort or reward for self or for other) in their motivation, such that:

Simple model

$$DW_i = \beta_{0j[i]} + \beta_{1j[i]}R_i + \beta_{2j[i]}E_i + \beta_3 B_i + \beta_4 R_i E_i + \beta_5 R_i B_i + \beta_6 E_i B_i + \beta_7 R_i E_i B_i \quad (3)$$

Interoception model

$$DW_i = \beta_{0j[i]} + \beta_{1j[i]}R_i + \beta_{2j[i]}E_i + \beta_3 B_i + \beta_4 I + \beta_5 R_i E_i + \beta_6 R_i B_i + \beta_7 R_i I + \beta_8 E_i B_i + \beta_9 E_i I + \beta_{10} B_i I + \beta_{11} R_i E_i B_i + \beta_{12} R_i B_i I + \beta_{13} E_i B_i I + \beta_{14} R_i E_i I + \beta_{15} R_i E_i B_i I \quad (4)$$

In the simple model, decision to work DW in the trial i was predicted by the fixed effects of reward R, effort E, beneficiary B and their interaction, with a random intercept clustered in each subject j. DW is a binary, factor variable. Random slopes for R and E were included as it is expected for these variables to vary across participants (Lockwood et al., 2017, 2022, 2021). The interoception model adds the interoception score I as a predictor together with its interaction with the other independent variables.

As a post-hoc analysis, we tested for Pearson correlations between the interoception scores and reward and effort beta values obtained for each participant from two mixed models where decisions to work were predicted by effort, rewards and their interaction. These reward and effort betas were used as a proxy for individual differences in participants' sensitivities to reward and effort in self and prosocial decisions separately. Crucially, each model considered beneficiaries separately for self and other. These two models were clustered with random-intercepts for each participant, and had random-slopes for reward and effort, such that:

$$DW_{self} = \beta_{0j[self]} + \beta_{1j[self]}R_i + \beta_{2j[self]}E_i + \beta_{3[self]}E_i R_i \quad (5)$$

$$DW_{other} = \beta_{0j[other]} + \beta_{1j[other]}R_i + \beta_{2j[other]}E_i + \beta_{3[other]}E_i R_i \quad (6)$$

Where DW is specific for self and for other trials separately. Thus, the interoception score was correlated with motivation to work represented by participants' beta estimates for reward and effort for self and other. For comparison between correlation coefficients, Fisher Z-transformation were performed using the online tool at http://vassarstats.net/rdiff.html.

2.4. fMRI acquisition and preprocessing

MRI data were acquired using a 3 Tesla Siemens Prisma MRI scanner. A structural scan was acquired at the start of the session using a magnetisation-prepared rapid gradient echo sequence with 192 slices; slice thickness: 1 mm; repetition time (TR): 1900 ms; echo time (TE): 97 ms; field of view (FOV): 192 × 192 mm; voxel size: 1 × 1 × 1 mm. Field map images were obtained immediately before the value representation task using a double-echo spoiled gradient echo sequence (TR: 488 ms; TE: 4.92/7.38 ms; voxel size: 3 × 3 × 3 mm; flip angle: 46°). Four runs of multiband T2*-weighted echo-planar imaging (EPI) volumes with BOLD contrast were collected during the value representation task (72 slices in interleaved ascending order; repetition time: 1570 ms; echo time: 30 ms; flip angle: 70°; field of view: 216 × 216 mm; matrix

size: 108 × 108; voxel size: 2 × 2 × 2 mm with 1 mm gap; multi-band factor: 3; in-plane acceleration factor: 2). Each run was 5:46 min in length, during which 220 functional volumes were acquired.

Preprocessing was performed following a standard protocol using fMRIPrep 20.1.0rc1 (Esteban et al., 2019), which is based on Nipype 1.4.2 (Gorgolewski et al., 2011). The details of preprocessing can be found in the **Supplementary Information**.

2.5. fMRI analyses

2.5.1. First level

First, a general linear model was created of the BOLD response at the participant level using SPM12, with regressors modelling the onset of each event, i.e. -£15, -£9, -£3, £3, £9, £15 for self and other, and the onset of the catch trials. The regressors were generated using delta functions convolved with SPM12's canonical haemodynamic response function. This first level model also included nuisance regressors modelling non-neural sources: (i) 24 movement-related regressors comprising the six estimated head movement parameters (x, y, z, roll, pitch, yaw), their first temporal derivatives, their squares, and their squared derivatives; (ii) the first six components of the anatomical CompCor extracted using the CompCor method (Behzadi et al., 2007) implemented in *fMRIprep*; and (iii) a variable number of single event regressors indicating volumes exceeding a threshold of 0.5 mm framewise displacement or 1.5 standardised DVARS and considered as motion outliers (mean% of volumes that were motion outliers = 1.04%, SD = 0.91%). The four runs were concatenated in the model and constant regressors were added to model each of the four run's mean. The null trials were used as the unmodelled baseline in all models. A high-pass filter with a cut-off of 128 s and SPM12's *AR(1)* correction for serial correlation were applied during the model estimation.

To calculate functional connectivity between regions of interest in various conditions using beta series regressions (Rissman et al., 2004). We also modelled the BOLD response for individual trials using the Least Square Single models approach (Mumford et al., 2012). We therefore fitted a GLM for trials of interest (672 trials; 56 presentations x 2 self/other x 6 amounts) in which the first regressor was the onset of the trial of interest, the second regressor the onset of all other trials within the session (including catch trials but excluding blank screens) and the other regressors were the nuisance regressors described above. To ensure that the single-trial regressors included in the analyses were not collinear with nuisance regressors, we calculated the variance inflation factor (VIF) for each trial and excluded trials with a VIF > 4 from all analyses. This removed on average 3% (SD = 6%) of trials across participants.

2.5.2. Second level

For group statistical analysis, whole-brain SPM analysis was performed using a single-sample *t*-test to examine the contrast of self > other, and other > self conditions averaged across all reward events, using a cluster-level probability threshold of $P_{FWE} < 0.05$, with clusters defined by the voxel-level threshold $P_{uncorrected} < 0.001$.

2.5.3. Similarity analysis

We measured multivariate similarity between self and other trials by calculating the Pearson correlation between the average parametric maps for each condition of interest, i.e., self and other, in five ROIs. These correlation coefficients for each participant were used to assess the relationship between individual differences in self/other neural similarity and performance on the interoception respiratory task and the prosocial effort task. Null correlations between similarity in these ROIs and participants' mean framewise displacement suggest that these values were not a product of participants' movements in the scanner (see **Supplementary Information**).

We based our analyses on the AI and the ACCg as ROIs. For the AI portions, we used the parcellations created by Deen et al. (2011) derived from a voxel-wise k-means clustering approach applied to resting-state. These parcellations divide the insular cortex in three different complexes for each hemisphere: posterior insula, dorsal AI and ventral AI. The 6 insula ROIs were acquired in 2 mm Montreal Neurological Institute (MNI) space directly from the author's website (https://bendeen.com/data/). Within the AI there is putatively more than one anatomical zone, with a particular distinction in function between dorsal and ventral AI. However, it is unclear which of these sub-regions may be linked specifically to vicarious reward processing. As such, we used right and left vAI and dAI ROIs.

For the ACCg region, we used thresholded masks taken from the resting-state connectivity-based parcellation by Neubert et al. (2015). This parcellation divides the frontal cortex into 21 different regions. The parcels were acquired in 2 mm MNI space directly from the author's website (http://www.rbmars.dds.nl/CBPatlases.htm). We created the ACCg masque by modifying the original parcel using the *imcalc* function in SPM12. Thus, the left hemisphere parcel for the area 24ab was duplicated onto the right hemisphere to create a bilateral masque. Furthermore, those voxels located in the posterior portion of the masque were removed to capture more closely what the literature refers to as ACCg, corresponding to the gyral portion of the anterior mid-cingulate cortex (Apps et al., 2016; Vogt et al., 1995). Notably this masque lies inferior to the cingulate sulcus and does not extend laterally. As such, given the morphological changes that occur when there is the presence or absence of an additional paracingulate sulcus (Paus et al., 1996), it is likely to capture area 24a/b in most hemispheres of most participants. Although this approach to deriving masks anatomically has benefits, it should be noted that other methods – such as deriving the masks from resting-state scans from each participant (Beckmann and Smith, 2004) – may have offered more precise masks at the individual subject level.

Similarity values obtained across these five ROIs were predicted by the interoception score Eq. (2)) and behavioural indexes in the prosocial effort task. For the latter, we used the reward and effort beta-weights obtained for each participant from the two mixed models previously described (Eqs. (5) and ((6)) where decisions to work were predicted by effort, rewards and their interaction separately for self and other. For each regression model where ROI self/other similarity was predicted by behavioural measures, participants who had similarity values below and above three standard deviations were excluded to account for potential artefactual similarity values. For the similarity values in RdAI, RvAI, LdAI and LvAI, robust regression models were used using the *rlm* function in R. For the similarity values in the ACCg, linear regression models were used instead as values in this region were normally distributed. All behavioural measures were normally distributed. Results obtained from these models were corrected for multiple comparison across the five ROIs using false discovery rate (FDR) (Benjamini and Yekutieli, 2001; Benjamini and Hochberg, 1995).

2.5.4. Functional connectivity analyses

We used beta series regressions (Rissman et al., 2004) to assess individual differences in the functional connectivity between the ROIs in each task condition. Specifically, for each participant, we calculated the mean value in ACCg and RdAI at each trial and calculated the regression coefficient between their combination across all trials for self and other. This allowed us to obtain a beta estimate parameter indicating the functional connectivity between ACCg and RdAI for each participant and condition. Paired *t*-test was used to test for differences in connectivity between self and other trials. Prior to this analysis, participants who had similarity values in both the ACCg and RdAI below and above three standard deviations were excluded ($n = 3$). Finally, a linear regression model was built having connectivity between RdAI and ACCg in other trials as a dependant variable, and the interoception score and other reward beta (Eq. (6)) as predictors.

3. Results

3.1. Behavioural results: reliance on interoceptive signals is associated with prosocial motivation

We first tested whether reliance on interoceptive respiratory signals (see Eq. (2)) was associated with motivation for self or/and others' rewards. A mixed effects model predicting choices to work or rest in the prosocial effort task was conducted including level of effort, magnitude of reward, and beneficiary (self vs other), together with their interactions (up to three-way), as fixed effects predictors. Random effects were also included, with random intercepts at the subject level, and random slopes on effort level and reward magnitude (see Eq. (3)). Model comparison revealed that including participants' interoception scores from the respiratory task as an additional between-subjects predictor of choices in the prosocial effort task (see Eq. (4)) improved model fit, both when penalising the model for additional parameters (AIC: simple model = 5115.5, interoception model = 5079.3; **Supplementary Fig. S1B**) and when performing a ratio test on model log-likelihoods (χ^2_{diff} = 52.28, df_{diff} = 8, $p < 0.001$). Thus, decisions to work in the prosocial effort task were better explained by a statistical model that included participants' reliance on internal signals.

We hypothesised that people who relied more on internal signals would be more motivated to act prosocially, and that this would be specifically linked to how incentivised people are by others' rewards. Consistent with this, we found a three-way interaction between interoception score, reward and beneficiary in the mixed effects model (b = 0.42, SEM = 0.11, z = 3.87, $p < 0.001$, Fig. 2A and 2B). Thus, during self trials participants were more likely to work as the rewards on offer increased, but this effect was present regardless of their interoception score (Fig. 2A). In the other (prosocial) trials, there was an effect of interoception, whereby people who relied more on interoceptive signals were more incentivised to work at higher reward levels for others. Conversely, externally-focused people were not incentivised to choose to work more often as the rewards that would be received by the other person increased (Fig. 2B).

In addition, there was a two-way interaction between interoception and beneficiary ($b = 0.65$, SEM = 0.12, $z = 5.3$, $p < 0.001$), such that participants who relied more on internal than external signals were more willing to work for others regardless of the rewards on offer or effort required (**Supplementary Fig. S2A**). Notably, despite previous research suggesting a link between interoception and people's sensitivity to effort, there was no such interaction in our data (see **Supplementary Table S1** for a full description of these results). Finally, the remaining findings of the mixed model replicated the results of previous studies using the prosocial effort task (Lockwood et al., 2017; Contreras-Huerta et al., 2020a; Lockwood et al., 2022, 2021), with significant effects of effort, reward and beneficiary, as well as interactions between reward and beneficiary, and effort and beneficiary (see **Supplementary Fig. S3A** and **S3B)**. In sum, we show that how much someone relies on internal signals is linked to how prosocial they are and not to their motivation to work for themselves. Specifically, more internally-focused individuals are more incentivised by the rewards that can be obtained for another person, and work more when they can obtain a bigger benefit for them.

Given that in the prosocial effort task the majority of people choose to work more at lower reward levels for themselves than for other people, this leads to a prediction that people with more positive interoception scores (i.e. more internally focused, relying more on interoceptive signals) would show less of a difference in decisions to work between self and other, as a function of reward. Furthermore, given the results of the mixed model above, greater interoception scores should be specifically associated with sensitivity to others' rewards and not with sensitivity to self rewards nor sensitivity to effort. Thus, as a confirmatory, post-hoc analysis, we extracted beta parameters for effort and reward from two mixed effect models where decisions to work or rest were taken separately for self and other trials (see Eqs. (5) and (6)). Thus, these indexes were a proxy for individual differences in sensitivities to effort levels and magnitudes of rewards for self-benefitting and prosocial decisions.

From the results described above, we expected interoception to be associated with sensitivity to others' rewards and not to effort or self rewards. We tested this using Pearson correlations between the interoception score, and reward and effort betas for self, other and their difference. Results revealed a significant negative correlation between interoception and the difference in reward betas, such that people who relied more on internal signals were more similarly motivated for self and other rewards, while people who were more externally focused valued more their own rewards compared to others' ($r_{(59)} = -0.39$, $p < 0.003$, Fig. 2C). To test if this effect of interoception on prosocial motivation was specific to sensitivity to rewards, we correlated difference in effort betas with interoception. We found null correlation ($r_{(59)} = 0.15$, $p = 0.26$), which was significantly different from the correlation between interoception scores and sensitivity to reward (Fisher's Z transformation, $z = -3$, $p < 0.003$). As expected, the interoception score also correlated with sensitivity to others' rewards ($r_{(59)} = 0.27$, $p < 0.04$, Fig. 2D), but not with self rewards ($r_{(59)} = -0.14$, $p = 0.29$; **Supplementary Fig. S2B**), and these effects were significantly different from each other ($z = 2.23$, $p < 0.03$), suggesting specificity of the social effects. Furthermore, interoception did not correlate with neither other ($r_{(59)} = -0.08$, $p = 0.54$) nor self effort betas ($r_{(59)} = -0.13$, $p = 0.32$), and these correlations showed a significant difference ($z = 2.19$, $p < 0.03$) and a non-significant trend from its effects on other reward betas ($z = 1.92$, $p = 0.055$), suggesting that interoception is specifically associated to sensitivity to others' rewards in the effort task. Taken together, these results reveal that people who rely more in internal versus external signals are specifically more driven by others' rewards when deciding whether to expend energy to act prosocially.

3.2. fMRI results: different roles of the ACCg and the right dorsal AI in vicarious rewards

Next, we examined whether the degree of similarity between neural patterns evoked by self and other rewards during the fMRI value representation task was predicted by prosocial motivation and interoception. Multivariate similarity analysis was calculated as the Pearson correlation between the average parametric map for self and other rewards (Kriegeskorte et al., 2008), representing the spatial similarity between the patterns of activation for self and other rewards across the different reward levels for each participant. Our hypotheses specifically related to the ACCg and AI due to their unique profiles of being linked to interoception, prosociality and vicarious reward processing (Lockwood, 2016; Harbaugh et al., 2007; Morelli et al., 2015; Apps et al., 2016; Critchley et al., 2004; Craig, 2002). We therefore included five ROIs: four AI portions - right dorsal AI (RdAI), right ventral AI (RvAI), left ventral AI (LvAI) and left dorsal (LdAI) -, and the ACCg (bilateral). Traditional whole-brain univariate analyses performed on the same data are reported in **Supplementary Tables S2 and S3**.

Five regression models were used to test whether the degree of similarity between self and other representations of reward was associated with interoception, one in each ROI, with the similarity of self and other representations being the dependant variable and participants' interoception score (reflecting their reliance on interoceptive signals, see Eq. (2)) as a predictor. Only one area, the RdAI, showed a significant effect ($n = 54$, $b = -1.01$, $F_{(1,52)} = 11.12$, $p < 0.008$ FDR corrected). No effects were found in any of the other ROIs. Within the RdAI, less similar responses between self and other rewards were linked to people relying more on their internal signals (Fig. 3A), suggesting that a greater specialisation of neural responses to others' rewards in this region occurs in people who are more aware and use interoceptive respiratory signals.

Next, we hypothesised that variability in vicarious reward responses would be related to how incentivised a participant was by others' rewards in the prosocial effort task. We used the same regression approach in each ROI, to test whether similarity in the neural response between

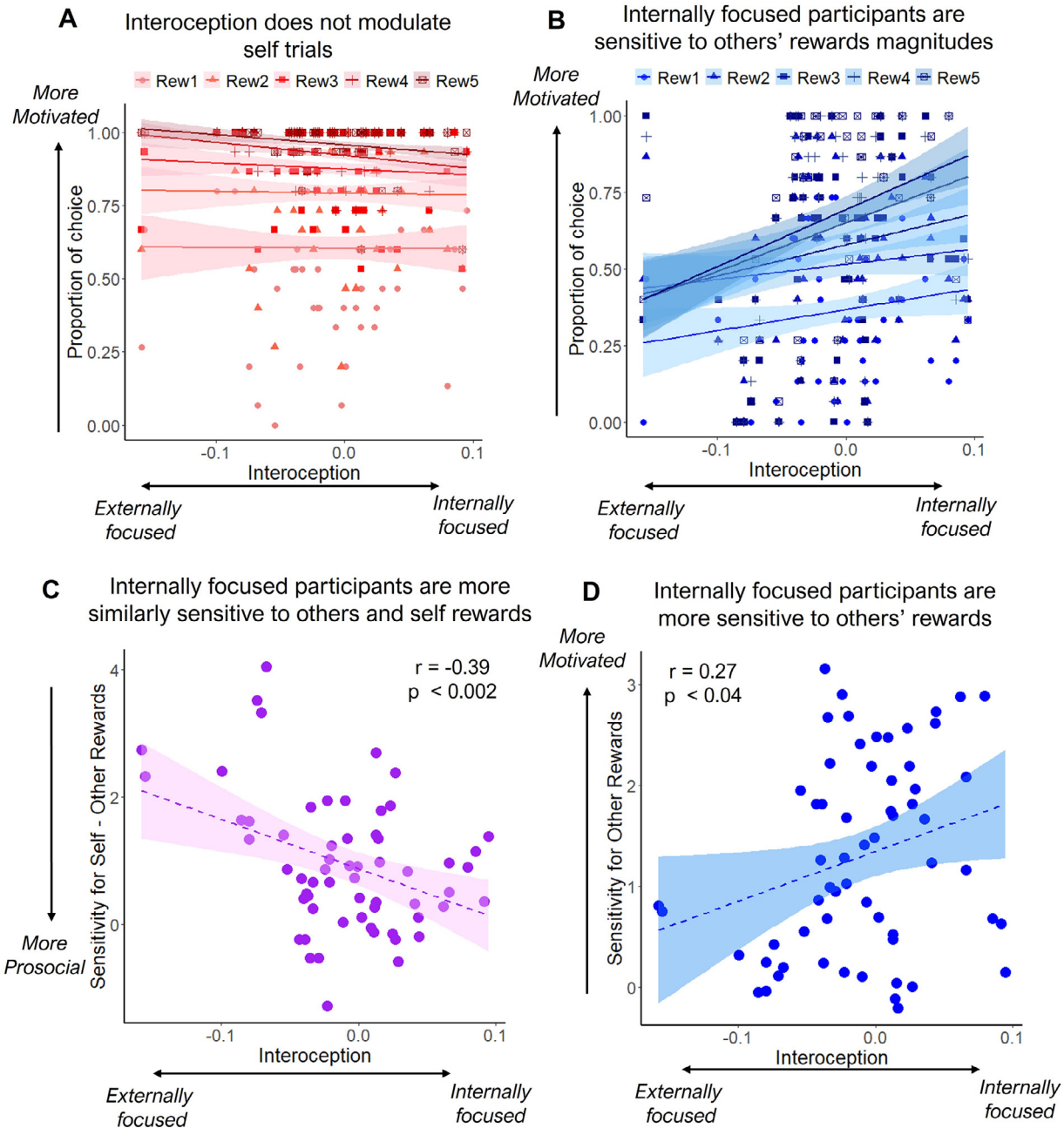

Fig. 2. Reliance on interoceptive signals is associated with prosocial choices. A. Interoception score did not influence motivation for rewards in self trials. Participants did not show different patterns in their proportion of choices to work over rest (y-axis) depending on their level of reliance on internal vs external signals (x-axis) across different reward magnitudes. B. Interoception influences motivation for others' rewards. As participants relied more on internal than external cues, they chose to work more as the reward to be received by another person increased. Participants who relied more on external cues were more reluctant to work to benefit others regardless the reward on offer. Shaded areas show the 70% confidence interval around the slopes. Individual points show the score of each participant for each condition. C. People who rely more on internal vs external signals (x-axis), weighted rewards more similarly for self and other when making decisions to work to obtain rewards. Y-axis depicts the difference between self and other betas from two mixed models predicting choices separately for self and other. Higher values indicate self rewards are valued more than others'. D. Participants who relied more on internal vs external cues (x-axis), were more sensitive to others' rewards (beta estimates, y-axis). Shaded areas show the 95% confidence interval around the slopes. Individual points show the score of each participant.

self and other reward was related to a beta-weight reflecting someone's sensitivity to others' rewards in the prosocial effort task (see Eq. (6)). We found a significant effect in only one ROI, the ACCg ($n = 55$, $b = 0.07$, $F_{(1,53)} = 7.36$, $p < 0.05$ FDR corrected, Fig. 3B). Within the ACCg, greater similarity in responses between self and other reward was predictive of increased incentivisation by others' rewards in the prosocial effort task. That is, people who showed a greater increase in choosing to help others as the rewards on offer increased, showed more similar neural patterns between self and other rewards. Note that both the results within the ACCg and RdAI were also present independently of the statistical method used to relate them, with similar results in both areas when Pearson correlation instead of regression models were used (see **Supplementary Information**). Furthermore, these results were largely consistent even when regression models were performed taking gains and losses

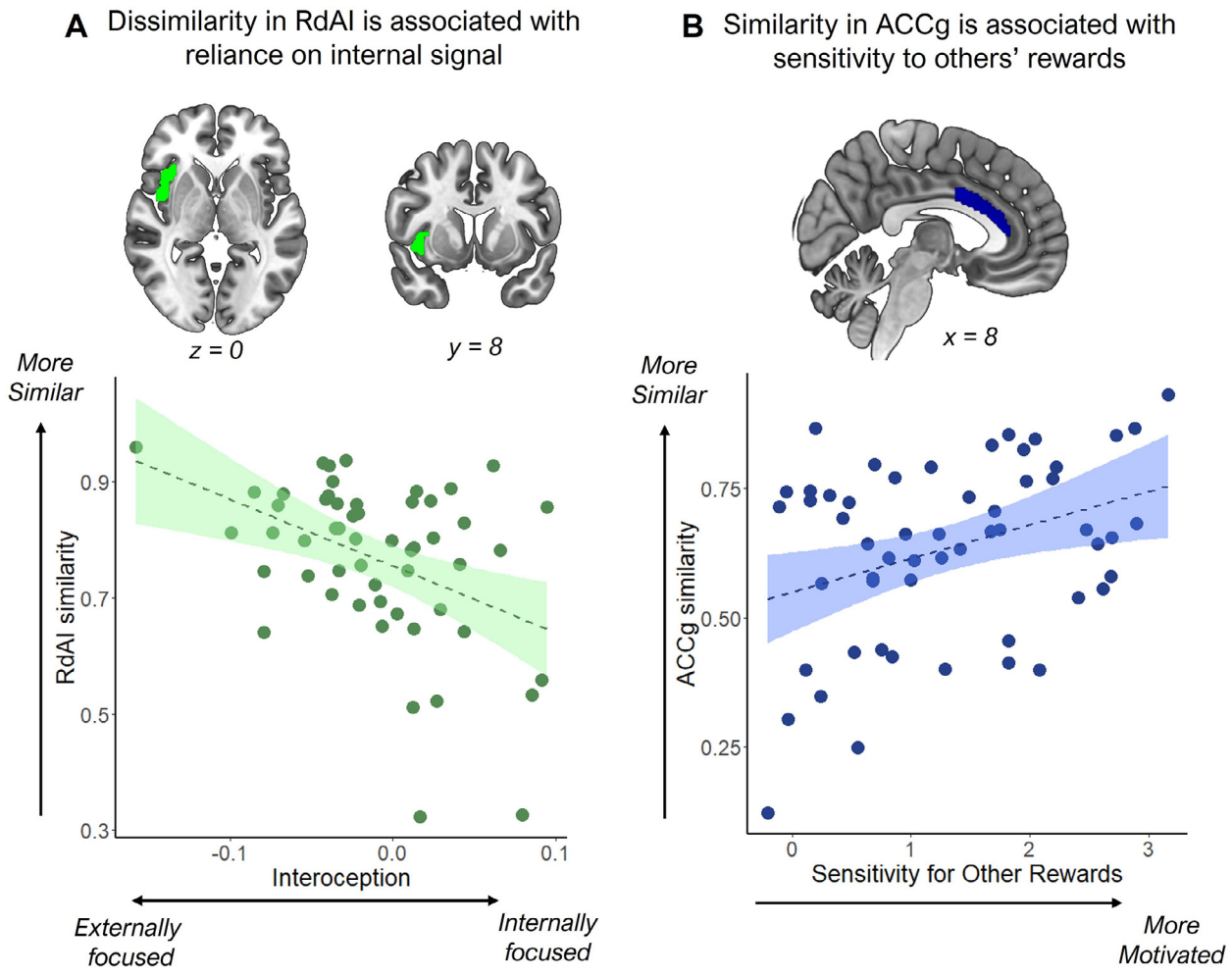

Fig. 3. Similarity between self and vicarious reward responses in the right dorsal anterior insula (RdAI) and the gyrus portion of the anterior cingulate cortex (ACCg). A. Interoception was associated with multivariate similarity between self and other rewards only in the RdAI, amongst five ROIs ($p < 0.008$ FDR corrected). People who relied more on internal relative to external signals showed more dissimilar neural responses to reward between self and other. Y-axis depicts the similarity values, with higher values meaning more similarity between self and other reward representation. B. Prosocial motivation - specifically how incentivised a participant was to obtain rewards for others - was associated with multivariate similarity between self and other rewards only in the ACCg ($p < 0.05$ FDR corrected). Participants who were more motivated to work for others' rewards showed more similar neural patterns between self and other rewards in the ACCg. X-axis corresponds to the reward betas obtained from a mixed model where decisions to work for others in the prosocial effort task was predicted by effort, reward and their interaction. Positive values indicate higher weights for others' rewards.

separately (see **Supplementary Information**). Finally, in both ACCg and RdAI, similarity between self and other was significantly above zero (**Supplementary Table S4**).

Analyses of the behavioural data revealed that interoception and sensitivity to others reward in the effort task were correlated, suggesting shared variance. However, analyses of the neural data demonstrated different effects of interoception and prosocial behaviour in distinct ROIs. In order to investigate whether the effects in the ACCg and RdAI remain when this shared variance is included in the model, two regression models were conducted, one in each region, in which the beta weight measuring incentivisation by other rewards from the effort task, and the interoception score, were both included as predictors of neural similarity. Both the effect of other reward in the ACCg ($b = 0.06$, $F_{(1,52)} = 6.25$, $p < 0.02$) and interoception in the RdAI remained significant ($b = -1.21$, $F_{(1,51)} = 14.74$, $p < 0.001$). As such, although people's reliance on interoception and motivation by others' rewards are related to each other, the responses in the ACCg and RdAI are linked specifically to prosocial motivation and interoception, respectively.

An advantage of using the prosocial effort task is that it also measures how sensitive people are to their own benefits (self reward), as well as how costly they find effort both when benefiting themselves (self effort) and others (other effort; see Eqs. (5) and (6)). To test the specificity of the effects above, we conducted two further regression models, one for the ACCg and one for the RdAI, including those three predictors. We found no significant effects in the ACCg nor in the RdAI for any of these predictors even at uncorrected thresholds (see **Supplementary Tables S4** and **S5** for details). Thus, vicarious rewards processing in the ACCg was specifically linked to how incentivised people were by others' reward in the prosocial effort task, and similarity in the AI was linked specifically to interoception.

Taken together, vicarious reward signals in the RdAI and ACCg were linked to different processes: interoception, and motivation to obtain benefits for others, respectively. However, behaviourally, reliance on interoceptive respiratory signals was associated with higher motivation to obtain rewards for others in the effort task. Crucially, this link was specific for others' rewards and not for self. Thus, the connectivity between RdAI (linked to interoception) and ACCg (linked to prosocial motivation) in reward processing might be stronger when evaluating others' than self rewards. Could connectivity between ACCg and RdAI underlie this link between prosocial motivation and reliance on internal signals?

We used beta series regressions (Rissman et al., 2004) to assess individual differences in the functional connectivity between the ROIs in each task condition (self and other). Specifically, for each participant, we calculated the mean value for ACCg and RdAI during each trial and calculated the regression coefficient of those values across all trials and within each condition of interest. This allowed calculation of a beta estimate indicating the functional connectivity between ACCg and RdAI for each participant and condition, enabling comparison of the degree of connectivity between ACCg and RdAI in the self and the other condition. Significantly higher connectivity was found in the other compared with the self condition ($t_{(52)} = 3.72, p < 0.001$, **Supplementary Fig. S4**), suggesting that these areas are functionally more associated when reward events occurred to another person compared to participants themselves.

Next, we tested whether connectivity between RdAI and ACCg in response to other rewards was associated with reliance on interoceptive signals and/or sensitivity to others' rewards in the prosocial effort task. Thus, a regression model in which the dependant variable was neural connectivity between these regions, having the interoception score and other reward betas in the prosocial effort task as predictors, did not reveal significant effects (see **Supplementary Table S6** for details).

4. Discussion

We often observe good and bad outcomes for other people. However, how this experience of vicarious processing relates on our interoceptive signals or how willing we are to engage in prosocial behaviours remains largely unknown. Our results show that even when simply seeing rewarding outcomes for others, variability in vicarious processing in the ACCg and AI is linked to someone's prosocial tendencies and their propensity to be aware and use respiratory interoceptive signals, respectively. In particular, whilst greater dissimilarity between neural responses to self and vicarious rewards was associated with increased reliance on interoceptive signals in the AI, greater similarity in the ACCg was linked to being more incentivised by rewards that could be obtained for others by prosocial acts. Functional connectivity between these regions was higher when participants observed others receiving reward outcomes than when they did themselves. Thus, these results indicate that processing of vicarious rewards is linked to levels of respiratory interoception and prosocial behaviour. In addition, at the behavioural level, how incentivised and motivated people are to obtain rewards for others relates to the propensity to be aware and use respiratory interoceptive cues, with more prosocial people being more focused on internal signals. This effect was specific, with reliance on internal signals not related to motivation to benefit oneself, or to the effort required to act prosocially. Taken together, these results suggest that variability in the patterns of neural responses to rewards are associated with interoception and prosocial behaviour.

Whilst both the AI and ACCg have been linked to vicarious reward processing (Lockwood, 2016; Apps and Ramnani, 2014; Morelli et al., 2015; Lockwood et al., 2015; Apps et al., 2013a), our results suggest that they might have different involvement when processing others' rewards, linked to interoception and motivation, respectively. Specifically, only the AI processing of rewards was associated with interoceptive respiratory signals. Previous research has suggested that the AI can be thought of as a secondary interoceptive cortex, due to the fact that it is heavily connected to regions that receive afferent signals from a number of sensory organs (Craig, 2009; Mesulam and Mufson, 1982; Singer et al., 2009; Tsakiris, 2010). It has been argued that the function of AI is therefore to integrate interoceptive signals with other cognitive and affective processes (Craig, 2009; Singer et al., 2009; Farb et al., 2013). However, a limited amount of research has demonstrated a link between reward signals, that have often been found in the AI, to interoceptive processes. By using a multivariate technique, we were able to show that vicarious reward signals, specifically in the RdAI, are linked to reliance on interoceptive respiratory signals, leading to a possibility that responses to vicarious rewards might reflect a stable trait for how much someone's internal signals guide their behaviour.

Interestingly, the association between interoception and AI was specifically located in its dorsal part in the right hemisphere. This is consistent with previous work that proposes functional and anatomical division in the AI (Uddin et al., 2017; Uddin et al., 2014). Thus, its ventral portion is believed to be mainly involved in affective reactivity toward salient outcomes impacting self and others, while its dorsal area (especially in the right hemisphere) has been more specifically linked to interoception, providing a bodily map for a wide range of mental processes (Critchley et al., 2004; Craig, 2009, 2002; Critchley, 2005; Uddin et al., 2017, 2014; Nomi et al., 2018, 2016). Posterior insula, involved in primary interoceptive representation of the physiological state of the body through thalamocortical pathways, has strong connections with dAI, and both share some of their connectivity fingerprints (Craig, 2009, 2002; Critchley, 2005; Uddin et al., 2017; Nomi et al., 2018, 2016). Notably, previous research that has linked the AI to interoception has mainly used cardiac based tasks, rather than respiratory. As the AI is largely association cortex, it does not receive direct input from different interoceptive systems, but instead receives input from a number of primary interoceptive regions, placing it as a region that may integrate information across different interoceptive modalities. Indeed, the RdAI has been suggested to be an integrative hub for exteroceptive and interoceptive signal, including respiratory and cardiac domains (Farb et al., 2013). Moreover, through its connections with areas such as the prefrontal cortex, the ACC and the rest of limbic system (Craig, 2009, 2002; Critchley, 2005; Uddin et al., 2017, 2014), the dAI is crucial for self-awareness, where multiple sources of information are integrated to represent body-ownership and sense of agency (Craig, 2009, 2002), which might be key for the distinction between self and other processes (Palmer and Tsakiris, 2018). Previous results indirectly suggest that the right AI, including its dorsal portion, have different mechanisms underlying vicarious and self outcomes (Brethel-Haurwitz et al., 2018; Corradi-Dell'Acqua et al., 2016; Crockett and Lockwood, 2018; Hu et al., 2021; Rütgen et al., 2015), supporting the notion that RdAI might have representations of vicarious rewards anchored to bodily changes, important for self-other differentiation.

Notably, our results suggest that response to others' reward in the ACCg might better reflect people's motivation to perform prosocial acts than their interoceptive propensity. Classical accounts have suggested different roles for these regions, with the insula more implicated in affective and bodily processes, while frontal areas, including the ACC, more involved in motivating behaviour (Craig, 2009; Mega and Cohenour, 1997; Vogt, 2009; Le Heron et al., 2018). However, the presence of vicarious reward signals in both the AI and ACCg did not seem to fit with such role division. Here we show that it is the degree to which the AI represents others' rewards as distinct from self rewards that is associated with interoceptive signals, and the same distinctiveness in ACCg that is associated with levels of motivation, specifically the willingness to help others. As such, our results expand knowledge of AI and ACCg functions, suggesting distinctive involvement in their vicarious signals of reward. Importantly, future directions should disentangle whether vicarious rewards representations have distinct mechanisms between gains and losses in their influence on behaviour. Previous studies have shown that reward gains and losses can facilitate different patterns of economic decisions, which can be linked to differences in neural activity (Rangel et al., 2008). Here, we took gains and losses together, given previous literature on ACCg and AI function on vicarious rewards (Lockwood, 2016; Apps et al., 2013a), with the overall pattern of result largely consistent when examining gains and losses separately. Therefore, future work should address these questions, aiming to understand similarity and differences between self and other gains and losses.

Although the classical account also posits that medial frontal cortex is involved in motivating behaviour, it is less clear that there would be a

link specifically between vicarious signals in the ACCg and prosocial motivation. Previous studies in macaques have suggested that lesions that encompass the ACCg as well as surrounding portions of the ACC prevent monkeys from executing previously-learned prosocial behaviours, whilst leaving self-benefitting behaviours intact (Basile et al., 2020). Local field potentials and single-unit recordings have shown that neurons in the ACCg respond specifically when monkeys make a choice to deliver rewards to another rather than to themselves (Chang et al., 2013; Dal Monte et al., 2020). However, despite a number of studies showing vicarious reward signals in the ACCg, and that variability in these signals is linked to self-reported empathy and disrupted in autism spectrum disorders (Lockwood et al., 2015; Wittmann et al., 2018; Balsters et al., 2017), very few studies in humans had linked these signals to directly-measured variability in prosocial behaviour. Here, we found that variability between people in the distinctiveness of ACCg vicarious reward signals was linked to how willing someone was to engage in prosocial acts, directly linking vicarious signals in this region with prosocial motivation.

Although these results in the ACCg and RdAI are correlational and require replication, they directly relate to broader discussions about whether common-currency or specialisation is more important for social and interoceptive processes (Ruff and Fehr, 2014; Apps et al., 2016; Lockwood et al., 2020; Palmer and Tsakiris, 2018; Ainley et al., 2016). Usually, studies researching social information processing examine overlap between clusters responding to self or other rewards in a univariate analysis, or examine clusters that show a difference between self and other at the population level. Less work has examined individual differences in how distinct self and other reward processing is within a region, and there are very few studies that have used multivariate techniques that directly test how similar responses are between self and other rewards.

Recently, it has been demonstrated that multivariate pattern analysis techniques may be more robust for examining individual differences than traditional univariate statistical approaches (Kragel et al., 2020). Here, by deploying this technique we showed that, within AI, specialisation for processing others' rewards – less similarity in patterns between self and other reward – was related to greater reliance on respiratory interoceptive cues. As such, specialisation and identification of self and other rewards might be a distinguishing feature related to how AI processes interoceptive information. In contrast, in the ACCg, there was evidence that greater similarity in patterns between self and other reward was linked to increased prosociality in the effort task. Although this might argue in favour of a common-currency account, it is notable that the multivariate technique examines spatial disparity in terms of patterns, rather than overlap in voxels as examined with traditional methods. This work highlights that neither account in isolation can explain how the brain processes social information.

The complexity of domain-general and socially-specific processes is also reflected in how the interaction between interoception and prosocial behaviour might be implemented in the brain. We did not find any strong evidence in our neural data for why interoception and prosocial behaviour are linked behaviourally. We did find that ACCg and RdAI are more functionally connected when people observe others' rewarding outcomes compared to self. Speculatively this suggests that the link between interoception and prosocial behaviour could be driven by the connectivity between regions. However, future work will need to unpack how the neural mechanisms that are present during prosocial behaviour, are linked to interoceptive processes and thus why those two processes relate.

It is important to note that the results suggest that the complexity of mechanisms in the brain when processing rewarding stimuli, and whether greater similarity or dissimilarity links the processing to bodily signals to social behaviour, may be much greater than existing accounts suggest. It is entirely plausible that more dissimilarity between self and other is necessary for some processes and greater similarity between self and other for others (Lockwood et al., 2020). As such, common-currency and specialisation rooted accounts will need to be adapted to account for multivariate similarity analyses, and individual differences in such patterns, to move forward our understanding of the implementation of social information processing in the brain (Lockwood et al., 2020).

The link between prosocial motivation and respiratory interoception, that was only identified behaviourally, aligns with theoretical accounts that posit that interoceptive signals might drive people to behave more prosocially (De Waal, 2008; Piech et al., 2017; Singer et al., 2009; Rilling et al., 2008). Although other studies have shown the contrary (Lenggenhager et al., 2013), here, we show that people who have a higher propensity to be aware and use respiratory interoceptive signals are more incentivised and motivated to help others, especially when the reward to be received by others increases, than people who rely more on external signals. We used measurements that allow us to identify that increased sensitivity to rewards for others was specifically associated with reliance on respiratory interoceptive signal, rather than changes in one's own experience of reward or effort. Thus, we expand the previous research showing a link between altruism in economic games and cardiac interoception to respoiratory interoception (Piech et al., 2017), which may indicate that different interoceptive sources are integrated at a higher level leading to a consistent modulation of social behaviour.

Our respiratory task focuses on the extent to which people are aware of their internal signals, rather than accuracy at reporting their respiratory state without awareness, a process which has previously been linked to the role of AI in self-awareness and body ownership (Craig, 2009, 2002). Even though most of the research linking affective and social processes to interoception has focused on the cardiac dimension not respiratory, more abstract processing of both interoceptive pathways may be linked to awareness and they may indeed be correlated (Garfinkel, 2016). Furthermore, both cardiac and respiratory processing partially depend on the insular cortex, and both, together with other interoceptive dimensions, seem to be affected in psychiatric traits such as alexithymia and anxiety (Harrison et al., 2021; Murphy et al., 2018; Weng et al., 2021), which in turn have been associated with deficits in social cognition and behaviour (Shah et al., 2017; Gambin and Sharp, 2018; Robson et al., 2019; Bird et al., 2010; FeldmanHall et al., 2013). It is possible that the link between interoception and prosocial behaviour might therefore be mediated by psychiatric and empathic traits, as suggested by previous studies (Contreras-Huerta et al., 2020a). Indeed, some of these traits have been linked to the ACCg/AI neural responses to self and other rewards previously (Lockwood et al., 2015, 2022), and these may also be involved in the behavioural and neural links found here. Future research could shed light on how different interoceptive modalities interact with social behaviour, whether this interaction occurs across all different levels of complexity of interoception, and whether this interaction is mediated by personality traits.

5. Conclusion

In this study, we show that vicarious reward processing, respiratory interoception and prosocial behaviour are closely linked. People who show a greater propensity to be aware and use respiratory interoceptive signals are more incentivised when they can obtain rewards for others and act more prosocially. Neural representations of passively observed vicarious reward in the AI and ACCg were able to predict people's degree of reliance on internal signals and their levels of prosociality, respectively. These results suggest that social behaviour is complex and relies on both shared representations and self-other distinctions in the ACC and the AI, which might work together when facing a social situation. These findings highlight how important our everyday observations of others' rewards may be for our subsequent internal and social lives.

Declaration of Competing Interest

The authors declare no competing interest.

Credit authorship contribution statement

Luis Sebastian Contreras-Huerta: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Michel-Pierre Coll:** Conceptualization, Methodology, Investigation, Data curation, Writing – review & editing. **Geoffrey Bird:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Hongbo Yu:** Investigation, Writing – review & editing, Project administration. **Annayah Prosser:** Investigation, Writing – review & editing. **Patricia L. Lockwood:** Methodology, Writing – review & editing. **Jennifer Murphy:** Methodology, Writing – review & editing. **M.J. Crockett:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Matthew A.J. Apps:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision.

Data availability

All data and scripts used for main analysis and figures can be found here https://osf.io/bkmea/?view_only=b68a7511d330486885a99252729a8919

Acknowledgement

This work was supported by grants from the John Templeton Foundation (Beacons Project and No. 61495), the Academy of Medical Sciences (SBF001\1008), the Oxford University Press John Fell Fund, and the Wellcome Trust Institutional Strategic Support Fund (204826/Z/16/Z) awarded to MJC; a National Agency for Research and Development (ANID) DOCTORADO BECAS CHILE (BECAS CHILE/2016 – 72170287) to LSCH; a Biotechnology and Biological Sciences Research Council David Phillips Fellowship (BB/R010668/1; BB/R010668/2) and a Jacobs Foundation Research Fellowship to MAJA; and a Medical Research Council Fellowship (MR/P014097/1 and MR/P014097/2), a Jacobs Foundation Research Fellowship, and a Sir Henry Dale Fellowship funded by the Wellcome Trust and the Royal Society (223264/Z/21/Z) to PLL.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.119881.

References

Ruff, C.C., Fehr, E., 2014. The neurobiology of rewards and values in social decision making. Nat. Rev. Neurosci. 15, 549–562.
Lockwood, P.L., 2016. The anatomy of empathy: vicarious experience and disorders of social cognition. Behav. Brain Res. 311, 255–266.
Adolfi, F., et al., 2017. Convergence of interoception, emotion, and social cognition: a twofold fMRI meta-analysis and lesion approach. Cortex 88, 124–142.
Critchley, H.D., Garfinkel, S.N., 2017. Interoception and emotion. Curr. Opin. Psychol. 17, 7–14.
Grynberg, D., Pollatos, O., 2015. Perceiving one's body shapes empathy. Physiol. Behav. 140, 54–60.
Lischke, A., Weippert, M., Mau-Moeller, A., Jacksteit, R., Pahnke, R., 2020. Interoceptive accuracy is associated with emotional contagion in a valence- and sex-dependent manner. Soc. Neurosci. 15, 227–233.
Shah, P., Catmur, C., Bird, G., 2017. From heart to mind: linking interoception, emotion, and theory of mind. Cortex 93, 220–223.
Lockwood, P.L., et al., 2017. Prosocial apathy for helping others when effort is required. Nat. Hum. Behav. 1, 0131.
Rilling, J.K., et al., 2002. A neural basis for social cooperation. Neuron 35, 395–405.
Contreras-Huerta, L.S., Lockwood, P.L., Bird, G., Apps, M.A.J., Crockett, M.J., 2020a. Prosocial behavior is associated with transdiagnostic markers of affective sensitivity in multiple domains. Emotion doi:10.1037/emo0000813.
Contreras-Huerta, L.S., Pisauro, M.A., Apps, M.A.J., 2020b. Effort shapes social cognition and behaviour: a neuro-cognitive framework. Neurosci. Biobehav. Rev. 118, 426–439.
De Waal, F.B.M., 2008. Putting the altruism back into altruism: the evolution of empathy. Annu. Rev. Psychol. 59, 279–300.
Fukuda, H., et al., 2019. Computing social value conversion in the human brain. J. Neurosci. 39, 5153–5172.
Greening, S., et al., 2014. Individual differences in the anterior insula are associated with the likelihood of financially helping versus harming others. Cogn. Affect. Behav. Neurosci. 14, 266–277.
Harbaugh, W.T., Mayr, U., Burghart, D.R., 2007. Neural responses to taxation and voluntary giving reveal motives for charitable donations. Science 316, 1622–1625 *(80-.)*.
Piech, R.M., et al., 2017. People with higher interoceptive sensitivity are more altruistic, but improving interoception does not increase altruism. Sci. Rep. 7, 15652.
Terasawa, Y., Fukushima, H., Umeda, S., 2013. How does interoceptive awareness interact with the subjective experience of emotion? An fMRI Study. Hum. Brain Mapp. 34, 598–612.
Apps, M.A.J., Ramnani, N., 2014. The anterior cingulate gyrus signals the net value of others' rewards. J. Neurosci. 34, 6190–6200.
Monfardini, E., et al., 2013. Vicarious neural processing of outcomes during observational learning. PLoS ONE 8, e73879.
Morelli, S.A., Sacchet, M.D., Zaki, J., 2015. Common and distinct neural correlates of personal and vicarious reward: a quantitative meta-analysis. Neuroimage 112, 244–253.
Lockwood, P.L., Apps, M.A.J., Roiser, J.P., Viding, E., 2015. Encoding of vicarious reward prediction in anterior cingulate cortex and relationship with trait empathy. J. Neurosci. 35, 13720–13727.
Apps, M.A.J., Green, R., Ramnani, N., 2013a. Reinforcement learning signals in the anterior cingulate cortex code for others' false beliefs. Neuroimage 64, 1–9.
Fan, Y., Duncan, N.W., de Greck, M., Northoff, G., 2011. Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. Neurosci. Biobehav. Rev. 35, 903–911.
Lindquist, K.A., Satpute, A.B., Wager, T.D., Weber, J., Barrett, L.F., 2016. The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. Cereb. Cortex 26, 1910–1922.
Contreras-Huerta, L.S., Baker, K.S., Reynolds, K.J., Batalha, L., Cunnington, R., 2013. Racial bias in neural empathic responses to pain. PLoS ONE 8, e84001.
Hill, M.R., Boorman, E.D., Fried, I., 2016. Observational learning computations in neurons of the human anterior cingulate cortex. Nat. Commun. 7, 1–12.
Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., Rushworth, M.F.S., 2008. Associative learning of social value. Nature doi:10.1038/nature07538.
Apps, M.A.J., Rushworth, M.F.S., Chang, S.W.C., 2016. The anterior cingulate gyrus and social cognition: tracking the motivation of others. Neuron 90, 692–707.
Lamm, C., Decety, J., Singer, T., 2011. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. Neuroimage 54, 2492–2502.
Hadland, K.A., Rushworth, M.F.S., Gaffan, D., Passingham, R.E., 2003. The effect of cingulate lesions on social behaviour and emotion. Neuropsychologia 41, 919–931.
Rudebeck, P.H., Buckley, M.J., Walton, M.E., Rushworth, M.F.S., 2006. A role for the macaque anterior cingulate gyrus in social valuation. Science 313, 1310–1312 *(80-.)*.
Critchley, H.D., Wiens, S., Rotshtein, P., Öhman, A., Dolan, R.J., 2004. Neural systems supporting interoceptive awareness. Nat. Neurosci. 7, 189–195.
Critchley, H.D., Nagai, Y., 2012. How emotions are shaped by bodily states. Emot. Rev. 4, 163–168.
Craig, A.D., 2009. How do you feel - now? The anterior insula and human awareness. Nat. Rev. Neurosci. 10, 59–70.
Pollatos, O., Herbert, B.M., Mai, S., Kammer, T., 2016. Changes in interoceptive processes following brain stimulation. Philos. Trans. R. Soc. B Biol. Sci. 371, 20160016.
Harrison, O.K., et al., 2021. Interoception of breathing and its relationship with anxiety. Neuron 109, 4080–4093 e8.
Naqvi, N.H., Bechara, A., 2009. The hidden island of addiction: the insula. Trends Neurosci. 32, 56–67.
Paulus, M.P., 2007. Neural basis of reward and craving-a homeostatic point of view. Dialogues Clin. Neurosci. 9, 379–387.
Paulus, M.P., Stewart, J.L., 2014. Interoception and drug addiction. Neuropharmacology 76, 342–350.
Corradi-Dell'Acqua, C., Tusche, A., Vuilleumier, P., Singer, T., 2016. Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. Nat. Commun. 7, 1–12.
Craig, A.D., 2002. How do you feel? Interoception: the sense of the physiological condition of the body. Nat. Rev. Neurosci. 3, 655–666.
Dillon, D.G., et al., 2008. Dissociation of neural regions associated with anticipatory versus consummatory phases of incentive processing. Psychophysiology 45, 36–49.
Holroyd, C.B., Yeung, N., 2012. Motivation of extended behaviors by anterior cingulate cortex. Trends Cogn. Sci. 16, 122–128.
Rolls, E.T., 2016. Functions of the anterior insula in taste, autonomic, and related functions. Brain Cogn. 110, 4–19.
Williams, Z.M., Bush, G., Rauch, S.L., Cosgrove, G.R., Eskandar, E.N., 2004. Human anterior cingulate neurons and the integration of monetary reward with motor responses. Nat. Neurosci. 7, 1370–1375.
Mega, M.S., Cohenour, R.C., 1997. Akinetic mutism: disconnection of frontal-subcortical circuits. Neuropsychiatry Neuropsychol. Behav. Neurol. 10, 254–259.
Vogt, B., 2009. Cingulate Neurobiology and Disease. Oxford University Press.
Le Heron, C., Apps, M.A.J., Husain, M., 2018. The anatomy of apathy: a neurocognitive framework for amotivated behaviour. Neuropsychologia 118, 54–67.
Mesulam, M.M., Mufson, E.J., 1982. Insula of the old world monkey. II: afferent cortical input and comments on the claustrum. J. Comp. Neurol. 212, 23–37.
Nord, C.L., Garfinkel, S.N., 2022. Interoceptive pathways to understand and treat mental health conditions. Trends Cogn. Sci. 26, 499–513.
Critchley, H.D., 2005. Neural mechanisms of autonomic, affective, and cognitive integration. J. Comp. Neurol. 493, 154–166.
Lockwood, P.L., Apps, M.A.J., Chang, S.W.C., 2020. Is there a "social" brain? Implementations and algorithms. Trends Cogn. Sci. 24, 802–813.

Lockwood, P., et al., 2022. Distinct neural representations for prosocial and self-benefitting effort. Curr. Biol. 32, 4172–4185.
Saxe, R., Haushofer, J., 2008. For love or money: a common neural currency for social and monetary reward. Neuron 58, 164–165.
Lockwood, P.L., Apps, M.A.J., Valton, V., Viding, E., Roiser, J.P., 2016. Neurocomputational mechanisms of prosocial learning and links to Empathy. Proc. Natl. Acad. Sci. USA 113, 9763–9768.
Lenggenhager, B., Azevedo, R.T., Mancini, A., Aglioti, S.M., 2013. Listening to your heart and feeling yourself: effects of exposure to interoceptive signals during the ultimatum game. Exp. Brain Res. 230, 233–241.
Sallet, J., et al., 2007. Expectations, gains, and losses in the anterior cingulate cortex. Cogn. Affect. Behav. Neurosci. 7, 327–336.
Liu, X., Hairston, J., Schrier, M., Fan, J., 2011. Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. Neurosci. Biobehav. Rev. 35, 1219–1236.
Apps, M.A.J., Green, R., Ramnani, N., 2013b. Reinforcement learning signals in the anterior cingulate cortex code for others' false beliefs. Neuroimage 64.
Murphy, J., Catmur, C., Bird, G., 2018. Alexithymia is associated with a multidomain, multidimensional failure of interoception: evidence from novel tests. J. Exp. Psychol. Gen. 147, 398–408.
Lockwood, P.L., et al., 2021. Aging increases prosocial motivation for effort. Psychol. Sci. 32, 668–681.
Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 1–28.
Crockett, M.J., Kurth-Nelson, Z., Siegel, J.Z., Dayan, P., Dolan, R.J., 2014. Harm to others outweighs harm to self in moral decision making. Proc. Natl. Acad. Sci. 111, 17320–17325.
Drucker, D.M., Aguirre, G.K., 2009. Different spatial scales of shape similarity representation in lateral and ventral LOC. Cereb. Cortex 19, 2269–2280.
Aguirre, G.K., 2007. Continuous carry-over designs for fMRI. Neuroimage 35, 1480–1494.
Esteban, O., et al., 2019. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods 16, 111–116.
Garfinkel, S.N., et al., 2016. Interoceptive dimensions across cardiac and respiratory axes. Philos. Trans. R. Soc. B Biol. Sci. 371, 20160014.
Gorgolewski, K., et al., 2011. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. Front. Neuroinform. 5.
Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37, 90–101.
Rissman, J., Gazzaley, A., D'Esposito, M., 2004. Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage 23, 752–763.
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, 2636–2643.
Deen, B., Pitskel, N.B., Pelphrey, K.A., 2011. Three systems of insular functional connectivity identified with cluster analysis. Cereb. Cortex 21, 1498–1506.
Neubert, F.X., Mars, R.B., Sallet, J., Rushworth, M.F.S., 2015. Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. Proc. Natl. Acad. Sci. USA 112, E2695–E2704.
Vogt, B.A., Nimchinsky, E.A., Vogt, L.J., Hof, P.R., 1995. Human cingulate cortex: surface features, flat maps, and cytoarchitecture. J. Comp. Neurol. 359, 490–506.
Paus, T., et al., 1996. Human cingulate and paracingulate sulci: pattern, variability, asymmetry, and probabilistic map. Cereb. Cortex 6, 207–214.
Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans. Med. Imaging 23, 137–152.
Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29, 1165–1188.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate : a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57, 289–300.
Singer, T., Critchley, H.D., Preuschoff, K., 2009. A common role of insula in feelings, empathy and uncertainty. Trends Cogn. Sci. 13, 334–340.
Tsakiris, M., 2010. My body in the brain: a neurocognitive model of body-ownership. Neuropsychologia 48, 703–712.
Farb, N.A.S., Segal, Z.V., Anderson, A.K., 2013. Attentional modulation of primary interoceptive and exteroceptive cortices. Cereb. Cortex 23, 114–126.
Uddin, L.Q., Nomi, J.S., Hébert-Seropian, B., Ghaziri, J., Boucher, O., 2017. Structure and function of the human insula. J. Clin. Neurophysiol. 34, 300–306.
Uddin, L., Kinnison, J., Pessoa, L., Anderson, M.L., 2014. Beyond the tripartite cognition–emotion–interoception model of the human insular cortex. J. Cogn. Neurosci. 26, 16–27.
Nomi, J.S., Schettini, E., Broce, I., Dick, A.S., Uddin, L.Q., 2018. Structural connections of functionally defined human insular subdivisions. Cereb. Cortex 28, 3445–3456.
Nomi, J.S., et al., 2016. Dynamic functional network connectivity reveals unique and overlapping profiles of insula subdivisions. Hum. Brain Mapp. 37, 1770–1787.
Palmer, C.E., Tsakiris, M., 2018. Going at the heart of social cognition: is there a role for interoception in self-other distinction? Curr. Opin. Psychol. 24, 21–26.
Brethel-Haurwitz, K.M., et al., 2018. Extraordinary altruists exhibit enhanced self–other overlap in neural responses to distress. Psychol. Sci. 29, 1631–1641.
Crockett, M.J., Lockwood, P.L., 2018. Extraordinary altruism and transcending the self. Trends Cogn. Sci. 22, 1071–1073.
Hu, J., Hu, Y., Li, Y., Zhou, X., 2021. Computational and neurobiological substrates of cost-benefit integration in altruistic helping decision. J. Neurosci. 41, 3545–3561.
Rütgen, M., et al., 2015. Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. Proc. Natl. Acad. Sci. USA 112, E5638–E5646.
Rangel, A., Camerer, C., Montague, P.R., 2008. A framework for studying the neurobiology of value-based decision making. Nat. Rev. Neurosci. 9, 545–556.
Basile, B.M., Schafroth, J.L., Karaskiewicz, C.L., Chang, S.W.C., Murray, E.A., 2020. The anterior cingulate cortex is necessary for forming prosocial preferences from vicarious reinforcement in monkeys. PLoS Biol. 18, e3000677.
Chang, S.W.C., Gariepy, J.F., Platt, M.L., 2013. Neuronal reference frames for social decisions in primate frontal cortex. Nat. Neurosci. 16, 243–250.
Dal Monte, O., Chu, C.C.J., Fagan, N.A., Chang, S.W.C., 2020. Specialized medial prefrontal–amygdala coordination in other-regarding decision preference. Nat. Neurosci. 23, 565–574.
Wittmann, M.K., Lockwood, P.L., Rushworth, M.F.S., 2018. Neural mechanisms of social cognition in primates. Annu. Rev. Neurosci. 41, 99–118.
Balsters, J.H., et al., 2017. Disrupted prediction errors index social deficits in autism spectrum disorder. Brain 140, 235–246.
Ainley, V., Apps, M.A.J., Fotopoulou, A., Tsakiris, M., 2016. 'Bodily precision': a predictive coding account of individual differences in interoceptive accuracy. Philos. Trans. R. Soc. B Biol. Sci. 371, 20160003.
Kragel, P.A., Han, X., Kraynak, T.E., Gianaros, P.J., Wager, T.D., 2021. Functional MRI can be highly reliable, but it depends on what you measure: a commentary on Elliott et al. (2020). Psychol. Sci. 32, 622–626.
Rilling, J.K., et al., 2008. The neural correlates of the affective response to unreciprocated cooperation. Neuropsychologia 46, 1256–1266.
Weng, H.Y., et al., 2021. Interventions and Manipulations of Interoception. Trends Neurosci. 44, 52–62.
Gambin, M., Sharp, C., 2018. Relations between empathy and anxiety dimensions in inpatient adolescents. Anxiety Stress Coping 31, 447–458.
Robson, S.E., Repetto, L., Gountouna, V.E., Nicodemus, K.K., 2019. A review of neuroeconomic gameplay in psychiatric disorders. Mol. Psychiatry 1, 67–81.
Bird, G., et al., 2010. Empathic brain responses in insula are modulated by levels of alexithymia but not autism. Brain 133, 1515–1525.
FeldmanHall, O., Dalgleish, T., Mobbs, D., 2013. Alexithymia decreases altruism in real social decisions. Cortex 49, 899–904.