When and why do we value the lives of others?

Insights from brain, body and behaviour

Jo Cutler

Thesis submitted for the degree of

Doctor of Philosophy in Psychology

School of Psychology

The University of Sussex

December 2019

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree. However, the thesis incorporates to the extent indicated below, material already submitted as part of required coursework and/or for the degree of MSc in Cognitive Neuroscience which was awarded by the University of Sussex.

Jo Cutler

This thesis is presented in a 'papers style' format.

Chapter 1: Introduction and literature review introduces the topic, covers existing literature and outlines the studies in the thesis.

Chapters 2 - 6 present 5 empirical papers, each written in a format that is suitable for publication in peer-reviewed journals:

Paper 1: Cutler, J. & Campbell-Meiklejohn, D. (2019) **A comparative fMRI meta-analysis of altruistic and strategic decisions to give.** *Neuroimage, 184*, 227-241.

The version presented here is the accepted version, reformatted to match the rest of the thesis. An earlier version of this meta-analysis was included in coursework for the degree of MSc in Cognitive Neuroscience. The version presented here and published in Neuroimage includes additional data, a different approach to the analysis and a novel methodological development.

JC and DC designed the study. JC identified the eligible studies, obtained the data from authors, ran the analysis and wrote the paper, with support from DC.

Paper 2: Cutler, J., Radua, J. & Campbell-Meiklejohn, D. (2018) **Adjusting for variable brain coverage in voxel-based fMRI meta-analysis.** *Preprint bioRxiv*, doi.org/10.1101/457028.

JC and DC created the concept of the method which relies on the output of software developed by JR. JC wrote the code and the paper with contributions to both from DC and JR. All authors contributed to manuscript revision, read and approved the submitted version.

Paper 3: Cutler, J., Miles-Wilson, J. J. & Campbell-Meiklejohn, D. **Scope insensitivity and proximity bias affect physiological responses to lives at risk.** *(in prep)*

JC and DC designed the study and JC wrote the stimulus presentation code. JC and JM collected the data with help from two students. JC ran the preprocessing, analysis and wrote the paper, with input from DC on all these aspects and from JM on analysis.

Paper 4: Cutler, J. & Campbell-Meiklejohn, D. **The neural basis of valuing the lives of others: value depends on the number of people and geographical location** *(in prep)*

JC and DC designed the study. JC wrote the stimulus presentation code, collected the data and wrote the paper with help from a student. JC also analysed the data with support from DC.

Paper 5: Cutler, J. & Campbell-Meiklejohn, D. **Responsibility and valence determine interest in outcomes for others** *(in prep)*

JC and DC designed the studies. JC then: wrote the code and created the online survey; collected the data, with support from several students; analysed the data; and wrote the paper.

***Final chapter*:** Discussion of the overall findings from the empirical papers, including applications, limitations and future directions.

Acknowledgements

Thank you to my supervisor, Dr Daniel Campbell-Meiklejohn, for your confidence in me and excitement about our projects. Working with you has taught me so much and made me a better researcher.

The data that make up my thesis would not have been possible without a huge amount of time and energy from others. Thank you again to all the authors and technicians who spent time looking for data to contribute to our meta-analysis. A huge thank you to Sana, Guneet, Jolyon, Anika, Aaron, Paige, Rhiannon, Stevie and Georgia for helping me to collect the data and for the enjoyable experiences working with each of you. Thank you to Professor Jamie Ward and Dr Ryan Scott, whose equipment we borrowed to collect data. To the whole team at the Sussex Clinical Imaging Sciences Centre (CISC), thank you for all your support and hard work setting up and running the fMRI scans, it's been a pleasure to spend time with you during testing. My thanks also go to the participants.

For support with analysis I would like to thank the creators of AES:SDM, Dr Joaquim Radua and Dr Anton Albajes-Eizagirre, and creators of PsPM, particularly Professor Dominik Bach for the valuable advice. In general, I would like to acknowledge the amazing contribution of everyone developing new software or packages and sharing them with detailed instructions and support. I am also grateful to Federica who has supported me during recent fMRI data collection.

Next, I would like to thank Sussex Psychology for being such a welcoming and engaging department. A special mention to my assessor, Dr Matt Easterbrook, as well as Dr Charlotte Rae and Dr Sarah King for your advice, support, and feedback. Thanks also to our technicians, Dan and Martha, and all the other staff who keep everything working behind the scenes. I want to acknowledge the Economic and Social Research Council, who funded my research and additional opportunities during my PhD. I am also grateful to have received funding from CISC, the Sussex Kindness Award (supported by Kindness UK), Sussex Neuroscience and the Doctoral School. More generally, the Doctoral School, and Shelley Jenkins in particular, have supported me in many ways.

My PhD experience would not have nearly as fun without all the amazing people I have met. Thank you to Jolyon, Lina, James, Ola, and 'Ex-office Friends' Jes and Jenny for sharing highs and lows, Jukka for the 4-Js writing club, and everyone else who has shared advice or simply been a friendly face.

Summary

Altruistic behaviours benefit others at a cost to the self. They can be motivated by valuing the lives and wellbeing of other people. However, this value can be inconsistent, creating biases in who is valued and receives help. In this thesis, a literature review differentiates motivations for prosocial behaviour and then studies using neuroimaging, physiology, and behaviour explore the situational and individual factors that determine the value of other people.

Paper 1 uses fMRI meta-analysis to distinguish between altruistic decisions, driven by intrinsic value, and strategic prosocial decisions, which could be for extrinsic gain. Results demonstrate overlap in reward-related regions, activations unique to each context and differences between the two. Paper 2 presents a solution to an issue in fMRI meta-analysis.

The following papers focus on altruistic contexts. Papers 3 and 4 identify the physiological (skin conductance) and neural (fMRI) correlates respectively of valuing lives at risk. Results show that this value can be biased by the number of people at risk, whether they are at home or abroad, and how many similar situations one has already seen.

Papers 1, 3 and 4 suggest we are affected by events that happen to others and our ability to help them efficiently. Paper 5 tests whether this translates to curiosity and choices to find out about their fate or choices to avoid this knowledge. Results from a series of behavioural experiments show 'wanting to know' about an outcome depends on who it affects, who caused it, and whether it is positive or negative.

Contents

List of abbreviations

ACC: anterior cingulate cortex
AES:SDM: Anisotropic Effect Size Signed Differential Mapping
AI: anterior insula
AIC: Akaike information criterion
BOLD: blood oxygen level dependent
dlPFC: dorsolateral prefrontal cortex
dmPFC: dorsomedial prefrontal cortex
ERP: event related potential
fMRI: functional magnetic resonance imaging
FSL: FMRIB's Software Library
GLM: general linear model
IFG: inferior frontal gyrus
MNI: Montreal Neurological Institute
NuAcc: nucleus accumbens
OFC: orbitofrontal cortex
PCC: posterior cingulate cortex
PFC: prefrontal cortex
pSTS: posterior superior temporal sulcus
ROI: region(s) of interest
SCR: skin conductance response(s)
SFG: superior frontal gyrus
sgACC: subgenual anterior cingulate cortex
STS: superior temporal sulcus
tDCS: transcranial direct current stimulation
TPJ: temporoparietal junction
vlPFC: ventrolateral prefrontal cortex
vmPFC: ventromedial prefrontal cortex

General introduction

Human generosity through charitable donations and fundraising, actions for positive social change, and government aid programs save and improve the lives of millions of people every year (UNICEF, 2018). Choosing to be prosocial in the form of trust, cooperation or kindness also builds relationships, improves wellbeing and can increase overall resources (Crocker et al., 2017; Curry et al., 2018; Park et al., 2017; Tabibnia and Lieberman, 2007). Given these benefits, furthering the scientific understanding of prosocial decision making is important for social, economic, mental and physical health outcomes.

1.1. Definitions

In this thesis, prosocial behaviour is defined as any act that benefits others, at an initial cost to the self. A prosocial decision is the choice to behave prosocially. The cost of prosocial behaviour could be in time, money, effort, risk, or a combination of these. In addition to benefits for the recipient, prosocial decisions have benefits for the decider. These benefits can be extrinsic: external rewards, such as money gained through reciprocal generosity, or intrinsic: internal rewards, such as positive thoughts or emotions (Cutler and Campbell-Meiklejohn, 2019). Altruistic decisions are a subgroup of prosocial decisions, defined by a lack of potential extrinsic rewards. Anonymous charitable donations are one example of altruistic behaviours (Moll et al., 2006).

Research on decision making suggests the costs and benefits associated with choice options are integrated in value calculations (Croxson et al., 2009). Value can be generally defined as the importance, worth, or usefulness of something, and the process of estimating this. The idea of subjective value differentiates experienced desirability from objective worth (Peters and Büchel, 2010). For example, the value of food changes depending on how much one has already eaten.

In the context of prosocial decisions, choosing to be prosocial suggests the subjective value of this option is greater than the value of not being prosocial. In this thesis, I identify why being prosocial has value. These sources of value can also be conceptualised as motivations to be prosocial or benefits of being prosocial, and I use the terms interchangeably. I also assume that prosocial behaviours occurring in contexts that isolate a motivation are driven by that motivation. For example, an "altruistic behaviour" is preceded by an "altruistic decision" with "altruistic motivations". My research focuses on *when* and *why* the people who benefit from prosocial behaviour contribute value, so increase the motivation to be prosocial.

1.2. Measuring value

Approaches to measuring value include behavioural, physiological, and neuroimaging methods. In this section I briefly outline the benefits of each and how I have applied them to the research questions. The following literature review outlines existing research that used these methods to study prosocial motivations.

Measuring prosocial decisions directly can be informative, given the reasonable assumption that choices generally reflect the option with the highest value (Rangel et al., 2008). To separate different motivations, based on behaviour alone, requires experimental conditions that isolate certain benefits and exclude others. However, choices in experiments can be biased if participants act in a socially desirable way, which does not reflect true motivations (Fernandes and Randall, 1992). A focus on behaviour may also ignore the fact that similar prosocial behaviours can be motivated by a wide range of factors. The differences between motivations are crucial to understand, predict and encourage prosocial behaviours (Hein et al., 2016). Fortunately, other tools from psychology and neuroscience can offer additional insights.

One way to apply neuroimaging measures is in combination with traditional decision-making tasks. I apply this method in Paper 1 (Cutler and Campbell-Meiklejohn, 2019) by using functional magnetic resonance imaging (fMRI) meta-analysis to compare tasks that isolate altruistic motivations, relying on intrinsic rewards, with contexts where strategic motivations to gain extrinsic rewards could also be at play. Similarities and differences in the neural correlates of these decisions suggests potential overlap but also distinction between the value computations in these two contexts.

Physiological methods, such as skin conductance, and fMRI can also measure value during passive responses to stimuli, removing the need for a decision or behaviour. This has several benefits. It removes the risk of socially desirable responding and is more like real-world contexts in which people learn information without a requirement to make decisions. Papers 3 and 4 measure skin conductance responses (SCR) and blood oxygen level dependent (BOLD) responses, using fMRI, to news stories describing people at risk of death, and then the outcome. If participants value the lives of the people in the stories, there should be a response when lives are at risk or lost.

In addition to measuring the value of lives overall, experimental manipulations can help identify the situational factors that increase and decrease this value. By varying the details of the stories, I show that the value of a life depends on the number of people at risk

and whether they are in the same country as the participant or abroad. Differences between conditions contribute to answering the question of *when* other people are valued most.

Given the focus on subjective value, it is also important to consider that different individuals may weight components in the value calculations differently and value could be differentially affected by situational factors. If patterns of individual differences relate to measurable differences in traits, such as the tendency to empathise, there is evidence that trait contributes to prosocial motivations.

Finally, behavioural paradigms are not limited to measuring decisions whether to be prosocial. Previous evidence suggests that people want to know about outcomes with positive value (Charpentier et al., 2018). Finding out more about other people and opportunities to help them is an important part of prosocial behaviour in the real world. Paper 5 presents results from behavioural tasks measuring interest in outcomes that affect others. This task also allows us to manipulate responsibility for the outcome and measure whether the value of outcomes depends on being responsible. Results from this study are relevant to understanding both the situational factors that increase the value of others' wellbeing (*when*) and motivations to improve other's wellbeing (*why*).

2. Literature review

2.1. Overview

For a prosocial behaviour to be worthwhile, there must be benefits, so the subjective value of being prosocial outweighs the costs. The benefits of prosocial behaviours can take many forms, and these translate into different motivations to help. For example, if a subjective benefit of helping is feeling good from having an impact, helping is at least in part motivated by achieving that benefit.

In the first half of the literature review, I aim to answer the questions: what are the possible motivations, do they exist and how can they be isolated? To do so, I differentiate prosocial motivations based on 4 dichotomies. In reality, it is likely that multiple motivations work together, even within a single decision. However, separating different motivations is of academic interest and has implications for promoting prosocial behaviour.

First, I separate *altruistic* motivations, where only intrinsic rewards are available, from *strategic* motivations to gain extrinsic rewards through prosocial behaviours. The rest of the literature review and papers 3–5 focus on altruistic motivations and the nature of the intrinsic rewards. The first subdivision is between motivations linked to the recipient themselves (*recipient-dependent* motivations) and those not related to the specific recipient (*recipient-independent* motivations). Recipient-dependent motivations can also be labelled empathy, in the widest sense of the term.

The next distinction is within recipient-dependent motivations and considers whether empathy leads to *other-focused* emotions or *self-focused* personal distress. The section on this distinction considers how both sets of emotions can promote helping, but in different ways. Finally, I distinguish *action-oriented* from *outcome-oriented* altruism. These are both other-focused but differ in whether the benefit of helping is linked to the action of personally making a difference or the positive outcome of improved wellbeing for the other. The overall structure of the literature review is shown in table 1.

Table 1

Divisions and subdivisions between motivations

Strategic	Altruistic			
	Recipient-independent	Recipient-dependent		
		Self-focused	Other-focused	
			Action-oriented	Outcome-oriented

The benefits associated with each motivation could have both emotional and cognitive components. For example, action-oriented altruism could lead to thinking about the difference one made, feeling good for making that difference, or both. As mentioned below, the relative importance of these levels may differ between motivation types. At the point of decision, both emotions and thoughts could be anticipated in relation to how one will feel or think after the prosocial behaviour (Erlandsson et al., 2016).

In the second part of the literature review, I cover definitions and previous findings on the situational factors and individual differences relevant to prosocial behaviour, with a focus on those used in the studies in the thesis. Combining the results from the studies with existing literature provides insight into how altruistic motivations are affected by context, how they differ between individuals and what these findings tell us about motivations.

2.2. Scope

Making a prosocial decision involves comparing the value of outcomes for others with outcomes for the self, but also comparing outcomes that vary in currency, time, abstractness, and certainty (Cutler and Campbell-Meiklejohn, 2019). Each of these factors alone relies on complex cognitive abilities such as theory of mind, to understand the mental states of other people (Frith and Frith, 2006), or thinking about the future (Schacter et al., 2017). In addition, the decision-making process must integrate these disparate signals into a common value to choose the most desirable option. Understanding all these factors presents a fascinating research challenge but is outside the scope of this thesis.

The focus here is on the benefits of being prosocial and the costs of being selfish, as these define the different categories of motivation I will outline below. The costs of being prosocial are the money, time or effort needed to achieve the prosocial outcome and the benefits of being selfish are keeping these resources. These contributions to the calculation

are not discussed in detail, although they relate to some of the motivations, for example the same prosocial benefit at a lower financial cost increases how efficient the prosocial option is.

I also focus the review on prosocial behaviour towards another person or human causes, as this is the most common operationalisation and most relevant for the studies in the thesis. Similar motivations and additional unique factors are likely to be relevant for prosocial acts towards animals and the environment (Bastian et al., 2011; Zaval et al., 2015). The focus is also adult populations and mostly healthy subjects, although some patient studies are mentioned. The developmental trajectory of prosocial behaviours, motivations and their neural basis are fascinating areas of study (Chernyak and Kushnir, 2018; Cowell et al., 2018; Knafo-Noam et al., 2018; Martin, Alia; Olson, 2015) and are only excluded for brevity.

2.3. Motivations for prosocial behaviour

2.3.1. Strategic

Prosocial behaviour could be considered a challenge for traditional theories in economics that suggest humans or homo economicus are entirely self-interested (Adamus, 2017). A related question is how such prosocial behaviours evolved if they require sacrifices that could decrease fitness. One answer to these questions is that prosocial behaviours can lead to extrinsic benefits for the person being prosocial, as well as the recipient. A possible motivation for prosocial behaviour is therefore to strategically gain such rewards. The value of other people in these contexts depends on their role in determining the extrinsic outcomes.

Prosocial acts may increase the gains of both people through reciprocity (Falk and Fischbacher, 2006; Fehr et al., 2002) and cooperation (Fehr and Fischbacher, 2004) or the resources of a broader group through public goods contributions (Chaudhuri, 2011). Likely to maintain these benefits, prosocial behaviour often has a positive effect on reputation, which promotes further beneficial interactions (Milinski et al., 2001; Nowak and Sigmund, 2005) and acting selfishly can lead to punishment (Fehr and Gächter, 2002).

Behavioural evidence

A variety of economic games and experimental tasks have been developed to study prosocial behaviour, when these strategic benefits are available, and findings support their role in motivating prosocial behaviour. The ultimatum game (Güth et al., 1982) involves a proposer who offers a split of money to a responder who can either accept or reject the

proposal. This offers the opportunity to punish selfish proposals, as rejections leave both players with nothing. In the trust game (Berg et al., 1995), the investor chooses an amount of money to entrust and this is multiplied and transferred to the trustee who decides how much to send back, possibly reciprocating the investor's generosity.

The public goods game (Rapoport, 1987) asks participants to put an amount of their money into a group pot which is multiplied and then divided between all players, regardless of contribution. The greatest total benefit is gained through prosociality but each individual gains the most by "free riding". Similarly, in the two person prisoners' dilemma (Luce and Raiffa, 1957) an individual player is better off if they defect, while their partner cooperates, but the total gain is enhanced by mutual cooperation. A common manipulation in these games is to have repeated rounds, introducing an incentive to build a positive reputation (King-Casas et al., 2005).

Importantly, prosocial behaviours in these games could be entirely motivated by the altruistic desire to help others, as discussed below, but it is impossible to rule out strategic motivations. However, manipulations to increase the size or likelihood of the extrinsic rewards increase levels of prosocial behaviour, supporting the role of strategic motivations in prosocial behaviour (Camerer, 2003; Camerer and Fehr, 2003).

Neural basis

Research using neuroimaging during prosocial decisions in these strategic contexts suggests their neural basis and a meta-analysis of these findings are presented in paper 1. Overall, these studies show involvement of regions associated with reward processing and social cognition, as would be expected in such tasks. The extrinsic rewards from strategic prosocial decisions are often dependent on the decisions of other people, for example whether they will cooperate or freeride. Neural regions linked to theory of mind are suggested to underlie this processing (Weiland et al., 2012).

Considering the reward aspect, social status, prosocial outcomes, cooperation, and money for the participant all activate the striatum (Izuma et al., 2008; Rilling et al., 2002; Saxe and Haushofer, 2008; Zink et al., 2004). This activity has been suggested to represent domain-general reward processing, across social and non-social sources, supporting the idea of a common neural currency for the benefits of a decision option. In strategic prosocial decisions, social and non-social rewards could combine if both intrinsic and extrinsic motivations apply. An interesting question is how these motivations are integrated in the brain, given behavioural findings that extrinsic rewards crowd out intrinsic ones (Frey and Oberholzer-Gee, 1997).

2.3.2. Altruistic

The possible extrinsic benefits to being prosocial described above explain how these behaviours can be motivated by strategic incentives to benefit oneself. However, these rely on a mechanism for the prosocial act to benefit the decider through reciprocity, reputation, or being part of the group that the benefits. In the lab, this is introduced through the structure of the game and in the real world, often the people we help are people we know, including friends, family and colleagues. This creates opportunities to "return the favour" or cooperate.

However, people are also prosocial in contexts where this is not the case. I will describe these choices, to be generous when there is no opportunity to gain extrinsic rewards, as altruistic. It is important to recognise that this term has been used with multiple meanings across multiple disciplines. The current definition reflects behavioural and preference altruism, rather than reproductive altruism, which requires a decrease in the actor's fitness in an evolutionary sense (Clavien and Chapuisat, 2012). If people are altruistic, this suggests they value other people in some way.

Behavioural evidence

Private donations to charities, working on causes that don't affect the donor, are examples of prosocial behaviour that strategic motivation cannot explain. In the UK in 2018, donations to overseas aid and disaster relief charities were the fifth most popular cause area and the second highest average donation amount. Average donations were £30 in the 4 weeks preceding the survey (Charities Aid Foundation, 2019).

In the lab environment, these contexts have mainly been created through the dictator game (Kahneman et al., 1986) in which the participant splits an amount of money with another person and there is no opportunity for reciprocity or punishment. If motivation was purely based on extrinsic reward, dictators would keep the full amount. Despite this, a meta-analysis of 616 dictator game treatments from 131 papers showed that on average, dictators gave 28.35% of the total. Of the 328 treatments with full range information available, 16.74% of dictators gave half of the total and 5.44% gave away everything (Engel, 2011).

These findings suggest that there is an intrinsic value to giving money away, which provides altruistic motivations for prosocial behaviour. I will refer to the benefit gained from acting altruistically as warm glow. This is distinct from what I will call action-oriented giving, described below, which aligns with the initial definition of warm glow (Andreoni,

1990, 1989). In line with more recent uses of the term (Erlandsson et al., 2016; O'Brien and Kassirer, 2019; Västfjäll et al., 2015), I use warm glow for any intrinsic reward from giving, regardless of its nature, and separate the different possible sources of reward below.

Evidence that people experience warm glow comes from findings that being altruistic enhances wellbeing (Dunn et al., 2008) with a recent meta-analysis suggesting a small to medium effect of being prosocial on happiness and no evidence of publication bias (Curry et al., 2018). Prosocial behaviours have also been linked to improved physical health outcomes (Crocker et al., 2017). These studies did not necessarily always isolate altruistic motivations, as participants could have been generous to people who could reciprocate, but evidence from two studies on volunteering found only those who help for other-focused reasons showed the benefits of decreased mortality (Konrath et al., 2012) and improved relationships (Crocker and Canevello, 2008).

Intrinsic rewards have also been shown to enhance prosocial performance and persistence (Grant, 2008) and including an intrinsic reward function in evolutionary modelling of prosocial behaviour improved the model (Wang et al., 2018). Interestingly, recent evidence suggests that people adapt to warm glow of giving more slowly than the good feeling of getting money (O'Brien and Kassirer, 2019). These studies all suggest that altruistic behaviours generate intrinsic rewards or warm glow. At the point of deciding whether to help, it is likely that people imagine helping to some extent (Gaesser et al., 2015) so anticipate warm glow if they decide to help. This introduces a mechanism by which the intrinsically rewarding outcome of being altruistic can motivate the preceding decision to be altruistic.

In contrast to warm glow as a result of helping, the emotional costs of not helping are guilt, shame and embarrassment. Anticipation of these negative emotions is likely to be integrated during decision-making in a similar way to anticipated warm glow. Research has often combined measures of avoiding negative emotion with seeking positive emotions. However, Erlandsson et al. (2016) suggest that seeking warm glow and avoiding guilt are separate, based on how each differentially related to personal responsibility.

Neural basis

Previous evidence from fMRI on the neural basis of altruistic decisions involving money is highlighted in paper 1, which provides results from a meta-analysis of these findings. These include activity in reward-related regions such as the nucleus accumbens and also the subgenual anterior cingulate cortex (sgACC) (Harbaugh et al., 2007; Moll et al., 2006).

In addition to fMRI research, evidence comes from methods that can suggest a causal relationship between neural structures and altruistic decisions. For example, a lesion study showed that altruistic donations were lower in patients with damage to the right posterior superior temporal sulcus (pSTS) and middle temporal gyri (Moll et al., 2018). The authors link the pSTS finding to diminished emotional empathy, which also follows damage to this area (Driscoll et al., 2012).

While most research looks at financial donations to others, the "pain versus gain" paradigm (FeldmanHall et al., 2012b) involves decisions whether to give up money to prevent the pain of another person. Evidence from combining this paradigm with transcranial direct current stimulation (tDCS) to the medial PFC (mPFC) supports our meta-analysis findings that this region is involved in altruistic decisions (Liao et al., 2018). Anodal stimulation, associated with enhanced activity, to mPFC increased generosity compared to cathodal stimulation, specific to conditions where there was a high probability of their help being successful.

The results and discussion of paper 1 focus on the rewards from giving. We do not discuss the costs of the rejected option, such as anticipated guilt. Research focusing on guilt finds activity in dorsomedial PFC (dmPFC), superior frontal gyrus (SFG), supramarginal gyrus, and anterior inferior frontal gyrus varies with the intensity of the guilt. Further, greater guilt was experienced when consequences were negative for others compared to the self. This contrast was associated with increased activity in the ventromedial PFC (vmPFC), dmPFC, precuneus, posterior cingulate and pSTS. The left anterior inferior frontal gyrus, left vmPFC and left anterior inferior parietal cortex showed an interaction between consequences and intensity (Morey et al., 2012). Interestingly, we found some of these regions also responded to lives being at risk or lost in paper 4.

Somewhat overlapping, and in other ways contradictory, results are given in a review of the neural correlates of guilt, shame and embarrassment (Bastin et al., 2016). Guilt, considered the most other-focused of the three emotions, was associated with activity in the ventral anterior cingulate cortex, areas of posterior temporal cortex and the precuneus. These regions also come up in contrasts in paper 4.

2.4. Motivations for altruism

The findings outlined above suggest altruistic motivations can drive behaviour but do not explain why such motivations exist. In the next section, I explore different motivations for altruism or the ways in which other people are valuable. I will argue that when considering the processes underlying motivations, it is useful to categorise

motivations based on the nature of the input to the value calculation. This is in line with the neural model of prosocial decisions suggested by Luo (2018) that starts with the signal input. This proceeds to influence the reward system and arouse emotions that prompt the planning of behaviour.

A key distinction in the valuation of other people is whether it depends on the recipient(s) or is independent of the person or people themselves. Next, I cover examples of *recipient-independent* and *recipient-dependent* information and evidence for their role in motivating altruism. Any of these factors could also motivate prosocial decisions in strategic contexts, instead of or in addition to the desire to gain extrinsic rewards.

2.4.1. Recipient-independent

Behavioural evidence

One factor suggested to motivate altruistic decisions that does not depend on considering the recipient's wellbeing is a preference for equity or equality (Bolton and Ockenfels, 1995). This factor could explain the finding that giving half of the total to the recipient is the most common non-zero offer in the dictator game (Engel, 2011). Another preference that could influence prosocial decisions without consideration of the recipient is for efficiency or creating the greatest impact. In dictator games, this can be established through multiplying the amount given to the recipient, a manipulation that does increase the amount given and the likelihood of giving something (Engel, 2011). There is some evidence that both efficiency and equality are integrated into judgements depending on their magnitude (Palmer et al., 2013a).

Moral rules are another way in which people may decide to be altruistic without processing information about the specific recipient. This could be the case if an option is interpreted as the "right thing to do" from cues independent of thinking about the recipient. The power of this was demonstrated in two studies where participants chose the option labelled as "nice", regardless of whether it was the most efficient or equitable option and generous people showed this effect most strongly (Capraro and Rand, 2017; Tappin and Capraro, 2018). Similarly, framing dictator game choices through moral words that suggest something is right ("donating") or wrong ("stealing") increases or decreases giving respectively (Capraro and Vanzo, 2019).

It could also be possible for learning processes to remove the need to consider the recipient before making an altruistic decision, if this has been rewarded in the past. This could have been through any of the types of reward discussed in this chapter leading to

habitual behaviours (Gęsiarz and Crockett, 2015) or learning through observation (Penner et al., 2005).

Recipient-independent motivations to be altruistic could come from expectations about what most people would do or perceived social norms. For example, participants' belief about the average behaviour of others, so what is appropriate, has been linked to choices in the dictator game (Matthey and Regner, 2014). Similarly, there is evidence people are more altruistic after observing others being generous (Nook et al., 2016). Finally, there is evidence that signalling prosociality to others is important in determining levels of altruism (Grossman, 2015).

Neural basis

Previous studies have also investigated the neural correlates of making altruistic decisions motivated by these factors. Considering equality, reward-related responses have been linked to equitable outcomes (Zaki and Mitchell, 2011) and paying to achieve equality (Dawes et al., 2012), whereas insula activity has been linked to inequitable outcomes and preferences against inequality.

Rule following during prosocial decisions was affected by tDCS to the right lateral PFC, independent of what the more prosocial outcome was (Gross et al., 2018). Other research has implicated the vmPFC, particularly on the right, as a key hub in the moral decision-making network for prosocial behaviours (Mendez, 2009). By no means is morality limited to recipient-independent factors and instead often includes information about the recipient or person in need. Moral rules simply provide one way in which decisions can be made without thinking about the recipient(s). The neuroscience of morality more generally is a huge field spanning moral judgements and inferences as well as decisions (Yu et al., 2019) and will not be covered in detail here (for reviews see Carlson and Crockett, 2018; Decety and Yoder, 2017; Van Bavel et al., 2015).

Considering learning, habitual prosocial behaviour could develop through the same processes as non-social learning, reward prediction-error signals in the striatum (Gęsiarz and Crockett, 2015). Evidence for this possibility has been shown for altruistic decisions in nucleus accumbens (NuAcc) (Kuss et al., 2013) and non-costly learning for others in the sgACC (Lockwood et al., 2015).

Finally, considering the expectations of others is likely to involve some similar processes and neural regions to those suggested below to underlie recipient-dependent motivations, such as theory of mind, but crucially in this case, these are applied to people

other than the recipient. Considering norms and expectations is also relevant to strategic prosocial contexts (Chang and Sanfey, 2013), but are categorised under altruistic motivations when there is no way that a prosocial reputation can lead to extrinsic rewards.

2.4.2. Recipient-dependent

The types of recipient-independent factors described above can motivate prosocial or altruistic decisions without any attention to or processing of information about the specific person(s) who will benefit from the decision. In contrast, perhaps the most common, long standing ideas in psychology and outside academia on why people are prosocial revolve around positive outcomes for the recipient(s). This relies on processing, at least to some extent, information about the person or people. I will now expand on the types of information and then consequences of this.

Empathy is the general term given to understanding others and is used with a variety of meanings. A common distinction is between a cognitive component, or theory of mind, and an emotional component (Shamay-Tsoory et al., 2009). Within the emotional component, there are further subdivisions depending on the relationship between the emotions felt by each person. These include: imagining how the other feels; imagining how one would feel in the other's situation; matching the emotion (sometimes described as automatic emotional contagion); feeling a complimentary emotion, such as pity if someone else is sad (often called sympathy); and feeling a positive desire to help in response to sadness (often called compassion) (Batson, 2009). A full review of the previous behavioural and neural work on empathy is outside the scope of this chapter (for reviews see Bernhardt and Singer, 2012; Decety, 2011; Engen and Singer, 2013; Lockwood, 2016; Marsh, 2018; Shamay-Tsoory, 2011). Here, I highlight some select findings on empathy in relation to prosocial behaviour.

Behavioural evidence

A concise and influential theory on the role of empathy in altruistic behaviour is simply that empathy motivates altruism (Batson et al., 1987). Support for this "empathy-altruism" hypothesis has often come from negative situations, where a helper's empathy motivates them to relieve the suffering of another, but can also apply in positive situations, where empathy generates vicarious joy during actions to make others better off (Batson et al., 1991). This has been suggested to create a positive loop, with altruism maintaining positive emotions in both the giver and the receiver (Telle and Pfister, 2015). In addition to sadness and joy, empathic anger has also been linked to helping behaviours, using state and

trait measures, and these effects were independent of the traditional measures of empathic sadness (Vitaglione and Barnett, 2003).

Evidence supporting empathy as a motivation for altruism comes from a range of study designs, including individual differences, and empathy is often used to explain the effect of situational factors on altruism. These studies will be described in the relevant sections below. Studies also use empathy manipulations, with one example showing increased altruistic behaviour in the dictator game, predicted by experienced empathy (Klimecki et al., 2016). However, there is some evidence that natural levels of empathy are in fact quite high and such manipulation methods show differences because of the control instructions, to stay objective, rather than the empathy instructions increasing caring above natural levels (McAuliffe et al., 2018). Empathy has also been supported as a good explanation from an evolutionary point of view for helping behaviour in situations where strategic motivations are not relevant (de Waal, 2008).

Neural basis

Neuroscience research on empathy more broadly has linked theory of mind to regions including the temporoparietal junction (TPJ), dmPFC, and pSTS and emotional empathy to the anterior cingulate cortex (ACC) and anterior insula (AI; Kanske, 2018). A general concept from this research is self-other overlap: similar regions are involved in processing the emotions of others and one's own emotions (Decety, 2015). Associations between activity in these regions and altruism suggest the neural basis of integrating information about the recipient into value calculations and support the importance of empathy in motivating altruistic decisions.

Results from fMRI on emotional empathy found positive functional connectivity from the ACC to the AI during altruistic decisions motivated by empathy (Hein et al., 2016). Observing people experiencing negative social experiences has been linked with activity in the theory of mind network, including dmPFC and, for empathic people, AI and dorsal ACC. Activity in AI and dmPFC predicted later prosocial behaviour (Masten et al., 2011). Similarly, ACC activity during concern for someone being socially isolated was attenuated through helping them (Kawamichi et al., 2015), though this was not costly helping.

Regions linked to theory of mind have also been implicated in altruism with suggestions that processing information about others, associated with dmPFC activity, and supressing information about the self is crucial (Majdandžić et al., 2016). There is also evidence of theory of mind and empathy regions increasing connectivity with reward regions in vmPFC during charitable donations (Hare et al., 2010) and the striatum during

donations to family members (Telzer et al., 2011). This suggests a mechanism by which the warm glow of altruism can be enhanced through processing information about the recipient.

Another region linking empathy and prosocial behaviours is the septal region where activity was shown to be modulated by taking a compassionate attitude when viewing sad expressions (Kim et al., 2009). This region also showed activity across empathy for pain, happiness and anxiety with the levels of this activity predicting daily helping, including to strangers (Morelli et al., 2014). Finally, the level of synchronisation between two participants' brains during a coordination task predicted levels of prosociality towards each other (Y. Hu et al., 2017). This finding supports the idea that neural mirroring promotes prosocial behaviour, although the measures were hypothetical.

2.5. Consequences of empathy

The translation of information about potential recipients into altruistic behaviour could be conceptualised as a multi-stage process, with aspects of empathy underlying the initial stages. Considering opportunities to relieve suffering, forms of empathy that generate awareness of suffering are likely to be an important initial step towards altruism. This may be necessary in some situations (although need could be detected without empathy, for example, direct requests for help) but not sufficient for altruism. There are two broad possible consequences of empathic understanding. These consequences may both lead to altruism but differ in whether the focus is on the recipient (*other*) or on the *self*.

2.5.1.   Other focus

The evidence described in the previous section suggests that empathy can lead to altruism and one possible route is through integrating information about others into value computations. The terms empathic concern and compassion describe the awareness of suffering evoking an other-focused motivation to help. Different orientations within this process are discussed below. Supporting the idea that empathy is not sufficient for prosocial behaviour, a study found concern about the welfare of others, not feeling their emotions, predicted prosocial behaviours (Jordan et al., 2016). Similarly, a there is evidence a concern mechanism motivates altruistic behaviour and requires minimal theory of mind (Nichols, 2001).

2.5.2.   Self focus

Despite evidence that empathy promotes altruism, understanding the suffering of others does not necessarily mean helping is motivated by other-focused concerns. Since

initial suggestions of the empathy-altruism hypothesis, it has been contrasted with a personal distress model that suggests people selfishly help to relieve their own distress (Batson et al., 1991, 1987). If altruism is entirely motivated by relieving personal distress, people will only help if it is the easiest way to reduce their distress. If escaping the situation is possible, this option will be taken instead of helping.

Based on evidence from several studies, Batson (2014) argues that when empathy is high, people do not take even easy opportunities to escape, against the idea of personal distress. However, other evidence shows people avoid opportunities to donate (Andreoni et al., 2017; Knutsson et al., 2013), will sacrifice money to avoid a dictator game decision (Dana et al., 2006), and even avoid empathy itself (Cameron et al., 2019).

Stress can have a positive effect on prosocial behaviour in some contexts (Buchanan and Preston, 2014). However, if self-focused negative emotions become too intense, and escape or avoidance is not possible, altruism may actually decline. For example, selfish behaviour in a virtual reality emergency was associated with neural signatures of personal distress (Zanon et al., 2014). Increased personal distress has also been proposed as a mechanism behind the bystander effect, where help is less likely when there are many potential helpers (Hortensius and de Gelder, 2018). Chronic experience of distress in response to the suffering of others can lead to burnout which undermines the capacity to help (Klimecki and Singer, 2012).

Given findings that similar brain regions process experienced and empathic emotions, the distinction between other-focused empathy and personal distress is less clear than early behavioural models suggested. It is likely that similar arousal processes underlie both emotions and the difference is how this is interpreted (Penner et al., 2005).

In real world altruism, often both motivations will apply and could relate to behaviours in different ways. For example, motivation to maintain one's own positive mood was linked to decisions whether to donate, whereas empathy was more predictive of the amount donated (Dickert et al., 2011b). A patient lesion study also found altruistic motivation depends on inhibiting one's own emotional experience to successfully generate empathic responses (Shdo et al., 2018). These findings emphasise the importance of balancing self and other focus.

2.6. Other-focused orientations

Evidence reviewed so far, and findings in this thesis, support the existence of altruistic motivations, not dependent on extrinsic reward and initiated by information

about the recipient, which promote helping, rather than escaping the situation. The final aspect of my classification is whether this desire is oriented towards the *outcome* for the recipient or the *action* of helping itself. These motivations were originally labelled pure and impure altruism respectively, with impure altruism also referred to as warm glow giving (Andreoni, 1990). However, the term warm glow has increasingly been used more broadly, as I have used it, to refer to any good feeling from giving, without excluding outcome-oriented motivations.

2.6.1. *Outcome orientation*

Outcome-oriented altruism is motivated by the benefit for the other person, separate from the act of being the one to help. This motivation has proven difficult to isolate in behaviour alone. One prediction of outcome-oriented motivations is that donations by oneself and donations by others are perfect substitutes for each other, if they create the same impact for the recipient(s). This means that if someone else contributes to a cause, an outcome-oriented donor should decrease their contribution by that amount, known as complete crowding out (Ottoni-Wilhelm et al., 2017).

Behavioural evidence

Experimental tests of crowding out have generally not shown support for outcome-orientation (Andreoni, 1990). However, an issue with this paradigm is that anything less than complete crowding out is considered evidence against outcome orientation, so the power to detect it is low. In a modified paradigm that measured preferences at multiple levels of charitable output, outcome orientations were shown to be more influential than previously estimated, although their contribution depended on the context (Ottoni-Wilhelm et al., 2017). Another criticism is that predictions of complete crowding out are specific to contexts where additional donations will not increase the size or strength of the positive outcome, and these are very unlikely in the real world.

Neural basis

Work on observing others gain, not because of altruism or prosocial behaviour, has identified the neural mechanisms of this vicarious reward (Mobbs et al., 2009). Like the overlap between experiencing and observing negative emotions, seeing others receive money activated areas including the ventral striatum, also active when receiving money. This vicarious reward was powerful enough to drive learning, although learning rates were slower than for the self (Lockwood et al., 2016). This study also found a specific reward-prediction error signal for others in sgACC.

A similar study measured activity in ventral striatum during gains for the participant, a charity, and both the participant and a charity. Interestingly, the group results did not show increased activity during gains for the charity only. However, the extent to which this was the case correlated with participant's self-reported empathic concern and how much they enjoyed winning for the charity. Empathic concern was also correlated with the amount donated to charity (Spaans et al., 2018).

Finally, an event-related potential (ERP) study found greater amplitude and duration of responses to failed, compared to successful, attempts to help hypothetical others, suggesting a successful outcome was valued (Gan et al., 2016). While these results combined suggest that positive outcomes for others can be rewarding, they do not consider costly prosocial decisions to achieve these positive outcomes. As the only studies that do this contrast outcome with action-oriented motivations, I outline them below.

2.6.2. *Action orientation*

Action-oriented altruism is motivated by the reward of personally having a positive impact on another person or cause. This idea was introduced to explain the lack of complete crowding out found in behavioural economics experiments and anecdotal patterns of behaviour (Andreoni, 1990, 1989). In the present classification, I am excluding prosocial decisions motivated by feeling like a good person from doing the "right" thing if this comes from cues independent of the recipient.

Behavioural evidence

As described above, one line of evidence for the presence of action-orientation is the lack of complete crowding out, although this measure has limitations and is not very relevant to real-world giving. A similar but more applicable pattern of behaviour is if donors fail to consider the marginal benefit of their gift. For example, donating to an appeal that has already raised enough to meet its aims, rather than one that has not yet reached its target.

An additional prediction to differentiate orientations is that action-oriented deciders will help in the way that makes their impact feel greatest, even if this diverges from creating the best outcome for the recipients. Anecdotal and scientific evidence support the idea that many altruistic behaviours are not those that would bring about the greatest possible impact.

In the charitable giving domain for example, most donors do not choose causes or organisations based on the efficiency of their impact (Van Iwaarden et al., 2009). It is common to choose organisations based on the percentage they spend on projects (Caviola

et al., 2014) and a lack of overhead costs promotes giving (Gneezy et al., 2014), even though this measure is often meaningless for comparing efficiency (Bowman, 2006). Similarly, despite evidence of the positive impact of directly sending cash to people living in poverty (Haushofer and Shapiro, 2016), most charities request donations for explicit items such as blankets or food, presumably as this is more effective in soliciting gifts. This could represent action-oriented giving if spending money on projects or set outcomes makes the donor feel better about their impact, at the expense of the objective impact for the beneficiary.

Many other situational factors, which I describe below and are tested in this thesis, also lead to biases in altruistic behaviour as they diverge from the common view that all lives have equal value (Slovic et al., 2011). These biases include preferring to help certain recipients over others (Everett et al., 2015) and the nonlinear relationships between the number of people in need and responses (Dickert et al., 2012). If decisions on whether and who to help are made based on our subjective experience of helping, rather than the objective change in wellbeing, there is evidence for action-oriented altruism. Finally, outside the lab, models of big data also support a role of donor joy in giving (Ribar and Wilhelm, 2002).

Neural basis

Identifying outcome and action-oriented motivations for altruism is one area where neuroscience can add insight beyond behavioural studies. One study gave participants voluntary choices whether to donate to their favourite charities and took compulsory 'tax-like' transfers to the same charities. There was greater striatal activity during the tax transfers, compared to baseline, but this was further increased by being responsible for the donation (Harbaugh et al., 2007). The findings suggest both the positive outcome and the process of deciding to give are reinforcing and may sum. Also, the participants who showed neural reward signatures during involuntary transfers were the most generous in real life.

Similar findings, that action orientation dominates for less altruistic individuals but outcome orientation is present for altruists, were shown using a reward prediction error paradigm with fMRI (Kuss et al., 2013). Participants decided whether to give and then some donations were cancelled before they reached the beneficiaries. The presence of action-focused motivation was suggested when a cancelled donation did not generate a reward prediction error. More generous participants seemed to be affected by whether the positive outcome of the donation was realised. They showed a reward prediction error in the nucleus accumbens when their donation was cancelled, just like when a payment for them was cancelled.

2.7. Fit with other models

In the above literature review I have differentiated prosocial motivations on several factors to create categories. At the highest level, prosocial behaviours can be driven by *strategic* motivations for extrinsic rewards or *altruistic* motivations for intrinsic rewards. Within altruistic behaviours, those motivated by *recipient-independent* factors are differentiated from those motivated by *recipient-dependent* information about the person or people being helped. Processing *recipient-dependent* information, through the various forms of empathy, can lead to *other-focused* emotions (empathic concern and compassion) or *self-focused* personal distress. Personal distress can promote altruism, but only if this is the easiest way to relieve the distress. Finally, other-focused motivations can be oriented towards the *outcome* for the recipient(s) or the *action* of helping. Studies measuring behaviour and neural activity support the presence of each of these motivations. The following section will consider the situational and individual differences that determine which is dominant in a given context for a specific individual.

While the combination of categories as a whole is novel, it incorporates previously established distinctions and builds on other models of prosocial motivations. For example, a review of 500 articles on charitable giving identified eight drivers (Bekkers and Wiepking, 2010), most of which can be matched to the categories I have described (in parentheses). These were: reputation (strategic); material costs and benefits (strategic); psychological benefits (altruistic); values and efficacy (recipient-independent); awareness of need (recipient-dependent); and "altruism" (other-focused). The final driver, solicitation, could elicit any of the motivations I have identified, including recipient-independent motivations if decisions are made without considering the potential recipient(s).

Other reviews list motivations for prosocial behaviour as: cooperation and competition (strategic); fairness, equity and equality (recipient-independent); and reciprocity (Adamus, 2017). I have not discussed reciprocating generosity as a motivation for prosocial behaviour in detail as it has limited relevance to prosocial behaviour towards strangers. Reciprocation could be driven by increased anticipated guilt through expectations (Chang and Sanfey, 2013) or conceptualised as the situational factor of having positive information available about the recipient (that they are generous), as discussed below.

Meier (2011) describes prosocial preference models, in which the utility of outcomes for others can influence one's utility directly (outcome orientation); indirectly (action orientation); or depending on inequality (recipient-independent). This paper also

highlights the contribution of self-identity (recipient-independent) and reciprocity (strategic). Finally, an experimental paper separated conceptualisations of fairness which can also be aligned with the present categories: impartiality (recipient-independent); charity (recipient-dependent); and reciprocity (strategic or responding to generosity) (Niemi et al., 2018). Activation patterns in theory-of-mind regions differentiated between these categories.

A core aspect of my categorisation of motivations is how the input information defines the benefits of being prosocial and the costs of not being prosocial. These are then compared with the cost of the prosocial act and the benefit of selfishly keeping resources instead. This idea of costs and benefits is not new and is applied from research on decision making outside of prosocial contexts. Gęsiarz & Crockett (2015) apply the three broad decision-making systems to prosocial behaviour. Cost and benefits of prosocial behaviours could either be predicted by a goal-directed system or affect future decisions through reinforcement learning. These systems likely work alongside a third system of relatively automatic responses, which developed through evolution, with emotional contagion an example relevant to prosocial behaviour.

The role of costs and benefits is also formalised in the sociocultural appraisals, values, and emotions framework for prosocial behaviour (Keltner et al., 2014). This model includes terms for: benefits for the self (extrinsic – strategic & intrinsic – altruistic); benefits for the recipient (outcome-orientation); and costs of inaction (punishment – strategic & guilt – altruistic). The sum of these terms is weighed against the costs of the prosocial action. All costs and benefits are scaled depending on what the default action is, defined by the individual differences. Terms are also scaled by several situational factors: recipient-independent factors; recipient-dependent factors; and the social encouragement of or resistance to prosocial behaviour. In the next section I consider these individual and situational factors

2.8. Situational factors

Much of the research described above used experimental manipulations to isolate contexts in which only certain prosocial motivations can drive behaviour, and tested differences in behaviour between these. For example, making donations anonymous and removing opportunities for reciprocity isolates altruistic motivations. This decreases levels of giving, compared to strategic contexts. Such situational differences are usually part defining and testing different motivations. In addition to these, other situational factors affect giving (Adamus, 2017) and these offer further insight into differentiating motivations.

As prosocial decisions are defined by benefitting others, who they are is likely to be important. Previous research has identified a number of factors relevant to the recipients that impact prosocial and altruistic decisions. Papers 3 and 4 expand the behavioural work on several key examples, described below, using physiological and neural measures.

2.8.1.   The number of people

Anecdotal evidence from events in the real world and behavioural studies outlined here suggest the relationship between the number of people suffering or in need and the response is not linear. Responses include emotional reactions such as sadness or outrage, media coverage, and altruistic behaviours, such as donations to charity. The balance of these responses is likely to depend on whether the people have already suffered or whether the situation can still be resolved. There is also evidence that stories of people who have already died can promote altruistic behaviours to help others affected by a similar issue (Slovic et al., 2017).

A nonlinear scaling of responses, as the number of people increases, is known as scope insensitivity (Dickert et al., 2015) or psychophysical numbing (Fetherstonhaugh et al., 1997). An extreme example is the singularity effect, in which the largest responses are to just one person, creating an inverse relationship (Schelling, 1968; Västfjäll et al., 2014). In a study on refugees, images of large groups, compared to small groups, negatively impacted compassion and political attitudes (Azevedo et al., 2019). Learning about this bias reduced contributions to singular beneficiaries to the same, lower level as statistical appeals, rather than raising giving to larger numbers of people (Small et al., 2007). Greater humanisation of single others, compared to groups, could also explain aversions to sacrificing one life to save others in moral dilemma situations (Majdandžić et al., 2012).

An alternative pattern is that as the number of people increases, the response also increases but not in a proportional way. This nonlinearity has been demonstrated in the policymaking (Olivola, 2015) and charitable giving (Cameron and Payne, 2011) domains. Finally, in some situations, these could patterns combine so responses increase up to a certain number of people, but then plateau or decrease beyond this point. There are two broad explanations for this collapse of compassion, an inability to feel emotional when large numbers of people suffering (Slovic, 2007) or a motivation to avoid doing this (Cameron and Payne, 2011).

Scope insensitivity mainly relates to responses towards a whole group of different sizes. A related but separable bias is proportion dominance, the effect of the reference group size on responses towards an individual or subgroup of people (Bartels, 2006). This effect,

also known as pseudoinefficiency (Västfjäll et al., 2015), describes situations in which the proportion of people saved or helped is valued over the absolute number of people.

Proportion dominance becomes non-normative when less people are helped as a result. For example, in a choice between two medical aid programs, choosing to support option A, to save 68 of the 80 lives in one camp, over option B, to save 70 of the 680 lives in a second camp, sacrifices two people to help a greater proportion (Bartels, 2006). Similarly, the number of people who cannot be helped introduced negative emotions that damped positive feelings about helping (Västfjäll et al., 2015). There is evidence that proportion dominance is caused by faulty deliberative reasoning (Mata, 2016), such that people perceive greater utility from helping a higher proportion (Erlandsson et al., 2014).

2.8.2. *Amount of information*

Another factor shown to affect levels of prosocial behaviour is the amount of information about the recipient. This often works in tandem with the singularity effect to create what is known as the identifiable victim effect (Lee and Feeley, 2016) although it should be noted that describing beneficiaries of charitable donations as "victims" can be demeaning, so is avoided here. Identifiability combined with singularity is often utilised in charity campaigns that describe a single recipient with personal information such as their name and a photo.

It could be argued that information is one possible mechanism behind scope insensitivity, as it is only feasible to have information about small numbers of people, but there is also evidence these factors are separable. For example, a meta-analysis found a small but significant effect of an identified person on prosocial responses, independent of number, particularly when a monetary appeal featured a photograph of a child suffering from poverty (Lee and Feeley, 2016). Similarly, a photo of the recipient was shown to increase activation in the nucleus accumbens during charitable donations (Genevsky et al., 2013).

Information about the need of the recipient has also been used as an experimental manipulation, to generate conditions of high empathy. In one example, information about need decreased negative ERP signals in response to costly transfers to others (Liu et al., 2018). Certain types of information may promote empathy more than others, for example a target is sad, rather than angry (Sassenrath et al., 2016).

Information could also be in the form of requests from the beneficiary, which were common when speaking was allowed in the dictator game and increased giving (Andreoni and Rao, 2011). However, when only dictators could speak, they were more selfish, unless

instructed to put themselves in the recipient's shoes. One possible explanation is that a lack of information allows justifications for selfishness in the dictator game, for example that the recipient could be richer than the decider (Aguiar et al., 2008). Finally, greater contributions to beneficiaries who have already been selected, compared to those who will be selected in future, showed the power of minimal identifiability (Small and Loewenstein, 2003).

2.8.3. Recipient identity

Information about the person or people who benefit introduces several additional factors that affect giving, including personal attributes. Much research has focused on the relationship between the personal attributes of the decider and the recipient, with similarity predicting giving (Loewenstein and Small, 2007). A common manipulation is whether beneficiaries are part of the deciders ingroup or outgroup, with increased giving towards ingroup members (Meier, 2011). Empathy was a stronger predictor of ingroup than outgroup helping (Stürmer et al., 2005) and interacting with members of ones ingroup and outgroup activates different neural networks (Rilling et al., 2008). Giving was higher to liked, compared to disliked, peers and patterns of brain activity differed between conditions (Schreuders et al., 2018).

Social distance can also be considered along a continuum, rather than a binary factor of ingroup or outgroup, with the closest others being good friends, family and partners. Feelings of interdependence towards friends predicted vicarious reward responses equal to rewards for the self (Varnum et al., 2014). Similarly, passively observing a friend, compared to a stranger, play a gambling game showed stronger ERP responses (Q. Ma et al., 2011). Dorsal striatal activity was related to affective empathy when helping a romantic partner, but cognitive perspective taking when helping a stranger (Kawamichi et al., 2013). This suggests different processes underlie helping others at different social distances. Similar effects of social closeness on (non-costly) helping behaviours have also been shown in monkeys (de Waal et al., 2008).

Much work on social closeness has contrasted very close others with strangers or compared ingroups with outgroups are that stigmatised or in conflict. Whether this factor is significant for more subtle differences in social distance is less clear. For example, ties to one's home country was the least important factor in predicting donations to developing countries (Hansen et al., 2014). Factors representing the need, including hunger, malnutrition, and child mortality rates, were much stronger predictors of giving.

2.8.4. Proximity

In addition to social closeness, physical proximity can also increase prosocial behaviour. At the extreme, being in the presence of someone suffering is a powerful motivator. A thought experiment asks whether we should save a child from drowning in a pond, even if it means ruining expensive shoes and clothes. This is contrasted with donating the cost of shoes and clothes to save the life of a child in a distant place (Singer, 1972).

Physical proximity was shown to increase donations, with greater expected impact underlying this effect (Touré-Tillery and Fishbach, 2017). Proximity also motivated donations following natural disasters, through counterfactual thinking that the donor could have been personally affected by close events (Zagefka, 2018).

Like the overlap between singular and identified beneficiaries, social and geographical proximity often align, so are difficult to separate in experimental designs. Results from manipulating salient identity suggested that identity mechanisms are responsible for physical proximity effects (Levine and Thompson, 2004). Nagel and Waldmann (2013) also found that the distance effect on moral obligation disappeared when confounds, including informational directness and personal involvement, were accounted for.

Proximity could also apply to closeness in time, of either the need for help or the impact of helping. Outside prosocial behaviour, research shows evidence of temporal discounting, a preference for positive outcomes closer in time (Frederick et al., 2002). How temporal discounting relates to prosocial decisions is likely to depend on the specific choice, including the delay length and whether it applies only to the cost to the self, only to the benefit for others, or both these outcomes.

Interestingly, some research has looked at the relationship between social discounting, based on social closeness, and temporal discounting. These two forms of discounting were associated with common activity patterns in brain regions including frontal areas, linked to control, and mesolimbic reward networks (Hill et al., 2017). An experimental manipulation of thinking about outcomes for others, or in the future, reduced social discounting (Yi et al., 2016). Patterns of choices were shown to be similar between those affecting others and affecting oneself in the future (Pronin et al., 2007).

2.8.5. Frame

Another application of research outside prosocial behaviour to altruistic decisions is the idea of framing outcomes as positive or negative. As described in prospect theory,

losses are often perceived as more salient than gains (Kahneman and Tversky, 1979). When giving money to another player in a dictator game, the receiver is not commonly framed as suffering or in need, although it may be suggested that they have experienced bad luck or unfairness in the allocation. The gift is likely to be conceptualised in a gain frame, increasing the welfare of the receiver above their current level of wellbeing.

In contrast, donations to charity could be framed as avoiding losses, in life, health, and wellbeing, or the same content could be presented as opportunities to save or improve a life. There has been debate and mixed findings on which frame is more effective. For example, participants helped refugees more when their action prevented a loss, rather than achieved a gain (Böhm et al., 2018), whereas other theories suggest positive messages about the impact of help are more powerful. Erlandsson et al. (2018) argue that these differences can be explained by the operationalisation of effectiveness. Positive appeals generate favourable attitudes towards the cause, but negative appeals are better or equally as effective in promoting donations.

Losses and gains may also be differentially affected by temporal discounting (Tanaka et al., 2014), demonstrating the potential for interactions between factors. Similarly, abstract statistical information in a charity appeal was most effective when combined with a negative frame, whereas the efficacy of anecdotal information was increased through a positive frame (Das et al., 2008). In contrast, another study manipulating message framing and whether statistical information used small (1 in 3) or large (700 million in 2.1 billion) numbers found positive frames better for large numbers but negative frames for small numbers (Chang and Lee, 2010).

While negative information or upsetting photos may be motivating, most real-world giving behaviours involve several points at which the person could disengage from the need. For example, someone could not open appeal letters or change channels during a television appeal. Research on the link between valence and information seeking is also relevant to prosocial decisions. When information about outcomes for the self has no instrumental value, results showed a bias towards information likely to be positive, and avoidance of information likely to be negative (Sharot, 2011). Findings from corresponding neuroimaging studies suggest an intrinsic value of information likely to be good news (Charpentier et al., 2018). Optimistic updating of beliefs has also been shown in relation to outcomes for others (Kappes et al., 2018).

2.8.6. Other

In the above review, I have outlined some of the key situational factors, with a focus on those relevant to the thesis. This is by no means a comprehensive list. Other factors linked to the recipient that increase giving include perceived deservingness (Petersen et al., 2011; Seu, 2016), lack of perceived responsibility for the situation (Zagefka and James, 2015) and "reference-dependence": whether the problem is due to a recent change rather than chronic (Small, 2010).

Considering the potential contributor, mood (both state and trait) can affect altruism (Boenigk and Mayr, 2015; Drouvelis and Grosskopf, 2014) but there were mixed findings on the directionality of this effect and interactions with the wider context (Weyant, 1978). The presence of other potential contributors negatively impacted helping. Explanations for this "bystander effect" include freezing (Hortensius and de Gelder, 2018) and strategic mentalising to calculate that someone else will help (Thomas et al., 2016). In contrast, the presence of others who are observers, not potential helpers, increased prosocial behaviour, with this effect stronger for friends than strangers (Li and Zhao, 2019). Other relevant factors that define the wider context of prosocial behaviour include the salience of social norms promoting helping or sanctions for not helping (Meier, 2011).

This chapter has focused on factors that increase the benefits of helping or the costs of not helping. There are also situational factors that increase the costs of helping such as high self-risk. Evidence from fMRI and tDCS support a role of the right dlPFC in integrating levels of this self-interest with other-need in decision making (J. Hu et al., 2017). In a dictator game where the endowments were at risk of being lost, participants actually increased their giving despite the risk that there would be a cost but no benefit (Engel and Goerg, 2018).

2.8.7. Relevance to motivations

Understanding the factors that increase and decrease the likelihood and extent of prosocial behaviours is crucial for organisations that rely on donations and volunteering. These factors also give further insight into the motivations driving prosocial decisions. Different factors enhance the various benefits of giving, and costs of not giving, that each motivation relates to. For example, the fact that information about the recipient affects behaviour suggests this *recipient-dependent* input is considered during decisions to help. Some situational factors create biases, where behaviour diverges from that which would be expected if motivations were *outcome oriented*. Examples are scope insensitivity and non-normative proportion dominance. These biases suggest *action-oriented* motivations

promote giving in the way that makes the donor feel best, rather than maximising the benefits for recipients.

A series of studies on the proportion dominance effect, identifiability effect and ingroup effect highlight the role of different mechanisms in explaining situational effects on giving (Erlandsson et al., 2017, 2015, 2014). Ratings of charitable appeals and justifications after donation decisions suggested emotional reactions drive the identifiability effect, perceived impact underlies proportion dominance, and increased responsibility for helping ingroups accounts for that effect.

2.9. Individual differences

The previous section highlights situational factors that can make people more or less prosocial. There is also huge variance between people in levels of prosocial behaviour and altruism. Individual differences along the selfish to selfless spectrum have been linked to biological and genetic differences (Sonne and Gash, 2018), with the heritability of helping behaviour estimated at 11% (Primes and Fieder, 2018). Despite situational influences, prosociality has been shown to be relatively stable across time and different economic games (Yamagishi et al., 2013), leading to claims of a 'cooperative phenotype' (Peysakhovich et al., 2014). This domain-general prosocial trait was also called a general benevolence dimension and there is evidence it increases with age (Hubbard et al., 2016).

Another related term is social value orientation, measured through questions on hypothetical giving decisions that vary in costs and benefits to self and other (Murphy et al., 2011). Answers categorise individuals as prosocial (cooperative or altruistic) or proself (competitive or individualistic). Prosocials reported more charitable donations, especially to causes helping the poor and the ill (Van Lange et al., 2007). There were also differences in brain activity, with only prosocials showing ERP sensitivity to other gain vs loss (X. Hu et al., 2017), and positive other-regarding reinforcement learning signals in vmPFC (Christopoulos and King-Casas, 2015). Finally, in a strategic cooperation task, different decision-making strategies were inferred from different patterns of neural activity between prosocials and proselfs (Emonds et al., 2011).

There has been much debate on whether being prosocial or selfish is the default or intuitive response (Rand, 2016). Evidence in both directions has come from reaction time and brain stimulation experiments, although there are issues with these inferences (Hallsson et al., 2018). It now seems that recognising individual differences may explain these mixed results. Faster decisions resulted in choices more closely aligned with general preferences (Cornelissen et al., 2011; Yamagishi et al., 2017). Proself individuals also

showed slower responses, and more activity in neural regions linked to cognitive control, during non-costly prosocial decisions (Kuss et al., 2015). However, a different study concluded that helping is more effortful for prosocial individuals (Ruci et al., 2018).

Identifying independent predictors of individual differences, rather than measuring these differences directly, provides further insight into the motivations behind prosocial behaviour and their neural basis. In this sense, psychological traits may be more relevant than demographic variables such as age, although demographics are often more practically useful for fundraising organisations (Bekkers and Wiepking, 2007).

2.9.1. Empathy

One of the most widely studied traits in relation to prosocial behaviour is empathy, which can be separated into cognitive and emotional components at the trait, as well as the state, level. Generosity in the dictator game was linked with greater theory of mind abilities (Lang et al., 2018). Perspective taking may also indirectly effect prosociality, through decreasing the relationship between negative emotions and personal distress in volunteers (Eisenberg et al., 1994). Perhaps perspective taking prevents self-focused responses dominating. Trait prosociality was also linked to resting current density in the TJP (Gianotti et al., 2019), a region often linked to theory of mind (Decety and Lamm, 2007). However, this region is also associated with attention. Prosociality has been linked with an ERP component associated with salience (Chiu Loke et al., 2011), in line with this alternative interpretation.

Another component of empathy is emotion sharing or neural self-other overlap in emotion processing. Extremely altruistic individuals, who donated a kidney, showed more similar representations of self and other pain, both when anticipated and experienced (Brethel-Haurwitz et al., 2018; O'Connell et al., 2019). Extraordinary altruists also had increased behavioural sensitivity and right amygdala volume and responsivity to fearful expressions, whereas psychopathic people showed the opposite pattern (Marsh et al., 2014). Similarly, people who struggle to recognise and experience the distress of others were less generous, had reduced activation in the AI and TPJ, and had greater alexithymia: an inability to recognise one's own emotions (FeldmanHall et al., 2013). Finally, activity in networks associated with self-other resonance positively correlated with dictator game offers (Christov-Moore and Iacoboni, 2016). In contrast, connectivity between these regions and areas linked to top-down control was negatively correlated with offers.

In addition to enhanced overlap for negative emotions promoting altruism, empathy can increase reward responses for others. Empathic care, consisting of both positive and

negative emotions, was associated with activity in nucleus accumbens and OFC and both this neural marker and one for personal distress predicted charitable donations (Ashar et al., 2017). Empathic concern also correlated with ventral striatum activity during gains for a charity, reported enjoyment when winning for the charity and larger donations to charity (Spaans et al., 2018). Similarly, striatal activity when watching a friend win correlated with prosociality (Morelli et al., 2018). Finally, the reward signal in the ACC gyrus was unique to others' rewards for participants high in emotion contagion, but also signalled rewards for the self in those low in emotion contagion (Lockwood et al., 2015).

More broadly, empathy correlated with prosocial intervention behaviours and the amplitude of a prefrontal ERP response to interventions (Balconi and Canavesio, 2013). Empathic concern, but not compassion, was linked with increased activity in the subgenual cingulate during experiences of guilt (Zahn et al., 2009). Trait empathic concern, rather than trait personal distress, predicted costly altruism and activity in the subgenual cingulate, caudate, and ventral tegmental area (FeldmanHall et al., 2015). Increasing trait compassion through training increased altruistic behaviour, activity in the inferior parietal cortex and dlPFC, and connectivity from dlPFC to NuAcc (Weng et al., 2013).

In addition to evidence from neuroimaging, research has linked other biological mechanisms with empathy and prosocial behaviour. Vagus nerve activity is one example, with evidence that this regulates the physiology underlying emotional expression and experience, emotional and cognitive empathy, and personal distress (Kogan et al., 2014). These authors show a quadratic relationship between vagal activity and prosocial measures including traits, emotions, and observer ratings, such that a medium level of activity predicts prosociality.

The heart is one organ innervated by the vagus nerve and direct measures of the heart have also been linked to prosocial behaviour, in particular heartrate variability, which differs between people. Those with high heartrate variability were better at identifying others' positive mental states but there was no difference for negative states (Lischke et al., 2017). Heartrate variability can also be regulated voluntarily and individual differences in how well participants could do so predicted differences in altruistic behaviours (Bornemann et al., 2016).

In summary, empathic concern has been associated with a variety of different prosocial behaviours. However, in one study, meaningful correlations were only found for spontaneous, informal helping behaviours towards someone in need who was present (Einolf, 2008). The relationship was much weaker for planned behaviours towards people

not present. Differences in neural activity between people high in empathy and 'systemisers' also seem to disappear during explicit empathy tasks (Riekki et al., 2018).

Although empathy is commonly linked with positive increases in prosocial behaviour, there can be downsides. Empathy can lead to helping behaviours more defined by social preferences and biased by partiality, drivers which may conflict with fairness and justice (Batson et al., 1995; Bloom, 2017; Decety and Cowell, 2015; Ugazio et al., 2014).

2.9.2. Guilt

Like empathy, guilt is a process underlying prosocial behaviour across people, but also differs as a trait between people. An association between traits such as empathy and guilt support their role in prosocial behaviour more broadly. While empathy is the process of understanding others' thoughts and emotions that can affect one's own emotions, guilt is the negative emotion associated with not helping, so anticipating guilt promotes helping.

Proneness to guilt at the trait level improved models of prosocial behaviour above empathy alone and this effect was stronger for women (Torstveit et al., 2016). It could be suggested that guilt is a product of empathy, if it relies on understanding others, but there is not strong evidence for this relationship (Lindsey et al., 2007).

2.9.3. Ambiguity

An interesting but often overlooked aspect of prosocial behaviour is uncertainty which can be separated into risk (known probabilities of outcomes) and ambiguity (unknown probabilities). In strategic games, the probability that a cooperative decision increases extrinsic reward depends on predictions about the others' behaviour. Tolerance to ambiguity, but not risk, predicted increased cooperation and trust in strategic games, but only when there was some level of ambiguity about the other player's actions (Vives and Feldmanhall, 2018). In the real-world, it is unlikely that altruistic decisions lead to outcomes for the beneficiary with 100% certainty. Individual levels of risk aversion were shown to be positively associated with inequity aversion in a distributive justice task (Palmer et al., 2013b).

2.9.4. Other

As with the section on situational factors, many more individual-difference measures have been studied in relation to prosocial behaviour than are covered here. Generally desirable traits that show positive relationships with prosocial behaviour include agreeableness (Habashi et al., 2016), gratitude (Yost-Dubrow and Dunham, 2018), self-control (Martinsson et al., 2012), and lack of avoidance attachment (Richman et al., 2015).

Interestingly, narcissism (Chen et al., 2019) and neuroticism (Pinazo et al., 2016) are also linked to increased prosocial behaviour. In support of strategic or recipient-independent motivations, trait reputational concern increases helping towards strangers (Kawamura and Kusumi, 2018), but so does belief in other-focused altruistic motivation (Carlson and Zaki, 2019).

Other traits show more nuanced associations with prosocial behaviour, for example justice sensitivity predicted the stability of altruism (Lotz et al., 2013). Emotional intelligence moderated the impact of successful or unsuccessful outcomes on helping behaviour (Agnoli et al., 2015). Two traits associated with mindfulness also both predicted helping behaviour (Cameron and Fredrickson, 2015). Present-focused attention was related to positive emotions, whereas non-judgmental acceptance negatively correlated with negative emotions, like stress and guilt, during helping.

Socioeconomic status is an interesting demographic variable as it combines financial and identity components. While wealthier individuals give more overall, the relationship between wealth and proportion of income donated is not so clear (Bekkers and Wiepking, 2007). There is evidence that individuals of a lower social status are more prosocial (Guinote et al., 2015) and suggestions that the relationship between class and helping depends on motivation, identity and inequality (Piff and Robinson, 2017). In the brain, stronger responses to others' pain was positively associated with donations in high subjective socioeconomic status, but negatively associated with donations in low status individuals (Y. Ma et al., 2011).

2.10. Interactions between situational and individual factors

The above sections summarise some of the main individual differences and contextual factors that affect prosocial behaviour, but these can also interact in different ways. A simple interaction is that the contextual factors may have different strength relationships with prosocial behaviour between individuals. For example, only some people showed an ingroup effect in prosocial behaviour and those who didn't actually preferred rewards for the outgroup (Hackel et al., 2017). There may be trait measures that account for these differences, for example powerful people dehumanised outgroup members to a greater extent (Lammers and Stapel, 2011). Individual differences in the strength of other situational effects are accounted for by decreased trait empathy. These include greater temporal discounting for distant others (O'Connell et al., 2013) and a greater detriment of cognitive load on empathic responses (Rameson et al., 2012).

Considering the situational factors relevant to the number of people, those low in analytic thinking donated more to one identified beneficiary, compared to statistical need or both. In contrast, those higher in the analytic thinking trait did not show this effect (Friedrich and McGuire, 2010). There was no evidence for a mediating role of emotion in these results. Trait numeracy ability also interacted with the format of donation requests (Dickert et al., 2011a). Only those low in numeracy ability were affected by mental images, whereas perceived impact of a donation correlated positively with donations, regardless of numeracy ability. Numeracy ability was also negatively associated with donations in response to requests with a loss frame (Mayr, 2012).

Another final possible pattern of interaction is when trait differences in prosociality predict different responses to situational manipulations. For example, in experimental conditions to create empathy and reciprocity-motivated altruism, selfish people only increased their altruistic behaviour in response to empathy inductions, whereas altruistic people only increased altruism in response to reciprocity inductions (Hein et al., 2016). Traits have also showed relationships with specific aspects of helping situations, such as empathy increasing perceived injustice (Urbanska et al., 2019) or agreeableness promoting donations towards targets seen as deserving (Yarkoni et al., 2015).

3. Overview of the thesis

The above review covers the existing literature on motivations for prosocial behaviour and the neural basis of these, as well as the situational and individual factors that increase or decrease prosociality. The following section highlights outstanding gaps in the literature and the contribution of the studies in this thesis.

3.1. Summary of papers 1 and 2

The studies described above support the existence of both strategic and altruistic motivations for prosocial behaviour and begin to suggest their neural correlates, but some results are mixed. Two previous studies comparing the strategic ultimatum game with the altruistic dictator game found different results (Weiland et al., 2012; Zheng and Zhu, 2013).

Paper 1 uses fMRI meta-analysis to identify the neural correlates of altruistic decisions, strategic decisions, commonalities across, and differences between these categories of prosocial decision. The paradigms that create opportunities for strategic motivations, through extrinsic rewards for being prosocial, do not exclude the possibility that altruistic motivations also play a role in these decisions. However, a comparison of contexts where extrinsic rewards are possible, with those where they are not, can identify the additional processes involved in strategic motivations for prosocial behaviour.

An fMRI meta-analysis also increases statistical power and overcomes issues with existing single-study fMRI papers, including small sample sizes (Button et al., 2013; Cremers et al., 2017) and low statistical thresholds (Eklund et al., 2016). However, results of meta-analysis are still affected by variable coverage in the original studies caused by lack of signal. Paper 2 presents a novel technique to account for this issue and demonstrates its effect on two data sets.

One of the key results from paper 1 shows that both types of prosocial decision, compared to choosing the selfish alternative, correlated with activity in reward-related regions, including the NuAcc and vmPFC. These findings support the idea that other people can have value in both an intrinsic and extrinsic way. Others offer opportunities for collaboration and cooperation, which can increase extrinsic benefits, such as money or other resources. For altruistic decisions, this shows the neural basis of the intrinsic reward or warm glow. This raises the question, why is being altruistic rewarding?

3.2. Summary of papers 3 and 4

As identified in the literature review, one possible reason that altruism is rewarding is that people value the lives of others. If this is the case, we can then consider *when* and *why* people value the lives of others. Based on the findings in the literature review, particularly on empathy and vicarious reward, we can predict that if people do value the lives of others they will:

i. Get distressed when bad things happen to others
ii. Feel good when good things happen to others
iii. Be interested in what happens to others
iv. Show similar biases in valuing outcomes for self and other (on the basis that people value their own life and wellbeing)
v. Take actions to improve outcomes for others

Papers 3 and 4 test the first four of these predictions to some extent by measuring physiological and neural responses respectively to others' lives being at risk and the outcome of whether they live or die.

We created the stimuli used in these studies from real news stories. In addition to testing overall responses to deaths and threats to life, the designs also manipulate whether the story happened in the UK or abroad and how many people were affected. These factors relate to two biases I described in the section on situational factors: geographical proximity and scope insensitivity. So far, these biases have only been researched using behavioural measures.

However, widely used self-report techniques are not well suited to measuring biases, particularly within-subjects. Participants may be unaware of their biases or the true motivations behind their decisions. A further issue with any measure that participants have conscious control over is the social desirability of the responses or behaviours of interest (Fernandes and Randall, 1992). Participants may act differently in experimental settings than they would in the real world, undermining the relevance and utility of the findings.

Overall, we find physiological arousal responses and activity in areas associated with negative emotions and empathy when lives are at risk and lost. In some contexts, deaths are differentiated from the people surviving. These findings are in line with people valuing the lives of others.

However, our results also show that skin conductance and fMRI measures reveal biases that participants do not report in behavioural measures. Specifically, physiological

and neural responses vary, based on geographical location, and show a non-linear relationship with the number of people. These findings suggest the extent to which the suffering of others affects us varies along these factors. The value of a life is not constant.

3.3. Summary of paper 5

Papers 3 and 4 give evidence that the lives of others are valued, from physiological and neural responses to passively viewing information about other people. The next question is how this translates into motivations to improve outcomes for others. As described in the literature review, altruism can be outcome oriented, focused on the change in wellbeing for the other, or action oriented, focused on personal responsibility for that change.

Paper 5 presents five experiments that aim to develop a behavioural measure of the role of responsibility for an outcome in determining interest in that outcome. The design was inspired by previous work showing that participants were more interested in outcomes that are likely to be positive, than those likely to be negative, when they affect the participant themselves (Charpentier et al., 2018). We first test whether this bias towards positive outcomes extends to outcomes affecting other people, then examine whether it depends on being responsible for that outcome.

The general method used in paper 5 involved participants completing rounds of a task to win money on behalf of themselves and another person. Another participant did this task simultaneously and a fourth person had already completed the task. Participants could then find out about outcomes. The multiple players provide a 2x2 design of who was responsible for the outcome (I play or they play) and who was affected by it (for me or for someone else). Outcomes in these 4 conditions also had either a positive or a negative frame. The main measure of interest was whether participants chose to find out the outcome of the trial.

We find evidence of outcome-oriented motivations to some extent, as participants did find out about outcomes affecting others that they were not responsible for. In some contexts, they showed a similar effect of frame (positive or negative) for outcomes affecting others and themselves. However, we also find evidence that being responsible for an outcome increases interest, in line with action orientations. Both interest and the extent of a bias towards positive information were increased when the participant was involved in some way – either responsible for or affected by the outcome.

Abstracts for papers 3, 4 and 5 are omitted to avoid repetition with this section.

A comparative fMRI meta-analysis of altruistic and strategic decisions to give

Jo Cutler* [1] and Daniel Campbell-Meiklejohn* [1]

[1] The Social Decision Lab, the University of Sussex, School of Psychology, Falmer, UK.

*Correspondence:

Jo Cutler – J.Cutler@sussex.ac.uk

Daniel Campbell-Meiklejohn – daniel.cm@sussex.ac.uk

The decision to share resources is fundamental for cohesive societies. Humans can be motivated to give for many reasons. Some generosity incurs a definite cost, with no extrinsic reward to the act, but instead provides intrinsic satisfaction (labelled here as 'altruistic' giving). Other giving behaviours are done with the prospect of improving one's own situation via reciprocity, reputation, or public good (labelled here as 'strategic' giving). These contexts differ in the source, certainty, and timing of rewards as well as the inferences made about others' mental states. We executed a combined statistical map and coordinate-based fMRI meta-analysis of decisions to give (36 studies, 1150 participants). Methods included a novel approach for accommodating variable signal dropout between studies in meta-analysis. Results reveal consistent, cross-paradigm neural correlates of each decision type, commonalities, and informative differences. Relative to being selfish, altruistic and strategic giving activate overlapping reward networks. However, strategic decisions showed greater activity in striatal regions than altruistic choices. Altruistic giving, more than strategic, activated subgenual anterior cingulate cortex (sgACC). Ventromedial prefrontal cortex (vmPFC) is consistently involved during generous decisions and processing across a posterior to anterior axis differentiates the altruistic/strategic context. Posterior vmPFC was preferentially recruited during altruistic decisions. Regions of the 'social brain' showed distinct patterns of activity between choice types, reflecting the different use of theory of mind in the two contexts. We provide the consistent neural correlates of decisions to give, and show that many will depend on the source of incentives.

1. Introduction

The decision to share resources is a cornerstone of any cooperative society. The motivations that drive these choices, however, will vary. Giving can be driven by intrinsic rewards, such as conditioned satisfaction from performing a generous act or the image of oneself as a 'good person'. On the other hand, giving can also be driven by strategic forethought of extrinsic rewards that might be gained through reciprocity, avoidance of punishment, or a public good. Some may argue that 'why' we give does not matter, as long as we do. However, understanding the 'why' is essential for determining the likelihood of prosocial behaviour in the absence of extrinsic benefit, such as when the beneficiary could never return the favour or when societies, which depend on prosocial behaviour, do not provide defined incentives. It can also help us understand how intrinsic and extrinsic drives interact in the decision process.

Over a decade of innovative neuroimaging studies have provided a fresh window into the old problem of why we give. Through this lens, we can see whether different motivations to help one another use different neural (and therefore cognitive) mechanisms. This then provides the basis for studying how these neurocognitive mechanisms may vary independently between contexts and individuals. This insight could also help to explain other phenomena. For instance, overlapping anatomy of intrinsic and extrinsic drives could underpin the effect of extrinsic incentives 'crowding out' altruistic motivations (Frey and Oberholzer-Gee, 1997), or make clear how intrinsic and extrinsic benefits sum in strategic decisions to help each other.

However, no systematic meta-analysis has examined the consistency of these findings on prosocial decision-making (see Filkowski et al., 2016 and Luo, 2018 for descriptive reviews, Gabay et al., 2014 for neuroimaging meta-analysis of ultimatum game responders and Bellucci et al., 2017 for trust games). With this meta-analysis, we investigated the consistent neural correlates of decisions to give and differences in these correlates that depend on whether there is potential for extrinsic gain through the interaction.

We define *altruistic* choices to give as generous acts with no opportunity to gain extrinsic rewards as a direct result of that interaction. Motivations for giving in these contexts rely on intrinsic rewards. Sources of intrinsic reward or 'warm glow' (Andreoni, 1990, 1989) can include vicarious reward experience (Mobbs et al., 2009); relief of empathic concern (FeldmanHall et al., 2015); self-enhancement from adherence to moral codes or social norms (Niemi et al., 2017); and conditioned reinforcement (e.g. from

parental feedback). Warm glow could also result from inferences of enhanced reputation (Izuma et al., 2010) in the eyes of an experimenter or omniscient religious figure, despite no defined benefit of that enhanced reputation. Intrinsic incentives to give are often studied with dictator games (Kahneman et al., 1986), donations to charities, or payments to prevent others from coming to harm (Table 1).

We define *strategic* choices to give as generous acts that might increase the probability of a defined extrinsic reward. Strategic choices can involve the intrinsic rewards of altruistic choices (Capraro, 2017), but add the possibility of extrinsic gain, which is thought to be the dominant weight in the decision process (Frey and Oberholzer-Gee, 1997). Extrinsic benefits could come through avoiding punishment (Fehr and Gächter, 2002); reciprocity of the recipient (Falk and Fischbacher, 2006; Fehr et al., 2002); collective contributions to a public good (Chaudhuri, 2011); enhanced gains from cooperation (Fehr and Fischbacher, 2004) or rewards from defined future interactions through having an enhanced reputation (Milinski et al., 2001; Nowak and Sigmund, 2005). Paradigms used to study strategic giving (Table 1) include the ultimatum game (Güth *et al.*, 1982); trust game (Berg *et al.*, 1995); public goods game (Rapoport, 1987); prisoners' dilemma (Luce and Raiffa, 1957); and repeated versions of these games which include opportunities for reputation building.

For altruistic choices, goal-attainment based on most intrinsic incentives begins at the time of the action, without delay. As a result, most intrinsic rewards are relatively certain. Warm glow can be modulated, in part, by the gift's (delayed and uncertain) impact on the recipient, but for most paradigms (all those included here), the participant does not observe this impact. Strategic rewards, on the other hand, are weighted to the uncertain effects of the gift on extrinsic outcomes. In these cases, goal attainment comes after the action, when others respond. Keeping money (which we label 'selfish', as opposed to being 'prosocial' by giving, cooperating, or trusting) in a strategic context enhances certainty and immediacy of reward, but could also lead to less money. In the altruistic context, a selfish choice does not change outcome certainty to the same degree, and will always leave the participant with more money.

Table 1.
Explanation and categorisation of tasks used in studies

Task	Description	Group
Dictator game	Participant either chooses an amount of their money to give or accepts / rejects a proposed split between themselves and the other player.	Altruistic
Charity donation task	Participant either chooses an amount of their money to donate or accepts / rejects a proposed split between the participant and a charity.	Altruistic
Pain vs. gain	Participant can give up varying amounts of money, the more given the less painful the electric shock given to a partner	Altruistic
Ultimatum game	Participant proposes a split between themselves and their partner that is only implemented if the partner accepts it.	Strategic
Trust game	Participant transfers an amount of money to the trustee that is multiplied by some factor (often 3). The trustee then chooses an amount to send back which decides the payoff for both players.	Strategic
Prisoner's dilemma	Participant and partner decide whether to cooperate or defect. They gain mutual benefit if both cooperate but individuals gain more by defecting if the partner cooperates.	Strategic
Public goods game	Participants invest an amount in a communal fund that is then multiplied and divided among all players, including those who did not initially contribute to the communal fund.	Strategic

The use of theory of mind – inferring others' mental states (Frith and Frith, 2006), also differs between altruistic and strategic decisions to give. In altruistic contexts, theory of mind is likely to be more weighted on how another will feel, rather than what they will do, considering appreciation, change of emotion, or disappointment of the other. This could occur via empathy processes (Lockwood et al., 2015) – feeling what the other is feeling (Decety et al., 2015), mentalising, or both, with variability across different people (Tusche et al., 2016). Assuming that the inferred appreciation of a gift by the recipient would increase motivation to give, greater theory of mind is expected during altruistic decisions to give, relative to keeping the goods. Strategic decisions might balance that difference to a

degree. Intention inferences can motivate either generous or selfish choices and therefore be equally associated with selfish and generous decisions.

In sum, both altruistic and strategic choices incur immediate costs that benefit others but differ in the sources, certainty, and immediacy of the associated reward. Theory of mind is likely to contribute to both decision types, but differently in each, with different dissociations between the prosocial and the selfish choice alternatives, see Figure 1.
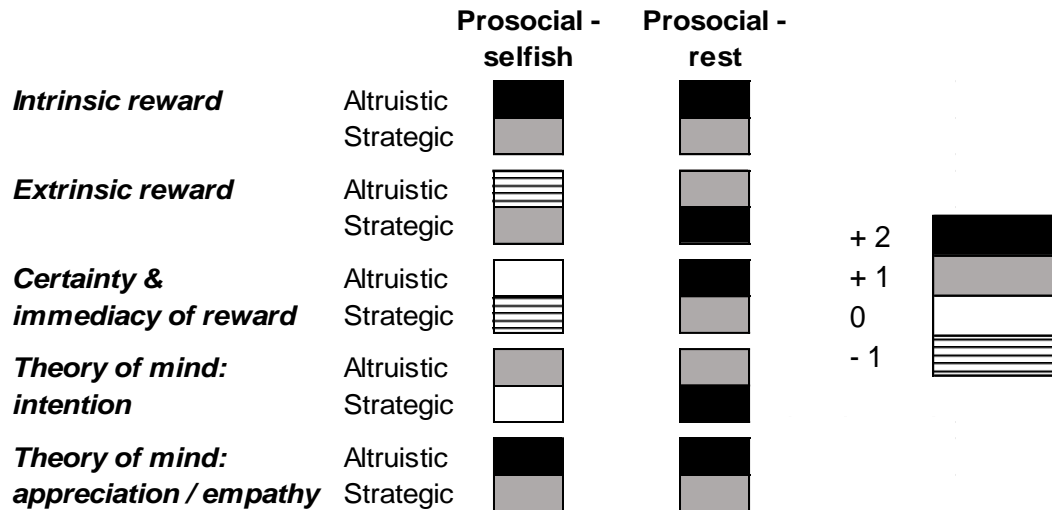
Figure 1. Theoretical presence in altruistic and strategic decisions to give compared to a selfish choice and to rest. This can act as a rough roadmap for interpreting neural differences between the two types of decision to give as well as each decision with the common controls used.

Two previous functional magnetic resonance imaging (fMRI) studies, which compared prosocial decisions in the altruistic dictator game with the strategic ultimatum game, report inconsistent findings and interpretations (Weiland et al., 2012; Zheng and Zhu, 2013). There was qualitatively no overlap of activation between them, creating concerns for the consistency and generalisability of reported differences. A third fMRI study using both ultimatum and dictator games to study childhood development of strategic social behaviour did not focus on the contrast between generous and selfish choices, but interestingly the difference in gift sizes between the two games (Steinbeis et al., 2012).

An fMRI meta-analysis integrates prior findings to increase the statistical power to detect reliable, consistent neural correlates of decision types (Müller et al., 2018; Wager et al., 2009, 2007). This is vital given that many fMRI reports are based on small sample sizes (Button et al., 2013; Cremers et al., 2017) and do not meet the new standards of high statistical thresholds recently shown to be required for confidence in effects (Eklund et al., 2016). In addition to the advantages associated with coordinate-based meta-analysis, our use of unthresholded maps enhances sensitivity (Radua and Mataix-Cols, 2012) and reduces assumptions of the spatial extent of activations. Our meta-analysis also allows new combinations and contrasts of different task types.

In this study, we do not make firm predictions of specific neural regions, remaining agnostic to previous findings as much as possible with a data-driven approach. One area that was identified as a region of interest a priori, however, was the ventromedial prefrontal cortex (vmPFC), as it is has been reliably linked to subjective value and decision-making (Bartra et al., 2013; Levy and Glimcher, 2012) including prosocial choices (Hare et al., 2010). It is also a large, cytoarchitecturally heterogeneous region (Mackey and Petrides, 2010) with varying connectivity (Sepulcre et al., 2010) and different sources of value show different activity patterns across it (Campbell-Meiklejohn et al., 2016; Sescousse et al., 2013). We therefore looked at activation across this region in more detail than the rest of the brain.

The present meta-analysis aims to answer four basic questions across the whole brain, with added focus on vmPFC:

i) What is common to altruistic and strategic decisions to give?

ii) What is consistent about altruistic decisions?

iii) What is consistent about strategic decisions?

iv) How do altruistic and strategic decisions differ?

There are examples of real-world decisions that do not fit into our defined groups, such as tax deductions from charitable giving or defined payments for blood donation. We also do not cover outcomes of decisions such learning a person appreciated a gift or rewards associated with choosing a winning strategy. The scope of this meta-analysis is fMRI paradigms for which the data at the point of a decision is available. By better understanding the neural basis of altruistic and strategic decisions as defined here, we provide a foundation for investigations in other contexts, as well as a milestone for research over the last fifteen years.

2. Method

2.1. Literature search and study selection

Literature searches using PubMed and Web of Science identified research added before September 2016. Keywords were either "fMRI" or "neur*" as well as one of: "altruis*", "charity", "charitable", "prosocial", "cooperation", "public goods", "social value orientation", "reputation", "dictator", "ultimatum", "trust game", "prisoner*". We identified additional potential articles from reference lists of selected articles or those offered by authors (Figure 2). Articles were considered if they reported novel fMRI data, not reported elsewhere, collected while participants made decisions, and analysed whole-brain data. For studies that either used psychopharmacological manipulations or tested populations other than healthy participants, we requested data from just the control group.

In addition, studies were screened for eligibility for the two groupings. Decisions in the altruistic group were defined as decisions that benefitted at least one other person at a cost to the self, with no potential for an extrinsic benefit. Paradigms were dictator games, pain vs gain or donation tasks. Decisions in the strategic group benefitted at least one other person but could also benefit the decision-maker. Common paradigms included trust and ultimatum games. Almost all selected studies focused on deciders rather than responders in these games. Responders, who may be prosocial due to reciprocity norms, have been studied previously (Gabay et al., 2014). One study (Garbarini et al., 2014) only had data from

the responder role in a trust game but with a clear incentive for participants to build up a prosocial and trustworthy reputation in order to receive future investments. All participants thought partners were human, which was true in some cases but not others. For all studies, the data were from the decision phase of the task.

Overall, we identified 35 altruistic and 45 strategic fMRI studies and requested statistical maps from the authors. Due to substantial methodological variation across studies, letters requested the decision of interest contrasted with as many as possible of a high-level control, baseline of some kind and the selfish choice. Obtaining multiple maps for a study maximised the likelihood of similarities between studies and shows whether specific contrasts affected results. If maps were not available, we requested coordinates for contrasts of interests or extracted them from manuscripts. Of these, a coordinate-based analysis was available for eight studies, while others did not report the required contrasts in manuscripts.

Of 35 altruistic studies, 18 authors provided usable maps (51% response rate) and three had coordinates, resulting in data from 557 participants in this group. Of 45 strategic studies, 10 authors provided maps (22%) and five had coordinates, giving data from 593 participants in this group.

2.2. Analysis

A combined image and coordinate-based meta-analysis of retrieved data was conducted using Anisotropic Effect Size Signed Differential Mapping software (AES:SDM, Version 4.31; Radua et al., 2014). When available, statistical maps enhance the sensitivity of the analysis and allow incorporation of both positive and negative differences. With coordinates, the software recreates estimated statistical maps from coordinates and their effect sizes using an anisotropic kernel. If the contrast of interest was available in both directions (e.g. cooperate > defect and defect > cooperate), the opposite peaks became negative t-values. The incorporation of study sample size increases the weight of larger studies.

A random-effects model, using the recommended 50 permutations, implemented all maps. This creates 50 random models with the same number of foci as the map of interest and tests the null hypothesis that the map's activations are the result of random distribution throughout the brain. Thresholding parameters used were those recommended by Radua et al. (2012) who found a voxel-level threshold of $p<0.005$ to approximate $p<0.05$ corrected and optimally balance specificity and sensitivity. Reported z-scores are specified as SDM-Z as they do not follow a standard normal distribution.

2.3.Contrasts and covariate

The available datasets used a range of control conditions broadly classified as rest (including one-sample tests), visuomotor controls, or selfish decisions. These controls vary significantly across multiple processes so we used an overall covariate in our models that codes for *comparator complexity* to control for differences, rating control conditions from 1 to 4 (see Supplementary Materials S1.2. for details).

For separate analyses of altruistic and strategic decisions, variance attributed to complexity was used to model contrasts between prosocial (generous) decisions and rest (a contrast where the covariate is at its modelled minimum) and between prosocial and selfish decisions (where the effect of the covariate is at its modelled maximum). By including the covariate in the model, we aimed to allow specificity in our interpretations of activity while maintaining statistical power (by including all available data). In addition, we specifically contrasted prosocial choices to the selfish alternative, only in the studies that had this comparison. This enabled tests of robustness that require a relatively homogenous group (see Supplementary Materials for details S1.3. and results S2.1.1.).

We used SPM12 (Statistical Parametric Mapping, http://www.fil.ion.ucl.ac.uk/spm) to extract areas of overlapping activation common to altruistic and strategic decisions, for example areas significant in both the altruistic > selfish and strategic > selfish maps. These maps were those from the overall analyses described above which use the comparator complexity covariate to model contrasts with rest and selfish decisions while incorporating all studies.

Differences between the decision types were calculated in both directions (altruistic > strategic and strategic > altruistic) using linear models in AES:SDM. For these comparisons, we entered comparator complexity as a covariate of no interest. Comparisons and overlap were also tested in the subgroup of studies with a selfish control.

Identification

Records identified through
Pubmed database searching
(n = 837)

Additional relevant records identified
through Web of Science
(n = 23)

Records after duplicates removed
(n = 720)

Additional relevant records
identified through reference lists
(n = 162)

Screening

Records excluded
(n = 395)

Titles relevant
(n = 325)

Eligibility

Full-text articles assessed for
eligibility
(n = 510)

Eligible papers identified for inclusion
(n = 80)

(altruistic = 35, strategic = 45)

Full-text articles
excluded
(n = 431)

Included

Studies with maps included
(n = 28)

(altruistic = 18, strategic = 10)

Studies with coordinates included
(n = 8)

(altruistic = 3, strategic = 5)

Reasons for unavailability of maps / coordinates:
(altruistic, strategic)
Same data as another paper included (n = 2, 2)
Data needed doesn't exist (n = 3, 2)
Author not able to send data (n = 7, 18)
No response from author (n = 0, 8)
Unable to align data to standard template (n = 2, 0)

Figure 2. Flow diagram of studies included and excluded at each stage of
identification and verification following PRISMA guidelines (Moher et al., 2009).

2.4.Labels and atlases

As AES:SDM uses a white matter template for labels, these were converted to grey matter labels with the Harvard-Oxford atlas in FSL (FMRIB Software Library, www.fmrib.ox.ac.uk/fsl). We also labelled regions according to labels used in relevant literature. The temporoparietal junction (TPJ), posterior superior temporal sulcus (pSTS), dorsolateral prefrontal cortex (dlPFC), ventrolateral PFC (vlPFC) and dorsomedial PFC (dmPFC), coordinates were used from the *a priori* regions of interest constructed by Telzer et al*.* (2011). However, we refer to their medial PFC here as the vmPFC and extend this region further ventral and posterior compared to the boundaries of that paper for our ROIs (see below). We also used connectivity-based parcellation atlases in FSL for further subdivision of the dorsal PFC (Sallet et al., 2013) and medial PFC and cingulate (Neubert et al., 2015).

2.5.ROI Analysis

To test the prediction that different types of prosocial decision may show different patterns of activation across the heterogenous vmPFC, values were extracted from across an anatomical axis. These regions were defined *a priori* by spheres used previously (Campbell-Meiklejohn et al., 2016) and similar to methods used in recent studies (De Martino et al., 2017; Nicolle et al., 2012; Sul et al., 2015; Yankouskaya et al., 2017). Effect sizes were extracted from single voxels along the same axis for our analysis (see details in Supplementary Materials S1.4. & Figure 8).

2.6.Accounting for dropout

A key region of interest, the vmPFC, is known to suffer from distortion and dropout during fMRI scanning due to factors including proximity to air and bone around the sinuses (Ojemann et al., 1997). While techniques have been developed to minimise this (Weiskopf et al., 2007) they are not universally employed and the inclusion of older studies in the meta-analysis meant that the coverage needed to be examined. This was done by binarising each map, after registration to a common template, based on whether there was signal in each voxel or not and summing these images to create coverage maps (Figure 3).

All studies

Studies with a selfish contrast

Altruistic

n = 21

n = 12

Strategic

n = 16

n = 12

n

0
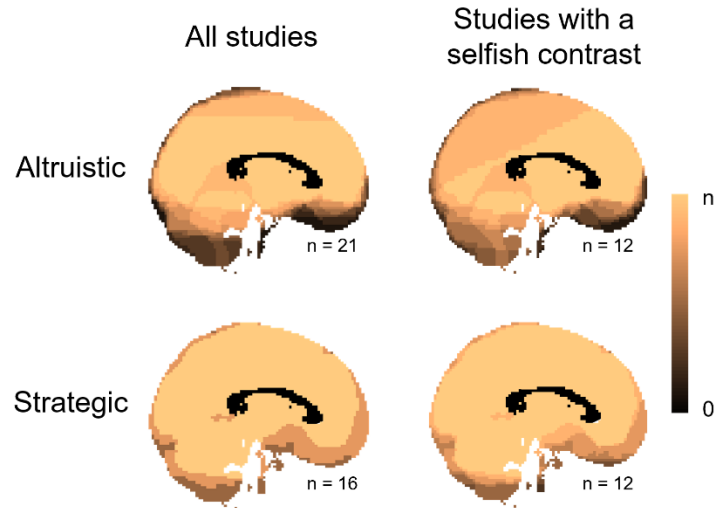
Figure 3. Coverage maps showing the number of studies with data in each region, x = 0, n = the number of studies available and the maximum possible coverage.

These maps show a decline in coverage around the anterior and inferior edge of the vmPFC, particularly for altruistic studies. With missing data represented as values of 0, this lack of coverage risked false negatives in reported results. To overcome this, we ran an *adjusted analysis* by modifying the calculations run by AES:SDM using SPM12 to only include studies with data present, on a voxel by voxel basis (see details in Supplementary Materials S1.5.). We did this only on the subgroup of studies with a contrast with selfish decisions.

Table 2.

Details of studies included in the meta-analysis

First author	Date	tN	F	N	Map or peaks	T	FWHM (mm)	Program	Sig	Contrast selected	Definition & additional info	Control / code
The focus of the study						The method used in the study						
Altruistic group												
Hare	2010	22	22/22	22	Map	3T	8	SPM5	<.05 C	PM by amount given	Free trials only	S / 4
Social value computations						Accept or decline proposed (doubled) payoff to charity						
Izuma	2010	29	12/23	23	Map	3T	6	SPM5	<.001 U	Donation vs fixation	Only when not observed	R / 1
Effect of an observer						Accept or decline the same payoff each time to a charity						
Telzer	2011	25	13/25	25	Map	3T	8	SPM5	<.05 FDR	Costly donation vs fixation	Across ethnicity	R / 1
Whether self-control and mentalising activation when giving to family and if moderated by family obligation values						Accept or decline proposed payoff to self and family - costly / non-costly reward / donation						
Brosch	2011	19	11/19	19	Map	3T	8	SPM8	<.05 FWE	Costly donation vs control	Control = self and charity -1	V / 3
Self-interest and openness to change values						Accept or decline proposed payoff to self and charity - reward, (non)costly donation, control						
FeldmanHall	2012	20	8/14	14	Peaks	3T	8	SPM	<.001 U	PM by amount given (+) and amount kept (-)	Only in real decisions	S / 4
Real vs hypothetical moral decisions						Real and imagined pain vs gain task						
Morishima	2012	30	17/30	27	Map	/	8	SPM8	<.05 FWE	Donation vs selfish		S / 4
Whether altruism correlates with TPJ structure and function						Choose between 2 proposed payoffs for self and other - advantageous or disadvantageous inequality						
Gunther Moor	2012	53	31/53	49	Map	3T	8	SPM5	<.001 U	Generous vs fixation	Across ages (22 age 10-12, 16 age 14-16, 15 age 19-21)	R / 1
Developmental differences in dictator game after exclusion						Dictator allocations to people who included or excluded participants during previous cyberball game						
Lelieveld	2012	26	17/26	18	Map	3T	6	SPM5	<.001 U	Prosocial decision vs selfish	Data doesn't separate response from decision	S / 4
Effect of receiver's emotional reaction						Participants who split 6-4 in dictator game split again after emotional feedback from receivers						
Genevsky	2013	30	14/22	22	Peaks	3T	4	AFNI	<.05 C	Donation vs not donating	Across conditions	S / 4
Neural basis of identifiable victim effect						Accept or reject requested donation to children with varying information (name, photo) available						

Kuss	2013	33	17/33	33	Map	3T	6	SPM8	<.001 U	Costly donation vs pure self-interest	Control: self-interest choices	V / 3
Motivation for donating using reward prediction error						Choose between 2 proposed payoffs for self and other – (non)costly donations, self-interest or efficiency						
Guroglu	2014	28	17/28	22	Map	3T	8	SPM8	<.001 U	Generous vs selfish		S / 4
Activations linked to inequity, regardless of benefactor						Choose between 2 proposed payoffs for self and other – fair or unequal benefitting either self or other						
Pulcu	2014	15	9/15	14	Map	3T	5	SPM8	<.05 FWE	Generous vs control	Healthy controls only, control: no gains	V / 3
Altruism activity in remitted major depressive disorder						Accept or decline proposed payoff to self and charity – (non)costly donations, both gain, P gains, no gains						
Koban	2014	22	10/22	17	Map	3T	8	SPM8	<.001 U	Sharing vs keeping	Human partner, conflict trials	S / 4
Activity related to interpersonal conflict for resources						Participant to share or keep resources when they and another player chose the same token (conflict)						
FeldmanHall	2015	19	11/17	17	Peaks	3T	8	SPM5	<.001 U	PM by amount kept (-)	With no covariates	S / 4
Personal distress vs empathic concern explanations						Pain vs gain task						
Hutcherson	2015	61	0/61	51	Map	3T	8	SPM5	<.05 FWE	Prosocial decision vs selfish	Decisions always created a selfish and a prosocial choice	S / 4
Propose and test a model of altruistic choice						Accept and rate 1-4 (strong no, no, yes, strong yes) split between self and other, 40% decisions reversed						
Strombach	2015	27	13/27	23	Map	3T	8	SPM8	<.005 U	Generous vs selfish	Across social distances	S / 4
Effect of social distance on prosocial decision activity						Choose between 2 proposed payoffs for self and other (diff social distances) - one selfish and one generous						
Telzer	2015	29	13/29	26	Map	3T	8	SPM8	<.05 FWE	Generous vs fixation	Across ethnicity and recipients	R / 1
Cultural differences in ingroup and outgroup helping						Choose between 2 proposals for self or other - ingroup donations, outgroup donations and pure reward						
Will	2015	28	16/26	26	Map	3T	8	SPM5	<.001 U	Generous vs fixation	Across includes and excluders	R / 1
Correlates of punishment and forgiveness after exclusion						Dictator allocations to people who included or excluded participants during previous cyberball game						
Kuss	2015	40	22/40	36	Map	1.5 T		SPM	<.05 FWE	Costly donation vs control	Across prosocials and proselfs, control: both benefit	V / 3
Impact of social value orientation individual differences						Choose between 2 proposed payoffs for self and other – (non)costly donations, self-interest or efficiency						
Tusche	2016	33	15/33	32	Map	3T	8	SPM8	<.05 FWE	High vs low donations	Scale $0-$50, 'selfish' = lowest ½ donations	S / 4
Differences in neural patterns for different giving motives						Decide on a scale how much to donate to a charity						
Will	2016	46	28/46	46	Map	3T	8	SPM8	<.001 U	Costly prosocial vs selfish	Across stable and rejected	S / 4
Sharing activity in chronically rejected and stable adolescents						Equitable or inequitable payoff for self and other in costly, non-costly and envious conditions						

Strategic group

Author	Year	N			Peaks/Map	Field		Software	Threshold	Contrast	Notes	Code
Delgado	2005	14	6/14	12	Peaks	3T	4	Brain Voyager	<.001 U	Trust vs defect	Across reputations	S / 4
Effect of responder's responses and reputation						Repeated trust game after reading description of responder and getting feedback on their decisions						
Krueger	2007	44	22/44	44	Peaks	3T	8	Brain Voyager	<.005 U	Trust vs control	Control similar to trust game	V / 3
Neural mechanisms of building trust						Alternating non-anonymous multi-round trust game						
Sharp	2011	10	0/10	10	Map	3T	4	AFNI	<.001 U	Trust vs keep / rest	Healthy controls only, across partners	S&R / 4&1
Reputation and reward processing in young boys with externalising behaviour problems						Trust game with peers rated aggressive, kind or neutral						
Fareri	2012	25	9/18	18	Map	3T	8	Brain Voyager	<.001 U	Cooperate vs control / one sample	Used data from both participants as alternated roles, control: lottery	V&R / 3&1
Effect of being excluded						Trust game after being excluded or not by different players						
Stanley	2012	51	22/40	40	Peaks	3T	6	SPM8	<.05 C	PM by amount sent	Human only, across exclusion	S / 4
Effect of race as a reputation proxy on trust						Trust game with people of different races						
Fouragnan	2013	20	0/20	18	Peaks	4T	8	SPM8	<.05 FWE	Trust vs keep	Across partners	S / 4
Effect of reputation on response to trust violations						Trust game with cooperative and individualistic partners with a reputation for being one or the other						
Smith-Collins	2013	27	27/27	24	Peaks	1.5 T	8	SPM8	<.001 U	Cooperate vs control / defect	Across races, control: remembering faces	S&V / 4&2
How brain updates representations of partners to learn						Two round trust game with the same players, receiving feedback on trustworthiness after first						
Bault	2014	29	12/29	25	Map	3T	5	FSL	<.05 FWE	PM by amount sent	Across social distances	S / 4
Dynamic model of social tie development						Repeated public goods game with the same partner						
Emonds	2014	38	22/38	35	Map	3T	4	Brain Voyager	<.01 U	Cooperate vs defect	Across ethnicity and group given to	S / 4
Differences between prosocials and proselfs in neural activity and impact of trust						Prisoner's dilemma game with measure of SVO and dispositional trust						

Author	Year	tN	F	N		T	FWHM	Software	sig	Contrast	Notes	Control
Garbarini	2014	16	8/16	16	Map	1.5 T	4	Brain Voyager	<.05 C	Giving back vs not / rest	Used trustee data due to reputational manipulation	S&R / 4&1
Gender differences in reputation building						Act as trustees in trust game with either reputation being monitored or not						
Fett	2014	45	0/45	45	Map	3T	8.8	XBAM		Trust vs control	Control: put cursor to target	V / 2
Modulation of mentalising by attachment style						Prisoner's dilemma game with measures of attachment style						
Ramsøy	2015	14	11/14	14	Peaks	3T	8	SPM8	<.05 FWE	Cooperate vs defect	Across frames	S / 4
Role of empathy as a heuristic in decision making						Iterated prisoner's dilemma game framed as competition or cooperation with trait empathy measures						
Schneider-Hassloff	2015	164	78/164	164	Map	3T	8	SPM8	<.001 U	Cooperate vs control / defect	Control: "#" appeared in matrix – no points or partner	S&V / 4&2
Modulation of mentalising by attachment style						Prisoner's dilemma game with measures of attachment style						
Chen	2016	104	50/104	93	Map	3T	5	FSL	<.05 FWE	Cooperate vs defect	Only placebo group with humans	S / 4
Oxytocin & vasopressin effects during cooperation						Iterated prisoner's dilemma following administration with human or computer partner (total N = 304)						
Fermin	2016	41	18/33	33	Map	3T	8	SPM8	<.005 U	Cooperate vs one sample / defect	Across prosocials and proselfs	S&R / 4&1
Differences between prosocials and proselfs in structure and function						Prisoner's dilemma with measures of SVO						

Note. tN = total sample in study, N = sample used in analysis, F = number of female participants as a fraction of tN or N depending on information provided, T = field strength of scanner in tesla, FWHM = full width half maximum size of smoothing kernel, sig = significance threshold used in study (this is relevant for studies with peaks but provided for those with maps to show peak thresholds are in line with other research), U = uncorrected for multiple comparisons, C = corrected for multiple comparisons, FWE = with family-wise error correction, FDR = with false discovery rate correction, PM = parametric modulation. Control: R = rest (including one sample t-tests), V = visuomotor control, S = selfish decision (including parametric modulation by generosity), Code = code for complexity of the control condition from 1 (minimum) to 4 (maximum).

3. Results

3.1. Mean analyses

3.1.1. Altruistic

Compared to being selfish, making a prosocial choice in an altruistic context showed significant activation in nucleus accumbens (NuAcc), subgenual (subcallosal) area of the anterior cingulate cortex (sgACC), vmPFC, orbitofrontal cortex (OFC), left dlPFC (Sallet et al., 2013 area 8B), pre-supplementary motor areas (pre-SMA), posterior cingulate cortex (PCC) and right cerebellum. Regions showing greater activation during selfish than altruistic decisions included bilateral dlPFC (areas 46 & 9), bilateral putamen, right caudate nucleus, bilateral posterior STS, bilateral frontal poles, and left amygdala (Figure 4a & Table 3).

Comparisons to 'rest' tell us how the brain is responding generally, in the decision context. These maps can be useful for making future predictions, but also aid the interpretation of the other contrasts. For instance, a region being 'more active' in one condition compared to another may counterintuitively actually reflect relatively less deactivation in that condition, relative to a common baseline.

Figure 4. Mean activations from altruistic meta-analytic maps from modelled contrasts using the complexity covariate modelled **A:** at maximum – selfish control and **B:** at minimum – rest control (thresholded with permutation analysis run in AES:SDM). Coronal images in radiological orientation (right = left).

Table 3.

Peak activations from modelled contrasts using complexity coordinate for altruistic vs. selfish

Peak label	BA*	coordinates			SDM-Z	Vox
		x	y	z		
Altruistic > selfish						
L striatum	25	-4	6	-12	3.38	1798
R orbitofrontal cortex	11	22	16	-20	2.81	360
R inferior temporal gyrus	20	44	-10	-36	2.37	158
L orbitofrontal cortex	11	-22	16	-22	2.61	97
Posterior cingulate gyrus		0	-38	2	2.39	92
L frontal operculum cortex	47	-40	28	4	1.96	67
L middle frontal gyrus	44	-54	24	30	2.04	65
L superior frontal gyrus		-16	34	44	1.72	46
L inferior temporal gyrus	37	-56	-56	-14	2.23	38
R parahippocampal gyrus		24	-14	-34	1.75	17
R cerebellum, crus II		46	-68	-42	1.67	16
R middle temporal gyrus	21	60	0	-18	1.67	13
L orbitofrontal cortex		-34	32	-8	1.85	12
R cerebellum, hemispheric lobule VI	37	34	-38	-34	1.71	12
R inferior temporal gyrus	20	60	-28	-24	1.85	10
Selfish > altruistic						
L supramarginal gyrus	22	-52	-46	14	3.38	526
R middle temporal gyrus		46	-34	-4	3.75	374
L temporal occipital fusiform cortex	19	-30	-62	-6	2.96	332
L frontal pole		-22	44	24	2.76	225
R putamen		26	-2	6	2.58	73
R caudate		18	-16	26	2.95	66
L putamen		-26	2	4	2.58	60
R inferior lateral occipital cortex		34	-72	12	3.08	57
R inferior frontal gyrus	45	50	34	6	2.53	29
L amygdala		-26	-10	-16	2.41	19
L superior lateral occipital cortex		-20	-78	18	2.43	17
R frontal pole		26	38	22	2.23	15
L precentral gyrus	6	-28	-12	54	2.32	11
L lingual gyrus		-26	-54	4	2.25	11
L precentral gyrus		-40	-10	40	2.37	10
L precentral gyrus		-50	-6	50	2.32	10
R precentral gyrus	6	54	0	46	2.41	10
R frontal pole	47	48	38	-6	2.17	10

Note. L = left, R = right, BA = Brodmann area, coordinates (x, y, z) in MNI space, SDM-Z= SDM z-value of activation, Vox = number of voxels in the cluster. *If cluster falls within *more* than one BA, none is reported either by AES:SDM software, or here.

Altruistic decisions contrasted with rest showed significant activation in ACC (Neubert et al. 2015 area 8m), right anterior insula (AI), bilateral dlPFC (area 46V), SMA and occipital cortex. Areas less active during the decision were vmPFC extending into left dmPFC, posterior insula, left precuneus, separate dlPFC regions (areas 8B & 46D), bilateral vlPFC, and temporal sulci, including TPJ and the pSTS in both hemispheres (Figure 4b & Supplementary Table 1).

3.1.2. Strategic

Strategic prosocial decisions, compared to selfish decisions, related to significantly higher activation in bilateral NuAcc, sgACC, vmPFC, ACC, right precuneus, right amygdala and regions of the cerebellum. Selfish strategic choices related to more activity in left TPJ, anterior middle temporal regions, right temporal pole and pre and postcentral gyri (Figure 5a & Table 4).
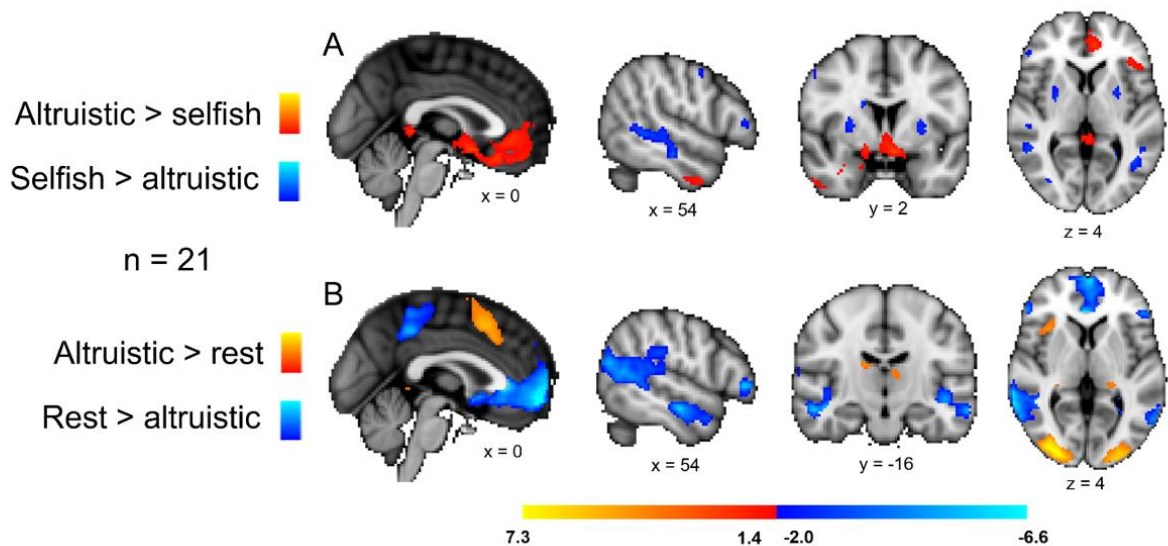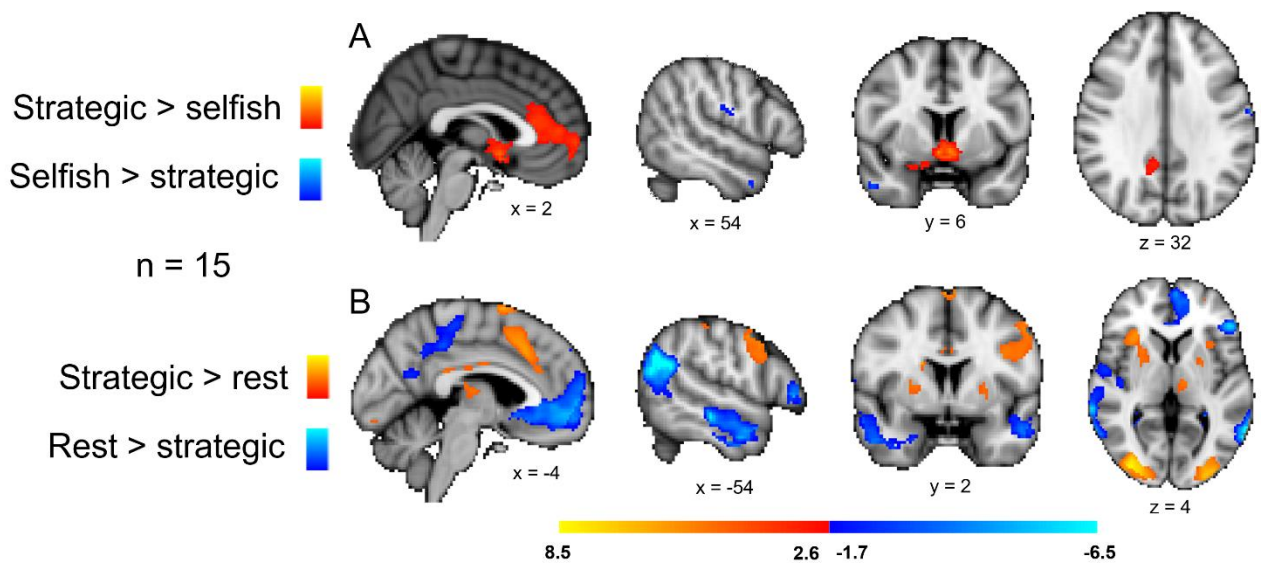

Figure 5. Mean activations from strategic meta-analytic maps from modelled contrasts using the complexity covariate modelled **A:** at maximum – selfish control and **B:** at minimum – rest control (thresholded with permutation analysis run in AES:SDM). Coronal images in radiological orientation (right = left).

Table 4.

Peak activations from modelled contrasts using complexity coordinate for strategic vs. selfish

Peak label	BA*	coordinates			SDM-Z	Vox
		x	y	z		
Strategic > selfish						
Paracingulate / anterior cingulate cortex		14	48	10	4.14	1017
L striatum		-6	2	-12	5.47	659
R precuneus cortex		12	-54	34	3.39	82
Cerebellum, vermic lobule VIII		18	-58	-42	4.45	26
R orbitofrontal cortex		22	28	-10	3.73	26
R amygdala		18	4	-18	3.11	26
Middle cerebellar peduncles		28	-46	-38	3.32	14
R superior lateral occipital cortex	39	42	-68	42	2.91	15
R cerebellum, crus I		28	-86	-28	3.19	11
L cerebellum, hemispheric lobule VIII		-14	-60	-44	3.12	11
R cerebellum, hemispheric lobule IX		6	-54	-46	2.96	10
Selfish > strategic						
R postcentral gyrus		54	-10	20	2.07	65
R temporal pole	21	52	6	-32	2.65	44
R brainstem		16	-28	18	3.03	41
R precentral gyrus		8	-20	56	2.09	36
L angular gyrus	39	-40	-56	22	2.63	32
R precentral gyrus		18	-12	60	2.51	29
L postcentral gyrus	43	-62	-8	32	2.47	13
L middle temporal gyrus	20	-58	-14	-28	2.06	12
R parietal operculum cortex		34	-22	20	1.84	10

Note. L = left, R = right, BA = Brodmann area, coordinates (x, y, z) in MNI space, SDM-Z= SDM z-value of activation, Vox = number of voxels in the cluster. *If cluster falls within *more* than one BA, none is reported either by AES:SDM software, or here.

Compared to rest, strategic prosocial choices evoked similar results to altruistic prosocial choices in the equivalent analysis: activation in ACC (area 8m), bilateral AI, widespread bilateral dlPFC regions, SMA and occipital cortex. In addition, strategic decisions showed activation in bilateral putamen and the right caudate. Deactivation was across vmPFC, left dmPFC, bilateral posterior insula, bilateral precuneus, left dlPFC (area 8B), left vlPFC, bilateral TPJ and posterior & anterior STS regions (Figure 5b & Supplementary Table 2).

3.2. Overlap

Overlaps were analysed using SPM to identify regions significant in both altruistic and strategic maps for a given contrast. These maps use the complexity covariate to model contrasts with rest and selfish decisions while incorporating all studies.

Compared to the selfish alternative, both types of prosocial choice showed overlapping activity in left NuAcc, ACC (sgACC and area 32), vmPFC and right OFC (Figure 6a). There was no overlap of areas responding more to the selfish choice (selfish > prosocial).
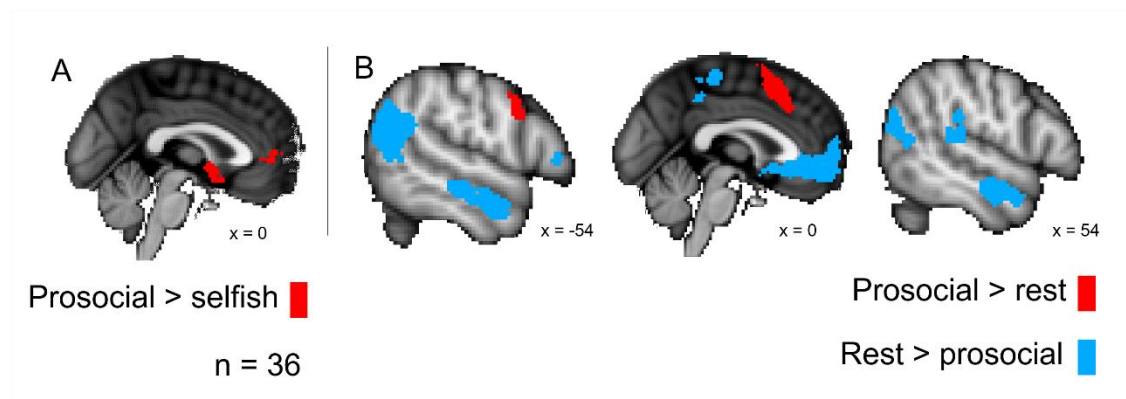
Figure 6. Overlap between altruistic and strategic decisions maps from modelled contrasts using the complexity covariate **A:** *vs.* selfish decisions and **B:** *vs.* rest.

When contrasted with rest, altruistic and strategic decisions both activated bilateral ACC (area 8m), right AI, bilateral thalamus, bilateral SMA, pre-SMA, bilateral occipital and parietal cortices. Altruistic and strategic decisions showed overlapping deactivations in vmPFC, right temporal pole, left precuneus, left dlPFC (area 8B), right posterior insula, left OFC and bilateral TPJ & posterior STS (Figure 6b).

3.3. Comparison

Areas more active during altruistic prosocial choices than strategic prosocial choices were the sgACC, left TPJ, left anterior STS, left inferior frontal gyrus, right temporal pole, right inferior temporal gyrus and bilateral thalamus. Strategic prosocial choices, compared to altruistic prosocial choices, showed more activity in the right NuAcc, left caudate, right dlPFC (area 46D), right posterior STS and right frontal pole (Figure 7 & Table 5).
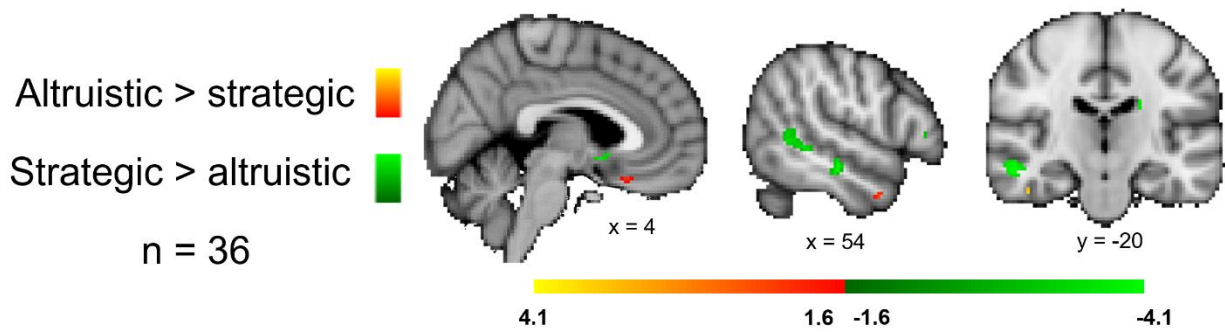
Figure 7. Comparisons between altruistic and strategic decisions controlling for complexity (thresholded with permutation analysis run in AES:SDM). Coronal image in radiological orientation (right = left).

Table 5.

Regions showing significantly greater activation in altruistic or strategic studies.

Peak label	BA*	coordinates			SDM-Z	Vox
		x	y	z		
Altruistic > strategic						
R temporal pole	21	52	6	-32	2.45	27
L angular gyrus	39	-40	-54	22	2.55	26
L inferior frontal gyrus		-40	32	2	2.40	21
L thalamus		-14	-28	18	2.79	14
Subcallosal cortex	11	4	22	-20	2.10	15
R thalamus		16	-28	18	2.37	12
L middle temporal gyrus	21	-66	-30	-4	2.25	12
R inferior temporal gyrus	20	46	-18	-30	2.57	10
Strategic > altruistic						
R middle temporal gyrus		52	-38	-4	3.98	110
R inferior frontal gyrus	45	50	34	6	3.42	76
R middle temporal gyrus		52	-22	-16	4.23	49
R striatum	25	4	6	-8	3.72	13
R frontal pole	46	30	44	34	3.17	13
L caudate		-20	-18	24	3.56	10

Note. L = left, R = right, BA = Brodmann area, coordinates (x, y, z) in MNI space, SDM-Z= SDM z-value of activation, Vox = number of voxels in the cluster. *If cluster falls within *more* than one BA, none is reported either by AES:SDM software, or here. Activations are after controlling for the complexity of the control task.

3.4. ROI analysis

Previous findings have shown that trajectories across the vmPFC can delineate decision types. Values were examined at the same 20 y and z coordinates (see Supplementary Figure 1) for x = 0, x= 4 and x = -4. Results reported here are from x = 4 as this plane showed the most striking differentiation but results for x = 0 and x = -4 were similar (see Supplementary Materials S2.2. and Supplementary Figure 5).

We ran linear models on each study's extracted effect sizes for each voxel using (i) all the data, including zero values and (ii) only the studies with data in that voxel, on a voxel-by-voxel basis. In both cases, a mixed-effects model was a better fit than a fixed-effects model.

Both analyses revealed an interaction between vmPFC voxel location and decision group, (i) $t(22)=3.11$, $p=.005$; (ii) $t(21.1)=3.31$, $p=.003$. Altruistic activation was greatest in posterior vmPFC, as suggested by the whole brain results, and decreased moving anterior, whilst strategic activation increased along this axis (Figure 8).
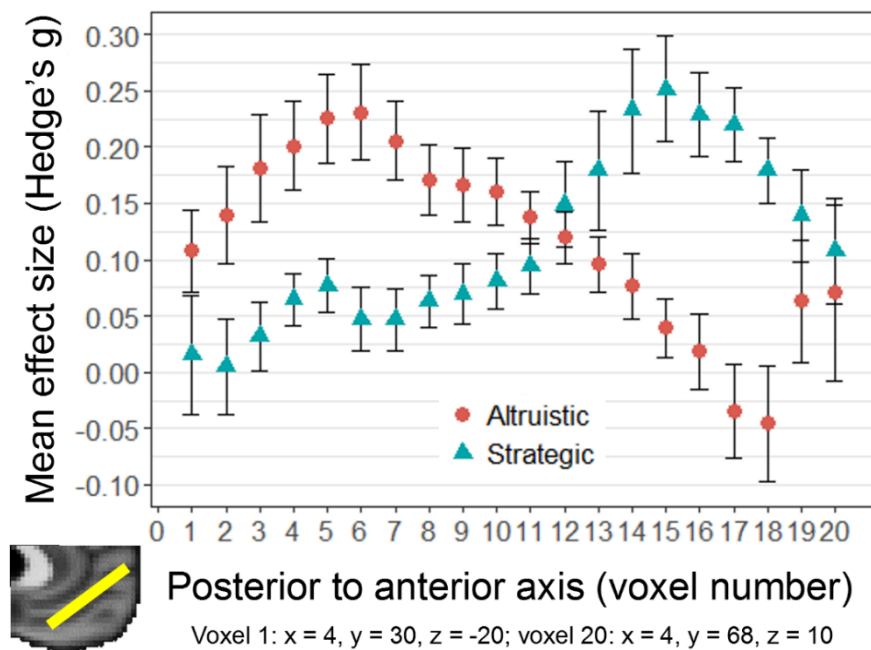
Figure 8. vmPFC ROI analysis: average effect sizes of activation for each group for each voxel at x = 4, not including studies with no data in that voxel (see Supplementary Figure 1 for exact location of voxels and Supplementary Figure 5 for graphs including all data). Error bars depict standard error.

3.5. Adjustment for dropout

As predicted, the analysis adjusting for missing data (see Supplementary Materials S1.5. for details) showed increased effect sizes across the lower vmPFC where dropout was most severe (Figure 3). Activations based on SDM-Z > 2.3 uncorrected were larger in the adjusted than the original analysis for the altruistic mean activations (Figure 9a). This threshold was chosen as a common value for thresholding, close to the average of the critical SDM-Z values generated in the permutation tests for the original analyses and AES:SDM analyses run with the 50% of maps with the best coverage.

For the comparison where altruistic > strategic prosocial choices, posterior vmPFC activations were either larger than the original analysis or shown only in the adjusted analysis (Figure 9b). In the original comparison where strategic > altruistic prosocial choices, no vmPFC activation was significant when thresholded using AES:SDM permutation analysis and very little showed SDM-Z > 2.3. However, when adjusting for dropout, a small region of activation in anterior vmPFC shows SDM-Z scores greater than 2.3 (Figure 9b).
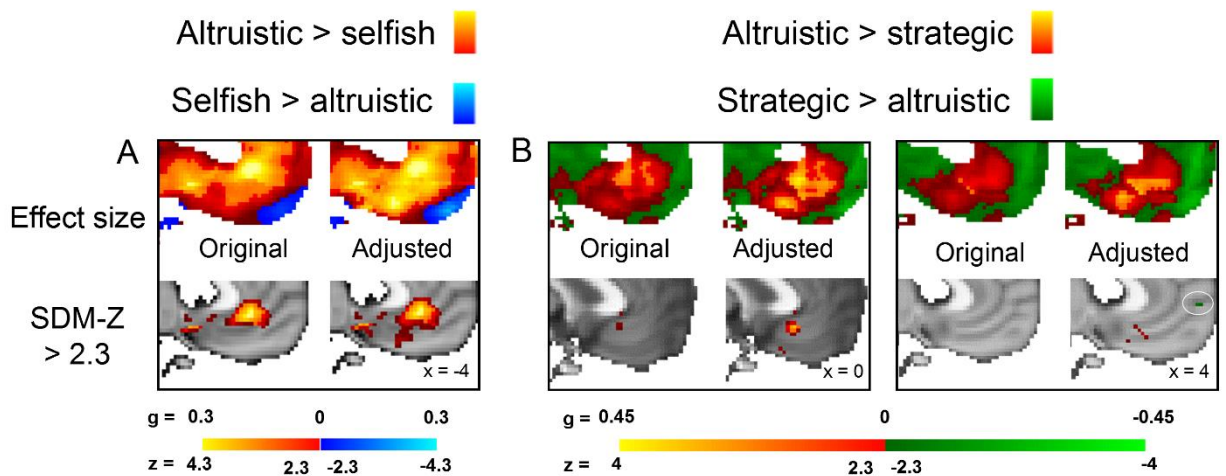
Figure 9. Adjusted analysis accounting for vmPFC dropout. Increased effect sizes (Hedges' G; top row) and larger regions showing SDM-Z > 2.3 (bottom row) in the adjusted analysis compared to the original. **A:** altruistic contrasted with selfish decisions (n = 12), **B:** comparison between altruistic and strategic decisions (n = 24). All results are from the subgroup of studies with a selfish contrast.

4. Discussion

This meta-analysis shows there is consistency in neural processes during decisions to give, using over a decade of fMRI research and over one thousand participants. Within each type of decision to give (altruistic and strategic), activations are present across a variety of tasks, suggesting common core processes. There are also differences between the two types of decisions to give, which are thought to reflect the differences in the source of reward, timing, certainty, and application of theory of mind.

We structure our discussion around the basic questions for the meta-analysis, though some discussion of contrasts inevitably crosses into more than one section. We focus on established social cognition and reward networks, given these are the most common networks discussed in the independent studies, but also highlight activations outside of this familiar territory.

All thresholded and unthresholded statistical brain maps are available at https://neurovault.org/collections/3987 and details of peaks (as well as maps) can be downloaded from https://doi.org/10.25377/sussex.c.4155923. Since a discussion of each individual activation is not feasible in this discussion, we encourage the reader to use these maps, together with provided tables, for their reference, inference, and convenience in future research.

4.1. What is common to altruistic and strategic decisions to give?

Sharing with others in either context consistently activates regions associated with reward. Compared to selfish decisions, NuAcc, vmPFC, OFC sgACC and ACC (area 32) are more active when an individual decides to give. These regions are established elements of the reward and value-computation networks (Bartra et al., 2013). A NuAcc activation, for instance, gives a moderately strong likelihood that a reward process is occurring (Ariely and Berns, 2010) and responds to multiple reward types (Levy and Glimcher, 2012) including money, intrinsic reward (Genevsky et al., 2013; Harbaugh et al., 2007; Moll et al., 2006) and strategic cooperation (Rilling et al., 2002).

These signals may represent different concepts in the two contexts. While intrinsic reward may accompany strategic prosocial choices, these activations could also signal anticipation of an extrinsic benefit. This would suggest intrinsic altruistic motivation shares neural circuits with extrinsic motivation, supporting the common neural currency proposal (Levy and Glimcher, 2012) for multiple aspects of prosocial decision-making. Speculatively, this overlap may also explain behavioural findings of decreased altruistic behaviour, which

was once freely given, following introduction of extrinsic incentives, by 'crowding out' the intrinsic motivations represented within the same neural circuits (Frey and Oberholzer-Gee, 1997). Such a conclusion however, would need direct experimentation.

Being selfish, relative to being prosocial, does not overlap much between strategic and altruistic contexts, save for a small region of the parietal operculum. This suggests the underlying processes and expectations are likely to be rather different. In altruistic contexts, the contrast may represent a preference for self over others or loss aversion (Tom et al., 2007). In strategic contexts, the choice is more likely driven by predictions of others' behaviour and the desire to reduce uncertainty.

As expected for any contrast with rest, altruistic and strategic generosity show overlapping widespread deactivation of the 'default mode' network and activation of the 'task-activated' network (Fox et al., 2005). These activations, such as in ACC and AI, could reflect domain-general processes such as deciding between multiple options (Kennerley et al., 2009). They could also reflect more specific processes common to both contexts, for example negative aspects of the prosocial choice including disutility (Zaki and Mitchell, 2011), distress (Zanon et al., 2014), and conflict (Koban et al., 2014). Finally, these activations could represent different processes in each context. For altruistic giving compared to rest, AI and ACC activity could underlie empathy (FeldmanHall et al., 2012a; Gospic et al., 2013; Hein et al., 2010). More relevant for risky strategic giving (Apps et al., 2016), ACC activation could reflect greater ambiguity or risk (Krain et al., 2006) or an increase in effort and prediction (Vassena et al., 2017).

Regions less active than baseline during both types of decision to give include vmPFC, TPJ and STS which are implicated in both social decision-making (Decety and Lamm, 2007; Frith and Frith, 2003; Schurz et al., 2014) and resting-state networks (Damoiseaux et al., 2006; Mak et al., 2017). This fits with previous demonstrations of quantitative overlap (Acikalin et al., 2017) and discussion on what these processes may share (Mars et al., 2012). Critically, these effects highlight the importance of the control task when making inferences; the 'activation' associated with a prosocial decision might actually be *less deactivation* which only becomes evident with a comparison to rest.

4.2. What is consistent about altruistic decisions?

During altruistic decisions, the consistent activation of reward networks demonstrates that despite ending up with less extrinsic gain, the reward system is active when giving.

In addition to regions common to both types of choices to give, generous altruistic decisions compared to selfish choices activate other regions, including left dlPFC (area 8B), posterior cingulate cortex (PCC), pre-SMA and right cerebellum. In an analysis restricted to studies with a selfish choice contrast, we also found bilateral precuneus. See below for discussion of these regions.

The reverse contrast identifies additional activation for selfish choices over altruistic ones in putamen, caudate, pSTS, dlPFC (areas 46 & 9), left amygdala and, in the subgroup with selfish contrasts, bilateral posterior insula. The putamen, caudate and amygdalae have all been linked to aspects of reward processing (Haruno and Kawato, 2006; Holland and Gallagher, 2004; but see also Miller et al., 2014) and have high levels of interconnectivity (Roy et al., 2009). These parts of the reward network may be less sensitive to intrinsic rewards and more sensitive to extrinsic benefits of keeping the goods.

There have been claims the dlPFC inhibits selfish tendencies to produce prosocial decisions (Feng et al., 2015; Knoch et al., 2009, 2006; Strang et al., 2014) but also findings supporting the reverse, that the inhibition is of prosocial impulses (Christov-Moore et al., 2016; Yamagishi et al., 2016). Our finding in area 8B corresponds with the former model but our pattern of results for areas 46 and 9 fits with the latter. This anatomical differentiation could explain the differences in previous conclusions if, for example, these stimulation studies targeted different subregions. Other theories (Buckholtz, 2015; Gershman et al., 2014) suggest the lPFC could be integrating norms into decisions – disrupting right lPFC affects adherence to rules in gift-giving contexts (Gross et al., 2018).

Altogether, these results provide the signature activations of altruistic giving decisions to use in future investigations of their connectivity and more specific roles in prosocial behaviour.

4.3. What is consistent about strategic decisions?

During strategic decisions to give, reward circuit activation could represent either extrinsic reward expectancy, which is usually higher for the prosocial choice contingent on the partner's behaviour, or intrinsic satisfaction. The signature of a generous strategic decision also includes activation of precuneus, right amygdala and regions of the cerebellum. Analysis limited to studies with the selfish contrast also shows activity in right dlPFC (area 46), left amygdala and the right frontal pole.

The precuneus shows similar, but not overlapping, activation for both altruistic (bilateral) and strategic (on the left) gifts compared to their selfish alternatives. Along with

the PCC being more active for altruistic than selfish decisions, these regions linked to mentalising (Cavanna and Trimble, 2006; Schurz et al., 2014; Waytz and Mitchell, 2011), empathy (Jackson et al., 2006) and guilt (Morey et al., 2012) did show results in line with increased perspective taking during prosocial choices.

Activation in the amygdala for strategic decisions fits with our interpretation of reward expectancy during selfishness in altruistic paradigms. In strategic tasks it could also be interpreted as encoding estimated immediate cost of the gift (Gospic et al., 2013). Previous studies have shown activity scales with the amount entrusted to another (Stanley et al., 2012) and is greater in real than imagined decisions (FeldmanHall et al., 2012a; Gospic et al., 2013). Outside of prosocial decisions, the amygdala is associated with many other relevant processes, including loss aversion (De Martino et al., 2010), fear (Tovote et al., 2015), aggression (Haller, 2018) and emotion processing (Janak and Tye, 2015) to name but a few. Due to the low resolution of a meta-analysis from averaging images and the broad scope of the tasks used, this study does not differentiate between specific explanations.

The consistent neural signature of keeping money in strategic contexts includes increased activity in left TPJ, anterior middle temporal regions, right temporal pole and pre and postcentral gyri. The specific analysis with selfish controls also showed activity in bilateral posterior temporal sulci, left temporal pole, left hippocampus, right supplementary motor area, right cerebellum and right posterior insula. If TPJ and pSTS activations represent mentalising in this context, these results suggest increased processing about the second player is associated with *not* trusting, valuing, or cooperating with them. For altruistic decisions, this was also the pattern for pSTS.

These findings overall show the consistent pattern of activations during strategic giving for extrinsic rewards. Like for altruistic decisions, future work can use this to reveal the exact roles and connectivity of regions.

4.4. How do altruistic and strategic decisions (statistically) differ?

Contrasts between altruistic and strategic choices make a clear case that that these groups of tasks rely on different processes and should not be considered interchangeable in the literature.

4.4.1. Altruistic > Strategic

Altruistic choices to give correlate with greater activation in a set of reward and social cognition regions. These include sgACC, TPJ and various regions of the temporal lobes. Limiting to selfish contrasts additionally revealed right pSTS. That *any* regions show greater

activity during altruistic choices challenges the idea that strategic decisions encompass all the elements of altruistic decisions.

The sgACC, which is involved in both altruistic and strategic prosocial behaviours independently, is particularly active when only intrinsic motivations are available. This region also activates during charitable donations (Moll et al., 2006), distinguishes altruism from decisions which benefit the individual (Pulcu et al., 2014) and signals prosocial learning prediction errors (Lockwood et al., 2016) as well as emotional processing in social contexts (Drevets et al., 2008). Activity in sgACC is also linked to a reduced propensity to harm others in utilitarian judgements (Wiech et al., 2013). The current meta-analytic finding adds weight to the idea that sgACC, together with a network of social cognition regions, may play a defining role in altruistic decisions to give.

4.4.2. Strategic > altruistic

Strategic decisions correlate with more activity in ventral and dorsal striatum, right dlPFC (areas 46 & 46D), right pSTS and right frontal pole. Isolating studies with a selfish contrast additionally showed posterior insula and fusiform face area activity.

Greater right NuAcc activity for strategic than altruistic choices suggests that intrinsic and extrinsic responses sum in this region, or extrinsic rewards (if supplanting intrinsic rewards) evoke stronger responses. This finding fits with behavioural findings that prosocial choices are more frequent in strategic than altruistic paradigms (Zheng and Zhu, 2013).

The left caudate also activates more for strategic than altruistic decisions, a pattern previously limited to the putamen (Weiland et al., 2012). A closer look suggests that differences in dorsal striatum between the task groups may be partly driven by greater activity for the selfish choice in altruistic contexts (a difference absent in strategic contexts). That may be due to the increase in extrinsic reward obtained by being selfish in an altruistic task whereas in strategic scenarios, selfishness often leads to a more certain but lower payment than optimum cooperation. The dorsal striatum has also been linked to habitual responses (Balleine et al., 2007; Burton et al., 2015; Everitt and Robbins, 2016) which could suggest that selfish decisions in altruistic contexts are more habitual and less goal-directed than strategic decisions to keep resources (and less sensitive to devaluation).

As expected, inferences from mentalising-associated regions are not straightforward. Compared to strategic gifts, altruistic gifts recruit more left TPJ and left anterior STS activity. Overall, when using all studies, strategic activity in right pSTS is greater than altruistic, potentially reflecting greater consideration of others' intentions.

However, when limited to comparisons with a selfish choice, the altruistic choice to give shows greater activity in this region. This discrepancy may result from relatively high but *similar* levels of mentalising activity prior to either response in strategic contexts. In contrast, the differential between gifts and selfish choices is higher in altruistic contexts – gifts likely involve more other-oriented processing than selfish choices (e.g. representing the other's need or appreciation).

Outside of mentalising, previous findings show TPJ activation and connectivity also peak at maximum conflict, near the maximum one is willing to give and when temptation to be selfish is greatest (Morishima et al., 2012; Strombach et al., 2015). This conflict is likely to be highest in an altruistic context, when there is no opportunity for financial gain. Overall, these social brain regions may have different roles which are recruited to varying extents depending on the context of task and what control is used. More work is required to understand precisely what they are doing in each context.

The results for right dlPFC areas 46 and 46D are in line with findings that levels of strategic behaviour correlate with differences between ultimatum and dictator game activation and cortical thickness (on the left) in dlPFC (Steinbeis et al., 2012). Moreover, right dlPFC transcranial direct current stimulation had opposing effects on giving behaviour under altruistic and strategic conditions (Ruff et al., 2013), suggesting the exact role of the region depends on the context.

4.4.3. The vmPFC

Analysis across a ROI axis in the vmPFC, particularly when adjusted for signal dropout, suggests altruistic activation *declines* moving anteriorly in the brain as strategic activation *increases*. Along the same axis, regions of vmPFC show differing connectivity patterns (Gallardo et al., 2017): posterior areas connect locally and to limbic regions while anterior areas are grouped with distant mentalising regions (Alcalá-López et al., 2017). Anterior areas are also directly involved in mentalising under conditions of ambiguity, a key feature of strategic tasks (Jenkins and Mitchell, 2010). A similar axis identifies the influence of certain social inferences on value judgements (Campbell-Meiklejohn et al., 2016; De Martino et al., 2017). Contrasts of this meta-analysis reflect within-participant comparisons. Looking between participants, selfish and prosocial individuals also differentially activate a similar vmPFC axis during consideration of benefits to the self and benefits to others (Sul et al., 2015).

With relevance to the nature of intrinsic reward, previous meta-analyses have suggested concrete primary (food or erotic) and abstract secondary (monetary) rewards

are represented more posterior and anterior respectively (Clithero and Rangel, 2013; Sescousse et al., 2013). By this, increased strategic activation further forward in vmPFC could be interpreted as a reflection of secondary (primarily monetary) rewards, whereas altruistic warm glow may activate areas similar to primary rewards.

There are also other differences between altruistic and strategic decisions, such as the certainty and timing of available rewards that may underlie differences this trajectory, requiring further study to rule these in or out as explanations.

4.5. Methodological contributions

We developed a novel method to adjust for dropout when combining effect sizes in a random-effects model that may be useful in future meta-analyses. Results suggests the role of vmPFC in prosocial decision-making may be underestimated due to a lack of coverage in the region. Future fMRI studies on this topic should always utilise methods to minimise this problem (Domsch et al., 2013; Fernandez et al., 2017; Weiskopf et al., 2007) and report the coverage of the region if focusing on the vmPFC.

This study also highlights the importance of a comparison to a selfish control for interpretation of findings. Generous and selfish choices in strategic tasks similarly involve mentalising and extrinsic rewards, making them more similar than altruistic gifts and their selfish alternatives. The latter put extrinsic and intrinsic motivations in direct competition. This point may explain discrepancies between the two previous studies comparing altruistic dictator and strategic ultimatum games (Weiland et al., 2012; Zheng and Zhu, 2013).

4.6. Limitations

Applying fMRI meta-analysis to prosocial decisions identifies consistent activations across studies, tasks and controls. However, averaging ignores connectivity (Hein et al., 2016) and individual differences such as social value orientation (Emonds et al., 2014, 2011; Fermin et al., 2016), attachment style (Schneider-Hassloff et al., 2015), age (Fett et al., 2014), gender (Krach et al., 2009) and personality (Garbarini et al., 2014). This paper includes the decisions of 1150 participants but there are more published studies than had maps or coordinates available for analysis.

Even if all published papers on prosocial decisions had been included, conclusions would be limited to specific experimental settings. The separation between altruistic and strategic decisions which arises from tightly controlled games is only an abstraction of real-world decisions. The real world provides opportunities for strategic and altruistic

motivations to work together (e.g. picking the ideal present, effective altruism and others) that neuroimaging studies are now only beginning to explore. Moreover the lab differs from the real world in other respects. For instance, is difficult to measure reputational motivation in lab-based altruistic studies. There is no way to completely prevent the participant considering the experimenter's view of them or their decisions so reputational concerns may differ from real-world donation contexts.

The selfish choice was the most common control condition used in the data we received. However, some research suggests that the contrast of prosocial with selfish choices can be complicated to interpret and can be influenced by context (Krajbich et al., 2015). The vmPFC region identified in this meta-analysis has been linked to the overall value of a prosocial decision and therefore how long it takes to make (Hutcherson et al., 2015). It was not possible to incorporate response time into the present analysis, due to under reporting of reaction times and reliance on group averages.

This meta-analysis also grouped different paradigms together under themes (i.e. multiple economic games). To precisely control for and differentiate the effects unique to individual paradigms would require tightly controlled individual studies or many times the number of available studies.

5. Conclusion

This study is the first systematic fMRI meta-analysis to demonstrate the consistency and differentiation of neural correlates for strategic and altruistic prosocial decisions. We identify the locations of reliable activations for two different types of prosocial choice.

While we knew both types of gift benefit others, we show that both are also consistently rewarding to the giver and share many neural substrates. In contrast, mechanisms for the decision not to give appears to vary, depending on the strategic context.

Despite some similarities, altruistic and strategic decisions to give are not interchangeable in the brain. The potential for an extrinsic benefit changes the neural mechanism at the point of choice in both level and location of activation. Of the regions involved in prosocial choices overall, sgACC activated more for altruistic than strategic decisions and NuAcc was more active for strategic than altruistic choices. That any region is more involved in altruistic decisions suggests that there is something additive and special about giving when the only benefit is a warm glow. Greater reward-network activity for strategic decisions is in line with and may account for higher rates of prosocial choices in these contexts, compared to altruistic ones.

Across regions linked to social cognition, results were mixed, reflecting the complexity and diversity of theory of mind application in our social lives.

The changing cytoarchitecture across the vmPFC also differentiates strategic and altruistic gifts. Activity increased for strategic and decreased for altruistic choices along a posterior to anterior axis in a way that relates well to what we know about the connectivity and functions across this region.

Methodologically, we provide an additional step forward in the methods of meta-analyses to account for variable signal dropout between fMRI studies.

With consistent effects in hand, the field can move forward with focused examination to distinguish between competing interpretations of each region's role in prosocial decision-making.

Adjusting for Variable Brain Coverage in Voxel-Based fMRI Meta-Analysis

Jo Cutler* [1], Joaquim Radua [2, 3, 4, 5, 6] and Daniel Campbell-Meiklejohn [1]

[1] The Social Decision Lab, the University of Sussex, School of Psychology, Falmer, UK.

[2] Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

[3] FIDMAG Germanes Hospitalaries, Barcelona, Spain

[4] Mental Health Research Networking Center (CIBERSAM), Madrid, Spain

[5] Department of Psychosis Studies, King's College London, Institute of Psychiatry, Psychology and Neuroscience, London, UK

[6] Centre for Psychiatric Research and Education, Karolinska Institutet, Department of Clinical Neuroscience, Stockholm, Sweden

*Correspondence:

Jo Cutler – J.Cutler@sussex.ac.uk

Meta-analyses of fMRI studies are vital to establish consistent findings across the literature. However, fMRI data are susceptible to signal dropout (i.e. incomplete brain coverage), which varies across studies and brain regions. In other words, for some brain regions, only a variable subset of the studies included in an fMRI meta-analysis have data present. These missing data can mean activations in fMRI meta-analysis are underestimated (type II errors). Here we present SPM (MATLAB) code to run a novel method of adjusting random-effects models for meta-analytic averaging of a group of studies and mixed-effects models for comparison between two groups of studies. In two separate datasets, meta-analytic effect sizes and z-scores were larger in the adjusted, compared to the unadjusted analysis. Relevantly, these changes were in regions such as the ventromedial prefrontal cortex where coverage was lowest. Limitations of the method, including issues of how to threshold the adjusted maps are discussed. Code and demonstration data for the adjusted method are available at https://doi.org/10.25377/sussex.c.4223411.

1. Introduction

Echo-planar images for functional magnetic resonance imaging (fMRI) are susceptible to signal dropout (Ojemann et al., 1997) leaving gaps in activation maps. The level of coverage can vary widely between individuals, scanners, and scan protocols. This presents a problem of false negatives (type II errors – wrongly concluding no effect exists) for both individual studies and for map-based fMRI meta-analyses.

The problem of missing data from lack of coverage is limited to a subset of regions, including ventromedial prefrontal cortex (vmPFC) and temporal lobes, due to factors such as nearby air and bone. False negatives are therefore localised to these regions and not uniformly distributed throughout the brain. Research on topics and tasks which rely on these regions, for example value-based decision-making in vmPFC (Levy and Glimcher, 2012), will be disproportionately affected by issues of coverage. In addition, if meta-analyses test for convergence, incomplete brain coverages may lead to incorrect p-values, because the test assumes that "false foci" are uniformly distributed across the brain (Albajes-Eizagirre and Radua, 2018).

While techniques have been developed to maximise coverage (Weiskopf et al., 2007), they are not uniformly successful or universally applied. In this report, we discuss an approach to reducing these type II errors in map-based fMRI meta-analyses.

Using meta-analysis techniques on neuroimaging data is vital to establish consistent neural correlates across studies (Müller et al., 2018; Wager et al., 2009, 2007). Several tools are available. One technique for meta-analysis is Anisotropic Effect Size Signed Differential Mapping software (AES-SDM, Radua et al., 2014) which combines coordinate-based meta-analysis with unthresholded maps (Radua and Mataix-Cols, 2012) to reduce assumptions of the spatial extent of activations.

Dropout in where signal is present can mean activations in fMRI meta-analysis are missed or underestimated. This will be, at least in part, due to voxels where no effect size was measured, being attributed the same variance estimates as voxels where effect sizes were measured. Here we present code which runs a novel method of adjusting both random and mixed-effects models, for meta-analytic averaging across a single group or comparison between two groups of studies respectively. The code adjusts each type of variance (within-study & between-study) in the models used in AES-SDM, which are usually assigned to every voxel, so only voxels where data was recorded are included.

2. Materials and methods

2.1. Data selection

This technique to account for variable coverage was developed as part of an fMRI meta-analysis on prosocial behaviour so a detailed description of selection methods is provided elsewhere (Cutler and Campbell-Meiklejohn, 2019). Briefly, meta-analyses were calculated for each of two groups of decision type, "altruistic" and "strategic", using random-effects models and these groups were compared with a mixed-effects model. The altruistic group contained 18 maps and 3 coordinates sets (n = 21, 557 participants) while the strategic group had 10 maps and 5 coordinates sets (n = 15, 593 participants). Due to different control conditions across studies, the adjusted analysis was only run on studies which contrast altruistic (n = 12) or strategic (n = 12) with selfish decisions.

To establish the wider relevance of the method, we conducted a second meta-analysis using data which researchers have made available through NeuroVault (Gorgolewski et al., 2015). It is vital to stress that this is in no way a comprehensive meta-analysis of any tasks and it is unlikely that a genuine meta-analysis would group these maps together. These maps were simply used as their CC0 license enables sharing as a demonstration set with the code (available at https://doi.org/10.25377/sussex.c.4223411).

Searches on NeuroVault were conducted for "choice" and "deci*" (for decision, decide etc.). Maps were selected if they had data from any decision task in the scanner with a contrast to no decision or a decision which varied on a parameter, for example complexity. This crude selection technique resulted in 18 maps (see Table 1).

Coverage was investigated by binarising each map, after registration to a common template, based on whether there was signal in each voxel and summing these images to create coverage maps (Figure 2). Both the dataset on prosocial decisions and the NeuroVault dataset on decisions showed decreased coverage around the periphery, particularly in vmPFC.

Table 1. Details of the studies with data available from NeuroVault included in the second meta-analysis

First author	Year	NeuroVault collection name	Map used	Map type	Sample size
Park	2017	Integration of individual and social information for decision-making in groups of different sizes	Brain regions for changing judgments to conform to the group's	T	23
Chang	2011	Great expectations: neural computations underlying the use of social norms in decision-making	Offer3_BetweenExpect	Z	17
Bang	2018	Distinct encoding of decision confidence in human medial prefrontal cortex	GLM1: Group-level main effect of coherence (un-masked)	T	32
Fleming	2018	Neural mediators of changes of mind about perceptual decisions	Parametric effect of reported confidence	T	22
Gonzalez Alam	2018	Meaningful inhibition: Exploring the role of meaning and modality in response inhibition	Figure 2. Semantic and Perceptual Inhibition Overlap	Z	27
Tom	2007	The Neural Basis of Loss Aversion in Decision-Making Under Risk	task001_cope03_parametric gain	T	16
Waskom	2016	Adaptive Engagement of Cognitive Control in Context-Dependent Decision Making	Context prediction error	Z	15
Op de Macks	2018	Supplement: Neural correlates of adolescent decision-making and outcome processing during the Yellow Light Game	Stop > Go (N=78, 11-17 years-old, community sample)	T	78
Suzuki	2015	Neural Mechanisms Underlying Human Consensus Decision-Making	Fig.4A (Main + Control)	T	20
Rahnev	2016	Causal evidence for frontal cortex organization for perceptual decision making	Stimulus/perceptual judgment epoch	T	17
Li	2017	Reason's Enemy Is Not Emotion: Engagement of Cognitive Control Networks Explains Biases in Gain/Loss Framing	Frame-Consistent>Framing-Inconsistent, unthresholded	Z	143
Aridan	PreP	Neural correlates of effort-based valuation with prospective choices	effort_gain_pos	Z	40
Cho	2016	Reward Sensitivity Enhances Ventrolateral Prefrontal Cortex Activation during Free Choice	Figure2 Free vs Forced TEPS unthresh	Z	33
Kameda	2016	Rawlsian maximin rule operates as a common cognitive anchor in distributive justice and risky decisions	Activity of the right temporo-parietal junction (RTPJ) during decision making.	T	30
van der Laan	2014	Sweet lies: neural, visual, and behavioral measures reveal a lack of self-control conflict during food choice in weight-concerned women	Choice periods SC vs NSC	T	20
Hunt	2014	Hierarchical competitions subserving multi-attribute choice	Integrated value competition	Z	19
Korn	2018	Heuristic and optimal policy computations in the human brain during sequential decision-making	GLM2: Participants' choices (binary parametric modulator: foraging versus waiting)	T	28
Fujiwara	2018	Ventrolateral Prefrontal Cortex Updates Chosen Value According to Choice Set Size	spmT 0001	T	18

2.2. Combined image and coordinate meta-analysis

Using maps in fMRI meta-analysis has a number of benefits including enhanced sensitivity and detection of consistent but subthreshold effects. When maps are unavailable, AES-SDM recreates estimated maps from coordinates and their effect sizes using an anisotropic kernel. If obtainable, peaks can be entered in both directions of the contrast.

Statistics other than T are transformed before all maps, including those recreated from coordinates, are aligned to a common template. The software then implements a permutation-based analysis. The recommendation is 50 permutations which creates 50 randomisations with the same number of foci as the map of interest. These preprocessing steps result in recreated NIfTI maps of effect sizes and within-study variance for each study. These maps are used in the both the original, unadjusted method by the software and the adjusted technique described here. Adjusted analysis uses custom scripts in SPM12 (Statistical Parametric Mapping, http://www.fil.ion. ucl.ac.uk/spm) which are available under an MIT license from https://doi.org/10.25377/sussex.c.4223411 and github.com/jocutler/adjusting-dropout-fMRI-meta.

2.3. Random-effects model

2.3.1. Unadjusted model

One widespread use of meta-analysis is to calculate mean effect sizes across studies. A common method, and the method used in AES-SDM, is a random-effects model. In the model, AES-SDM weights each study by the inverse of the total (within-study and between-study) variance. The between-study variance, τ^2, is obtained by the DerSimonian-Laird estimator (Dersimonian and Laird, 1986) as:

$$\tau^2 = \max\{0, Q_w - (k-1)\} / [\sum_i w_i - (\sum_i w_i^2 / \sum_i w_i)]\}$$

Where w_i is a weighting calculated as the inverse of the ith study's within-study variance, k is the number of studies and Qw is calculated as:

$$Q_w = \sum_i w_i (y_i - \bar{y}_w)^2$$

Where y_i is the ith study effect size estimate, w_i is a weighting calculated as the inverse of the ith study's within-study variance and $\bar{y}_w$ is the weighted estimate of the overall effect size calculated as:

$$\bar{y}_w = \sum_i w_i \, y_i \, / \sum_i w_i$$

In simplified terms, used in the code:

$$\tau^2 = \text{numerator} \, / \, C$$

Where:

$$\text{numerator} = \max \{0, Q - DoF\}$$

With Q calculated as above, degrees of freedom (DoF) the number of studies -1 and:

$$C = [\sum FE \text{ weightings} - (\frac{\sum FE \text{ weightings}^2}{\sum FE \text{ weightings}})]$$

Where FE weightings are the inverse of the studies' within-study variance. These are referred to as fixed-effects (FE) weightings as they are the ones used for a fixed-effects model, which just takes into account within-study (not between-study) variance.

In practice, these equations demonstrate that the between-study variance (τ^2) depends on the sum of the weightings (w_i) which are the inverse of the within-study variance.

2.3.2. Adjusting random-effects models

In fMRI meta-analysis, the within-study variance is a single number which is applied as the variance across every voxel in the brain within the mask of interest. The effect size for that voxel is the transformed effect size created during preprocessing. However, if signal dropout has occurred, the effect size is zero. This means that voxels with no recorded signal are attributed variance but no effect size. When the FE weightings and effect sizes are each summed during calculation of the between-study variance, these voxels are contributing to the total variance without contributing an effect size. This could underestimate meta-analytic effect size due to inflated variance.

To account for this, calculations for the meta-analysis can be adjusted so only studies where data was recorded contribute weightings to the calculation of τ^2. This can be done either at the single-voxel level with a spreadsheet (Figure 1) or across the whole brain, voxel-by-voxel, by masking variance maps with their coverage. The DoF value is also adjusted to be the number of studies with data – 1.

It is important to note that maps recreated from coordinates should *not* be adjusted *unless* the coverage is known. If the coverage is unknown, effect sizes with values of zero do

not necessarily imply lack of signal and could meaningfully demonstrate the voxel is too far from any peaks to be attributed effect size. Of course, the maps which these coordinates were generated from could also suffer from signal dropout but this cannot be confirmed. This is another reason to obtain maps wherever possible.

If the coverage is known, for example if the paper states the cerebellum was not analysed, a coverage map reflecting this could be created and entered into the analysis as the mask for that study. This was not done for any of the studies in the current analyses.

Once τ^2 has been calculated as a single number for the between-study variance, it is added to the within-study variance for each study to provide total variance. The inverse of this total variance provides the random-effects (RE) weightings for each study, by which the effect-size estimates are multiplied.

The issue of variable coverage affects results again at this stage as the overall meta-analytic effect size (Hedges' g) is calculated by the sum of the weighted effect sizes divided by the RE weightings summed:

$$g = \sum RE\ weighted\ effect\ sizes\ /\ \sum RE\ weightings$$

The meta-analytic variance map, used to calculate standard error and z-scores, is calculated as the inverse of the RE weightings summed:

$$variance = 1 / \sum RE\ weightings$$

Again, voxels with no effect-size estimate due to missing data will contribute zero to the effect sizes but increase the sum of the RE weightings in both of these calculations. A greater value of summed weightings leads to underestimation of g and overestimation of variance.

The same principle can be applied here as with the FE weightings, where studies are only included in the weightings sum if they have an effect size present.

2.4. Mixed-effects model

2.4.1. Unadjusted model

In addition to calculating the mean effect size for a group of studies, meta-analysis can calculate the difference between two groups using a mixed-effects model. The calculations and method of adjusting are similar to random-effects models, except the DoF equals the total number of studies across groups -2, the calculation of Q is:

$$Q_{total} = Q_{group\ 0} + Q_{group\ 1}$$

And the calculation of C is:

$$C_{total} = C_{group\ 0} + C_{group\ 1}$$

Where Q and C for each group separately are calculated as above. Groups are referred to as 0 and 1 to match dummy coding in AES-SDM.

Study	Effect size estimate	Within Var	Map?	Fixed effects (FE)					Random effects (RE)							
									Unadjusted (All)				Adjusted (Adj)			
				Within Var	FE Weighting (Wt , 1/Var)	Es*FE Wt	Es^2*FE Wt	FE Wt^2	RE All Tau-sq	RE All Total Var	RE All Wt	Es*RE All Wt	RE Adj Tau-sq	RE Adj Total Var	RE Adj Wt	Es*RE Adj Wt
1	0.00	0.05	0	0.05	20.92	0.00	0.00	437.67	0.20	0.25	4.02	0.00	0.29	0.34	2.95	0.00
2	0.00	0.06	1	0.06	16.39	0.00	0.00	268.74	0.20	0.26	3.82	0.00	0.29	0.35	2.84	0.00
3	0.63	0.04	1	0.04	23.81	14.93	9.36	566.89	0.20	0.24	4.12	2.58	0.29	0.33	3.00	1.88
4	1.28	0.09	1	0.09	11.17	14.30	18.30	124.81	0.20	0.29	3.44	4.41	0.29	0.38	2.63	3.37
5	0.00	0.06	1	0.06	16.17	0.00	0.00	261.60	0.20	0.26	3.81	0.00	0.29	0.35	2.84	0.00
Sum:				All:	88.47	29.23	27.66	1659.72	All:		19.20	6.99	Adjusted:		8.59	5.25
Sum:				Adjusted:	55.90	29.23	27.66	1129.37								

N all.	5
N adj.	3

Calculating between-study variance		
	All	Adj
Q	$\sum(Es^2*FEWt) - (\sum Es*FEWt)^2 / \sum FEWt$	
DoF	N all. - 1	N adj. - 1
Numerator	max(Q - DoF, 0)	
C	$\sum FEWt - \sum FEWt^2 / \sum FEWt$	
Tau-sq	Numerator / C	

Between-study variance		
	All	Adj
Q	18.01	12.38
DoF	4	2
Numerator	14.01	10.38
C	69.71	35.70
Tau-sq	0.20	0.29

Random effects statistics		
	All	Adj
Effect size	0.36	0.61
Variance	0.05	0.12
Stnd error	0.23	0.34
SDM-Z	1.59	1.79

Figure 1. Demonstration of an adjusted random-effects model on a single voxel for five studies. Shaded rows (studies 1, 3 & 4) are those which are still taken into account in the adjusted analysis. "All" refers to the unadjusted analysis with all studies. Study 1 is from coordinates so this is included despite having an effect-size estimate of zero as this could be meaningful (see section 2.3.2 for details and exception to this rule). Studies 3 and 4 are maps with non-zero effect-size estimates in the voxel of interest, so are included. Studies 2 and 5 are maps with zero effect-size estimates, suggesting missing data due to signal dropout, so these are not included in the adjusted analysis. The fixed-effects section is purely used to calculate the between-study variance (τ^2: tau-sq) with the calculations for this shown in the box "calculating between-study variance". Although τ^2 increases from 0.20 to 0.29 in the adjusted analysis, likely linked to having less studies, the meta-analytic effect size (g) increases substantially from 0.36 to 0.61 as the sum of the RE weightings (which the sum of the weighted effect sizes is divided by) decreases from 19.20 to 8.59. An interactive spreadsheet in this format can be downloaded from https://doi.org/10.25377/sussex.c.4223411 to run this analysis on any voxel for any set of studies.

Calculating τ² follows the same process as above to produce a single number for the between-study variance across all the studies in both groups. This is added to the within-study variance and the inverse of this total variance provides the study's mixed-effects (ME) weighting.

The meta-analytic effect size (Hedges' g) is then calculated with the formula shown above for *each group of studies separately* - the sum of ME weighted effect sizes divided by the sum of ME weightings for that group. The meta-analytic variance is also calculated as above for each group of studies separately: the inverse of the ME weightings sum for that group.

The Hedges' g effect-size map for the difference between groups is the calculated by *subtracting* the two separate effect-size maps:

$$g_{1-0} = g_{group\ 1} - g_{group\ 0}$$

The meta-analytic variance for the difference between groups is calculated by *summing* the two separate variance estimates:

$$variance_{1-0} = variance_{group\ 0} + variance_{group\ 1}$$

2.4.2. *Adjusted model*

As with random-effects models, voxels with zero effect sizes due to dropout are attributed within-study variance and so increase the sums of FE and ME weightings. This is likely to underestimate average effect sizes. The same adjustments can avoid this problem in a mixed-effects model by excluding weightings of voxels where no effect size is present, unless the map was recreated from peaks with unknown coverage. This adjustment can again be done for a single voxel or across the whole brain. The DoF becomes the total number of studies, across both groups, which contribute weightings – 2.

2.5. Z-maps and thresholding

Once the effect-size and variance maps have been adjusted, maps of standard error and z-scores can be produced. As the input came from permutations in AES-SDM, z-scores are "SDM-Z" because they do not follow a normal distribution.

Standard error (SE) is the square root of the meta-analytic variance, either for a single group or the difference between groups (shown):

$$SE_{1-0} = \sqrt{variance_{1-0}}$$

SDM-Z is the effect size (Hedges' g) divided by the standard error, either for a single group or the difference between groups (shown):

$$SDM\text{-}Z_{1-0} = g_{1-0} \, / \, SE_{1-0}$$

Thresholding in AES-SDM uses a voxel-level threshold of $p<0.005$ which approximates $p<0.05$ corrected and balances specificity and sensitivity (Radua et al., 2012). However, in SDM-Z maps from the adjusted method, voxels have differing DoF meaning thresholding is not straightforward.

In the prosocial decisions meta-analysis (Cutler and Campbell-Meiklejohn, 2019), maps were thresholded with SDM-Z > 2.3. This was chosen as a common value for thresholding, close to the average of the critical SDM-Z values generated in the original, unadjusted analyses (with all studies) and AES-SDM analyses run with the 50% of maps with the best coverage. Here, we apply the same threshold to the NeuroVault data. We recognise this is not a perfect solution but it provides continuity and this analysis is not meaningful, regardless of thresholding method, other than for demonstrating the impact of the adjustment.

3. Results

In both random and mixed-effects analyses, in both datasets, the adjusted method increased effect sizes across the lower vmPFC where coverage was worst (Figure 2). Activations based on SDM-Z > 2.3 uncorrected were larger in the adjusted than the unadjusted analysis.

3.1. Prosocial decisions

In the prosocial decisions data, effect sizes and the size of SDM-Z > 3 activations increased in the adjusted analysis for altruistic vs. selfish (Figure 2b iii) and altruistic vs. strategic (Figure 2b iv). For altruistic > strategic in posterior vmPFC and strategic > altruistic in anterior vmPFC, some activations were shown only in the adjusted analysis. This dissociation fits with findings of a posterior to anterior vmPFC axis differentiating altruistic from strategic decisions (Cutler and Campbell-Meiklejohn, 2019).

3.2. Decision-making (NeuroVault data)

Results from the second dataset from NeuroVault on decision-making further support the use of the adjusted method to account for coverage. Again, it is vital to stress that this analysis purely provides a second application of the adjustments for signal dropout and results are not a comprehensive meta-analysis or a meaningful group.

Effect sizes across vmPFC increased in the adjusted analysis in both the random-effects model on all 18 studies' decision-making condition vs. control (Figure 2b i) and the mixed-effects model comparing 2 randomly-allocated groups of 9 studies each (Figure 2b ii). Similarly, regions where SDM-Z > 2.3 were larger or only present in the adjusted analysis.

Figure 2. Coverage maps and results. **(A)** Coverage maps showing the number of studies with data in each region, x = 0, n = the number of studies available and the maximum possible coverage. **(B)** Increased effect sizes (Hedges' g; top rows) and larger regions of SDM-Z > 2.3 (bottom rows) in the adjusted analysis accounting for vmPFC signal dropout, compared to the unadjusted analysis with all studies. Decisions (data from NeuroVault; n = 18) **(i)** average (random-effects) and **(ii)** comparison (mixed-effects). Prosocial decisions **(iii)** altruistic average (n = 12; random-effects) **(iv)** comparison between altruistic and strategic decisions (n = 24; mixed-effects).

4. Discussion

This paper presents a novel method of adjusting voxel-based fMRI meta-analysis technique AES-SDM (Radua et al., 2014) to account for variable brain coverage between studies. Increased effect sizes and larger regions with substantial z-scores were found in the adjusted, compared to the unadjusted analysis, using two datasets.

Our results suggest that regions that suffer from signal dropout, such as vmPFC, can show false negatives if coverage is not accounted for. The role of these regions may be underestimated or overlooked, preventing a complete understanding of their function. In meta-analyses, lack of coverage in some studies may obscure an activation despite high consistency in the studies with data in that region.

The uneven spatial distribution of signal dropout may also increase false positive results in the rest of the brain. Specifically, the test for convergence used in current coordinate-based meta-analyses assumes a uniform distribution of false positive study peaks (Albajes-Eizagirre and Radua, 2018), but this assumptions is incorrect if some brain regions have no data in some studies. This additional problem should not happen in the upcoming version of SDM, which no longer conducts tests for convergence.

To overcome these issues, our method adjusts random and mixed-effects meta-analysis models to only include variance, at each calculation stage, from voxels with study-level effect-size data. This means the number of contributing studies ranges between one and the total number of studies. The voxel degrees of freedom will be the number of contributing studies -1 (random-effects model) or -2 (mixed-effects model). For some voxels with only a few studies contributing, single studies could dominate and prevent the benefits of meta-analysis from being realised. In comparisons between two groups, differences between the sizes of each group could be driven by coverage.

Perhaps the biggest challenge from varying degrees of freedom across voxels is thresholding the adjusted meta-analytic map. In the prosocial decisions meta-analysis, SDM-Z > 2.3 was applied to the adjusted and unadjusted analysis as a comparison (Cutler and Campbell-Meiklejohn, 2019). Although this is liberal for z-scores which follow a normal distribution, SDM-Z scores do not follow a normal distribution (Radua et al., 2012) 2.3 was close to the threshold SDM-Z score generated in the unadjusted analysis with all studies and the half with the best coverage. Here, we also threshold the second meta-analysis with SDM-Z > 2.3 for continuity.

This is not a perfect solution and this adjusted method is perhaps best used alongside unadjusted analyses thresholded in a more appropriate way. The usefulness of a simple SDM-Z threshold may be limited to visual comparisons between unadjusted and adjusted maps. Comparing adjusted and unadjusted effect-size maps (Hedges' g) can also reveal the impact of coverage deficits.

Evidence for the importance of adjusting calculations is presented using two datasets. One is a comprehensive meta-analysis of prosocial decisions. The second group is studies with data on NeuroVault for decision tasks. Decision-making was chosen as it activates vmPFC (Levy and Glimcher, 2012), a region prone to reduced coverage. This data was purely used as the availability enables sharing to demonstrate the code and results are not be meaningful in any other way. That the NeuroVault data did not show large vmPFC activations is likely due to variety in tasks, which would not be grouped together in a genuine meta-analysis.

5. Conclusion

Using meta-analysis techniques on fMRI data establishes consistent findings across samples, scanning sites and tasks, providing many benefits and overcoming issues with single studies. Including statistical maps enhances some of these benefits and methods like AES-SDM, which combine maps with coordinates, increase the chances of including a study. However, fMRI meta-analyses are prone to false negatives if lacking coverage leaves missing data in study-level statistical maps. Several key regions, including vmPFC and temporal areas suffer from signal dropout, meaning false negatives are unevenly distributed throughout the brain.

To account for and overcome these issues, we present a novel method of adjusting calculations for random and mixed-effects models for meta-analytic group averages and comparisons respectively. By adjusting models to only include variance for voxels with data present in the study's effect-size map, we demonstrate increased meta-analytic effect sizes in regions with the worst coverage. This highlights that failing to account for coverage underestimates effect sizes or may miss activations altogether.

Scope insensitivity and proximity bias affect physiological responses to lives at risk

Jo Cutler, Jolyon J. Miles Wilson and Daniel Campbell-Meiklejohn

The Social Decision Lab, the University of Sussex, School of Psychology, Falmer, UK.

1. Introduction

Enhanced emotional and physical responses to the experiences of other people are linked with behaviours that save and improve lives (Brethel-Haurwitz et al., 2018; Marsh et al., 2014). Individual responses to people suffering include empathy and personal distress, which can both motivate helping for different reasons (Batson, 2014; Batson et al., 1991, 1987). These individual responses also contribute to a collective response, which can impact political decisions and outcomes for those affected (Slovic et al., 2017).

However, the suffering of others can also be overwhelming (Cameron and Payne, 2011). Events and issues that prompt emotional and prosocial responses are too numerous to all be acted on. Each day there are reports about the most recent natural disasters, violent conflicts and health issues. These feature people who have already died and people who are still at risk, in need, and often suffering. Some reports of these tragedies lead to huge responses, both from the public and on an individual level (Slovic et al., 2017). In contrast, others are largely ignored, or the response is not proportional to the scale of the tragedy. Here we present a novel stimulus set, adapted from news stories about people at risk of death and the outcomes.

Previous work has identified multiple factors that shape emotional and behavioural responses to lives at risk and to the deaths of strangers. These will be described here as biases when they lead to responses that are not in line with the commonly held moral view that all lives are equal (Dickert et al., 2012).

One important factor is where the people are. While overseas aid and disaster relief charities receive sizable donations, they are less popular than causes that affect people in one's own country (UK data, Charities Aid Foundation, 2019). Differences in responses to events in different locations have been explored and explained along a range of dimensions, which often overlap. We designed the stimuli to be inclusive of all of them.

Another factor that creates biased responses is the number of people in need or suffering. A range of non-linear shapes have been suggested to reflect the relationship between the number of people and emotional or behavioural responses. One possibility is that responses show a negative correlation with the number of people, such that the greatest responses are to a single person in need. This singularity effect often coincides with increased information about the person, such as a name or photo, which also promote giving (Genevsky et al., 2013), leading to the so-called "identifiable victim effect" (Lee and Feeley, 2016).

At the opposite end of the number spectrum, there is evidence that information about large numbers of people can decrease giving (Schelling, 1968; Västfjäll et al., 2014), weaken compassion and worsen political attitudes (Azevedo et al., 2019). There are two competing explanations for this *compassion collapse* when the number of people suffering gets very large. First, that it is due to an inability to empathise with too many people (Slovic, 2007) and second, that people are able to empathise but motivated to avoid processing this much suffering (Cameron and Payne, 2011).

A decrease in compassion towards high numbers of people could lead to an overall drop in responses or mean that the additional value of each life decreases. This can mean responses do increase with the number of people but in a non-linear way, known as *scope insensitivity* (Dickert et al., 2015) or *psychophysical numbing* (Fetherstonhaugh et al., 1997). Taken from findings in economics, this bias describes any situation in which incremental increases (or decreases) in value decrease as the total absolute value increases (Kahneman and Tversky, 1979). Applied to the valuation of other's lives, this means the difference between 10,001 and 10,002 people suffering feels much smaller than the difference between 1 and 2 people suffering. Evidence for this bias has been shown in decisions whether to give to charity (Cameron and Payne, 2011) and on life-saving policy interventions (Olivola, 2015).

Most previous studies on this topic have used behavioural measures of willingness to pay or amount donated to quantify these biases. While such measures do show evidence of biases, these may be underestimated if participants feel social pressure to report lower levels of bias than their true beliefs (Fernandes and Randall, 1992). Having to answer questions may also put participants in a more deliberative or analytical state of mind, which has been linked to more utilitarian judgements (Li et al., 2018). These tasks are not representative of passively viewing information about others suffering, for example on television news. Much of the information we receive about the suffering of others is not directly as part of an appeal for help. Instead we often hear news of an event and then must seek opportunities to help if we are moved to do so.

In this research, we use a physiological measure of skin conductance to overcome limitations of self-report. Previous applications of physiological measures to relevant topics include differentiation of perspective taking approaches through a composite physiological stress measure (Buffone et al., 2017) and the development of prosociality (Miller, 2018). Research measuring skin conductance responses (SCRs) specifically has shown arousal when observing others in pain or in need: is greater for someone of the same race

(Forgiarini et al., 2011); increases with age (Sze et al., 2012); and predicts costly helping (Hein et al., 2011).

As SCRs have been linked to anticipation of negative outcomes and uncertainty (Bach et al., 2009; Critchley et al., 2001; Feldmanhall et al., 2016) and negatively framed outcomes (Ring, 2015), we designed our task to maximise these elements. Due to the noisy nature of SCR data, it was necessary to have many trials for each condition. However, this also gave us an opportunity to look at how responses, and the impact of number and location, change over time. This reflects the repeated exposure to news of tragedies in the real world.

We also measured participants' scores on the interpersonal reactivity index (IRI), designed to measure empathy (Davis, 1983). This allowed us to test whether the effect of the number of people and location of the news story differed between those high and low on this trait. Differences would suggest that the emotions the scale measures as a trait are involved in the effect of our variables. While empathy is generally linked with altruism and concern for others (FeldmanHall et al., 2015) there is also evidence that empathy can bias responses leading to greater partiality in who receives help (Batson et al., 1995; Bloom, 2017; Decety and Cowell, 2015; Ugazio et al., 2014). If this is the case, we would predict participants scoring high on the IRI would differentiate events in the UK from those abroad to a greater extent than people who score lower on this trait.

2. Method

2.1. Participants

189 participants (138 female, 50 male, 1 non-binary) in total completed the study. Ages ranged from 18 to 65 with an average age of 21.2. All participants were UK residents. 122 participants self-identified as having a UK nationality, 65 reported a non-UK nationality and 2 did not disclose their nationality. Participants with dual nationalities including the UK were categorised in the UK group, as were those who reported Irish nationality without specifying the Republic of Ireland. The majority (149) were Psychology students taking part for course credit as well as the available prizes, 37 were students in other disciplines and 3 were not students. The procedure was approved by the Sussex Sciences & Technology Cross-Schools Research Ethics Committee.

Participants were compensated up to £5 for their time, depending on the control task, and were also entered into two prize draws of £25. Participants were tested in two rounds, with the outcomes of the news stories (*alive* or *dead*; see below) either truthful (95 participants) or swapped (94 participants) to balance how surprising the outcomes were. Participants in the truthful round also completed an extra control task which related to an additional prize draw of £50.

Usable skin conductance data was available from 186 participants for the news stories task, 184 for control task 1 and 90 for the additional control task 2 completed by just one round of 95 participants. Data from the other participants was either lost during testing due to technical issues or had issues discovered during analysis.

2.2. Stimuli

2.2.1. News stories

In the social 'news stories' task, 120 summaries of events were presented in the form "[number] people in [country] were identified as at risk of death from [cause]". This **presentation** sentence appeared and then 1 second later, a ">" appeared in the right bottom corner indicating that participants could press the space bar as soon as they had read the sentence and were ready to continue. Their press was followed by an **anticipation** period of 4 seconds ± 30% jitter where "..." was presented on the screen. Then a fixation cross appeared for 0.5 seconds before the **outcome phase**: either "they are alive" or "they are dead" appeared and remained on screen for 2s ± 30%. A fixation cross period of 2s ± 30% seconds preceded the next trial.

Stimuli for the news stories paradigm were taken from real news stories in which varying numbers of people were at risk of dying. Initially 156 were obtained and rated by 32 independent participants through an online survey for course credit. For these ratings, the number of people and the country were hidden to isolate the cause of death. The stories were rated at both the anticipation phase, for example "X people were identified at risk of death from a fire" and the outcome phase, for example "X people survived breast cancer". Participants rated these statements on how emotional they felt, whether the emotion was positive or negative, and how much empathy they felt for the people involved.
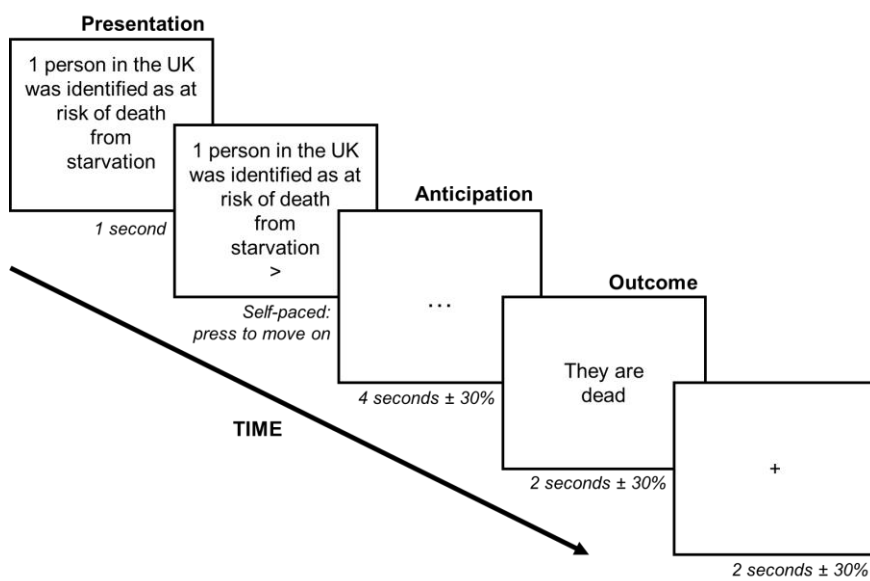
Figure 1. Structure of the task. After the ">" appeared, participants could press space bar when they finished reading the sentence to move on. During the anticipation phase the dots appeared one after the other.

These pilot ratings were combined with the number of people involved and whether the story took place in the UK or in a named foreign country to balance the chosen stories across conditions. The final stimulus set consisted of 120 stories: 5 for each of 24 conditions in a 6 (**number bin**) x 2 (**location**: *UK* vs. *abroad*) x 2 (**outcome**: *positive*: alive vs. *negative*: dead) design. Event type (natural disasters, accidents, violence, disease or long-term issues) was also balanced across the conditions.

Table 1
Stimuli design

Bin	News stories task				Control task 1		Control task 2	
	UK		Abroad		Money		Credits	
	Alive (+)	Dead (-)	Alive (+)	Dead (-)	Keep (+)	Lose (-)	Keep (+)	Lose (-)
1	1		1		£1		1	
2	2 - 10		2 - 10		£2		2 - 10	
3	11 - 50		11 - 50		£3		11 - 50	
4	51 - 500		51 - 500		£4		51 - 500	
5	501 – 10,000		501 – 10,000		£5		501 – 10,000	
6	10,001 +		10,001 +				10,001 +	

Note. (+): positive outcome, (-): negative outcome.

2.2.2. *Control task 1: money*

To provide a non-social comparison, two control tasks presented stimuli with similar manipulations of number and outcome. Both matched the structure and timing of the news stories trials, as shown in figure 1. These tasks were also both in a loss frame, meaning the best outcome was just to keep the resources.

Control task 1 involved the possibility of losing some or all of the £5 payment for participation. One trial was chosen at random and the payment of £5 minus anything lost on that trial was implemented. Each trial in this task presented a **number** of pounds (*£1, £2, £3, £4* or *£5*) that could be lost. After 1 second a ">" appeared in the bottom right corner of the screen to indicate participants could press a button to move on. An anticipation phase (4s ± 30% jitter, "…" on the screen) followed this, before the **outcome** (2s ± 30%) of whether the participant kept (*positive*) or lost (*negative*) the money and a fixation cross (2s ± 30%) before the next trial.

Half of trials for each amount lead to a positive outcome and the other half a negative outcome. Participants were told that overall, there was a 50:50 chance on each trial. This created a 5 (**number**) x 2 (**outcome:** *positive*: keep or *negative*: lose) design with 5 trials in each of the 10 conditions. All 189 participants completed this task in a counterbalanced order with the news stories task.

2.2.3. Control task 2: credits

A second control task was only completed by one of the two rounds of participants (95 of the total 189). This involved the possibility of losing credits which translated into participants' chances to win an additional prize of £50. Participants started the task with 500,000 credits but each trial of this task could take some away. The **presentation** phase showed an amount of credits that might be lost, matched to the 6 number bins in the news stories task. The timing and details of the **anticipation** and **outcome phases** were identical to control task 1. This task had 60 trials in total, 5 in each of 12 conditions, fulling crossing *number bin* (*1-6*) with *outcome* (*positive*: keep) or (*negative*: lose), with participants aware of the 50% chance of each outcome.

2.3. Measures

2.3.1. Willingness to pay

To provide a behavioural measure of the value of a life in each condition, we adapted a willingness to pay measure (Frederick and Fischhoff, 1998). Participants were asked to imagine they were a "UK government minister in charge of allocating resources to stop people dying from flooding both in the UK and abroad". Participants then filled in the amount, from £0 - £100,000, to allocate to 12 projects, 6 in the UK and 6 in a named foreign country, to save different numbers of people. The numbers corresponded to the average numbers in each bin of the news stories stimuli.

Instructions read "You must consider the following proposals and allocate funding. For each project, there is a maximum of £100,000 available which you should allocate if the project is very important. You must not spend unnecessary money because the money can also go to other worthy causes. If a project is not worthwhile you can decide to give it £0. However, you cannot use the money you save on one project now for another one of the projects listed, imagine any money you save would be used on different future projects."

2.3.2. Interpersonal reactivity index (IRI)

Participants completed the 28-item version of the IRI measure (Davis, 1983). This includes 7 items of each subscale; Perspective Taking, Fantasy, Empathic Concern and Personal Distress and in analysis we used the total for all subscales.

2.3.3. *UK identity*

Participants' level of identification with the UK were measured by 3 items adapted from Vroome et al. (2014). These were "My UK national identity is important to me", "I identify strongly with the UK", and "I feel really connected to the UK". Each were rated on a 5-point scale from "does not describe me" to "describes me extremely well". An average of these items was used in analysis.

2.4. Procedure

On arrival at the lab, participants washed and dried their hands thoroughly, removed any jewellery from their hand and arm and were seated in an experimental cubicle. With consent, we attached the electrodes for the skin conductance recording (index and middle finger) and pulse oximeter (ring finger) immediately. This allowed maximum time for them to settle on the skin before recording, while we obtained full informed consent. Participants then completed several questionnaire measures not reported here via Qualtrics Online Survey Software (Qualtrics, Provo, UT).

Before starting the main tasks, the maximum payment of £5 was placed in cash beside participants, clearly in view, to maximise the salience of the loss frame used in the control task. Participants then completed the main tasks with instructions presented immediately before each started. All tasks were run in MATLAB (2017b, Mathworks, Inc., Natick, MA) using Psychophysics Toolbox extensions version 3 (Brainard, 1997). The lights were turned off for the duration of the experimental tasks.

The order of the tasks was counterbalanced such that some participants completed the control task then the news stories task and for others it was the opposite. Three further tasks not reported here were completed next. For the 95 participants who completed control task 2, this and a further task not reported here followed, then all participants completed further questionnaire measures, including willingness to pay and UK identity, as the final part of the study.

During the debrief, participants were asked a) whether there were any of the stories that they recognised during the presentation period, so knew the outcome before it was presented; b) whether there were any foreign countries or c) causes of death they have a personal link to; and d) whether they believed the news stories were real. One trial was chosen at random to determine the payment of up to £5.

2.5. Data acquisition

Skin conductance and heart rate data were collected using a Biopac MP36 sampling at 1000Hz. All equipment was put on the left hand when possible and most left-handed participants were happy to complete the simple button presses required for the study with their non-dominant hand. We measured skin conductance with two EL507 EDA isotonic gel electrodes with 11mm Ag/AgCl contact attached to the head of the index and middle fingers. Heart rate data were collected using a pulse oximeter on the ring finger. Here we do not report the heart rate data as there was too much noise, even after manual cleaning.

2.6. Analysis

2.6.1. *Skin conductance preprocessing*

Parameters for the response to each event were extracted using the dynamic causal modelling (DCM) function (Bach et al., 2010; Staib et al., 2015) from the PsychoPhysiological Modelling Toolbox version 4.0.1 (Bach et al., 2018; Bach and Friston, 2013) in MATLAB (2017b, Mathworks, Inc., Natick, MA). The toolbox enables separation of responses to stimuli with inter-stimulus intervals shorter than the skin conductance response. DCM was chosen over the general linear model function for the option to model events with flexible timing. The model searches a specified time window for the responses and estimates the dispersion and latency, as well as the amplitude. Default settings were used, except the minimum dispersion was increased to 0.3.

Due to the novelty of our paradigm, and small number of previous papers measuring SCR to complex social stimuli, we generated 15 different timing models for the events in a trial. These varied in which events were modelled (for example presentation onset or an anticipatory response to the preceding fixation cross), whether each was fixed or flexible and the specified time windows. For example, some models included a flexible event that started at the presentation and continued throughout the anticipation period, whereas others modelled the presentation and anticipation separately.

All these models were run on data from the news stories task and control tasks separately. Models were compared through the residual sum of squares for each model for each participant. Importantly, this whole process was completely blind to conditions and represents the fit to all events across all trials. Of participants with usable data, the same two models showed the best fit for 126 (62+64) of 186 on the news stories task, 131 (72+59) of 184 on the monetary control task and 71 (41+30) of 90 on the credits control task. These two models just differed in whether the outcome event started at the onset or

the fixation cross preceding it by 0.5s. Latency parameters for this outcome event changed by 0.5 seconds in almost all cases, suggesting the same peak was being detected. The parameter estimates (response amplitude, latency and dispersion) from the model starting the outcome at its onset were used for all participants in all further analyses, as this was the better fit in slightly more cases (175 vs. 153).

The chosen model consisted of a flexible *presentation* event that started as soon as the stimulus was presented and continued until participants pressed to indicate they had read the sentence and the screen moved on. The next flexible *anticipation* event started at this button press and continued until just before the outcome appeared. The third and final flexible *outcome* event started as the outcome appeared and continued through the outcome being displayed and following fixation cross before the next trial. The end time of one event was always 1ms before the start time of the following event, meaning the modelled time was continuous.

The SCR amplitude, latency and dispersion estimated by PsPM DCM relate to the sudomotor burst that causes a change in skin conductance. The area under the SCR curve, as measured with traditional methods, is a product of the amplitude and dispersion but here we analyse each separately. Despite possible physiological reasons that participants may not show any SCRs, we did not exclude any participants as "non-responders", as it is impossible to separate these explanations from psychological reasons to not show responses during the current tasks.

2.6.2. Group-level models

We extracted and Z-scored amplitude parameter estimates for each presentation and each outcome event, for each participant, as recommended for SCR data (Bach, 2014; Staib et al., 2015). This was done on each task separately, to account for the different order of tasks, so absolute values of responses cannot be compared between tasks. We also extracted the raw dispersion parameters. We did not analyse parameters for responses during the anticipation phase as it is not clear how these relate to responses to the presentation while no new information is available.

We then analysed all trial-by-trial SCR parameters and behavioural willingness to pay responses using lme4 (Bates et al., 2015) in R (R Core Team, 2017) with RStudio (RStudio Team, 2015). For analysis of physiological measures, due to the skewed nature (even after Z-scoring in the case of the amplitudes), a gamma distribution was used, based on a diagnostic test and a log link function used, in favour of any further transformations

(Lo and Andrews, 2015). This required Z-scores to be rescaled to become positive. All continuous variables were mean-centred.

For the number of people (news stories task) or credits (control task 2) we also transformed the raw number of people with $\log_{10}$ before mean-centring, as physiology could never change on the same scale as the number of people variable. The number of pounds (£1-£5) was analysed on that scale which is comparable to the average number of credits or people from each bin, once transformed with $\log_{10}$.

As identified in the introduction, previous research has found non-linear effects of the number of people on responses or behaviours. To test for these in our data, we include polynomial terms for the number, creating a linear, a quadratic and a cubic shape. These were removed in order of complexity if they did not improve the model fit but if a quadratic or cubic term was kept, so was the lower term(s).

All analyses started with the maximal random-effects model (Barr et al., 2013) but this was too complex so we used the Analysis of Factorial EXperiments (afex) (Singmann et al., 2019) package to enable removal of the correlation between random effects of a binary factor (*UK* vs. *abroad*). The model reported in each case is the optimal converging one, selected using the Akaike information criterion (AIC; Akaike, 1974).

3. Results

3.1. Behavioural results

The optimal multilevel model of the responses to the willingness to pay questions only included the number of people, as a fixed and random term. This showed a highly significant relationship between number of people and willingness to pay (figure 2A; t = 16.22, p < .0001). Although participants allocated more money to save larger numbers of people, the increase was not linear as the amount per person declined drastically as the group size increased (figure 2B).

Adding location (*UK* or *abroad*) did not improve the model (ΔAIC = 1, $\chi^2(1)$ = 1.26, p = .26) and was not a significant predictor when included. The lack of differentiation between the value of a life in the UK and abroad is further supported by a strong positive correlation between willingness to pay in the two locations (r = .91, p < .001).
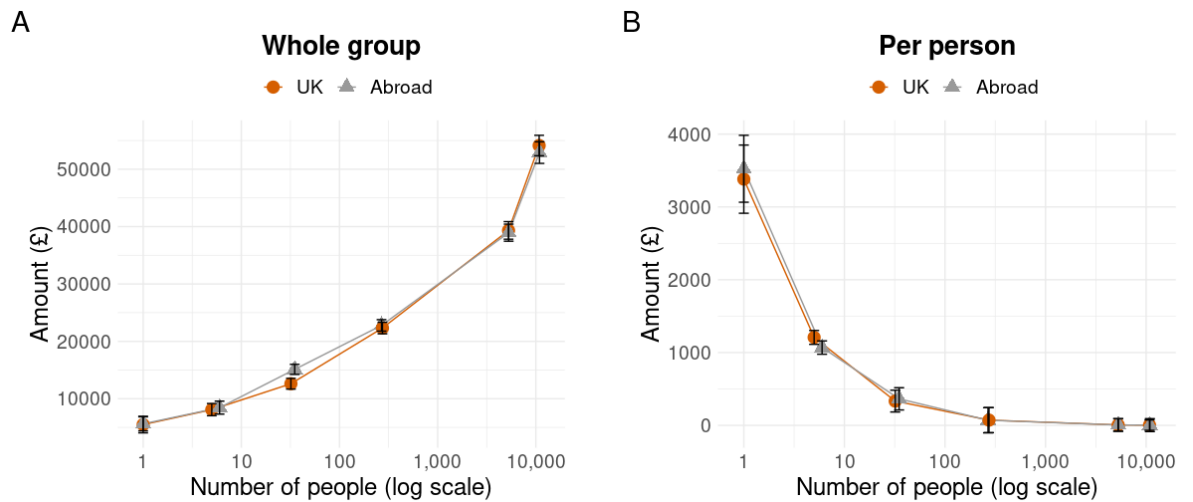

Figure 2. Amounts given on the willingness to pay task to save A: groups of different sizes (n = 186) and B: each life in groups of different sizes (excluding when the full amount was given).

3.2. Skin conductance results: presentation phase

3.2.1. People (news stories task)

The best model of SCR amplitudes at the presentation phase included main effects of trial number, location, and number of people ($\log_{10}$ scale with both linear and quadratic shapes) plus the interaction between trial number and the number of people. These factors were all included as fixed and random terms, except for the quadratic shape for the number of people as a random term.

Results showed a significant main effect of location, such that responses to learning people are at risk in the UK were larger than to learning people are at risk abroad ($t = 2.39$, $p = .017$). On a $\log_{10}$ scale, the number of people showed a significant negative quadratic relationship with SCR amplitude ($t = -2.23$, $p = .026$). As shown in figure 3, responses initially increased as the number of people increased but then declined at higher numbers of people. This quadratic effect was moderated by trial number ($t = -4.12$, $p < .0001$; see below).

In addition to analysing the amplitude parameters, we also analysed the dispersion parameters as these two measures combine to determine the area under the curve of the SCR response. For the presentation phase, the best model of dispersions was like the one for amplitudes in including main effects of trial number, number of people, and location (*UK* or *abroad*), plus an interaction between trial number and number of people. However, there was only a linear shape for the number of people, no quadratic. Each of these variables was included as a fixed and a random term.

Results revealed a significant linear relationship between SCR dispersion and the ($\log_{10}$) number of people with larger dispersion for higher numbers of people ($t = 7.14$, $p < .0001$; figure 3). There was also a significant effect of trial number, with dispersions decreasing over time ($t = -6.34$, $p < .0001$). The fixed-term interaction between these factors was not significant ($p = .20$) but removing the fixed and random interaction significantly worsened the model fit ($\Delta AIC = 15$, $\chi^2(2) = 18.65$, $p < .001$). Finally, there was no evidence for a difference between the dispersion of events in the UK, compared to abroad ($p = .81$).
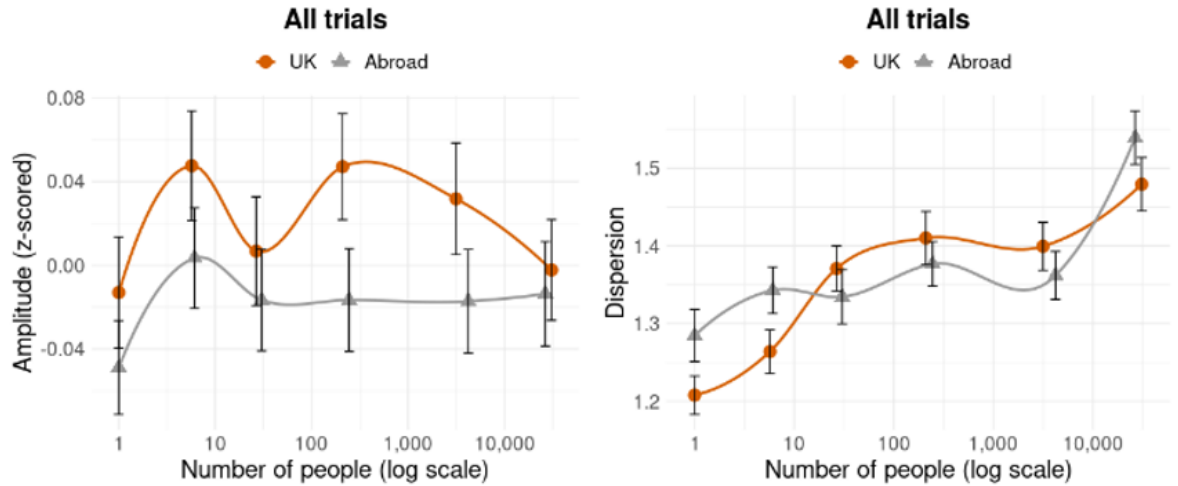
Figure 3. Left: amplitude and right: dispersion estimates of SCRs to the presentation of lives at risk in the news stories task.

To explore the interaction between trial number and the number of people for SCR amplitudes, we split the 120 trials into two halves (figure 4) and ran the model without trial number on each section. This showed a significant inverse quadratic relationship with ($\log_{10}$) number of people in the first half (t = -4.15, p < .001) of trials but not the second (p = .153; note the random term for number had to be removed to achieve convergence in the second half).

The increase in responses to lives at risk in the UK, compared to abroad, did not interact with participants' nationality (p = 0.96) or level of identification with the UK (p = 0.14). The model fit did not improve when adding nationality (ΔAIC = 4, $\chi^2(2)$ = 0.015, p = .99) or UK identity (ΔAIC = 1, $\chi^2(2)$ = 2.45, p = .29).

To test for individual differences based on IRI, we added the mean-centred total score to the model, interacting with the key variables. IRI scores showed a significant three-way interaction with the quadratic term for the number of people and trial number (t = 2.25, p = .024). To interpret this interaction, we plotted the data for high and low IRI participants separately, based on a median split, for the first and second half of the task (figure 4). This shows high IRI participants started the task with responses peaking at numbers in the hundreds then declining, whereas lower IRI participants responses initially peaked at numbers in the thousands. As the task went on, the peak for all participants shifted to numbers in the tens, although for high IRI participants there were still strong responses to the largest numbers of people abroad.
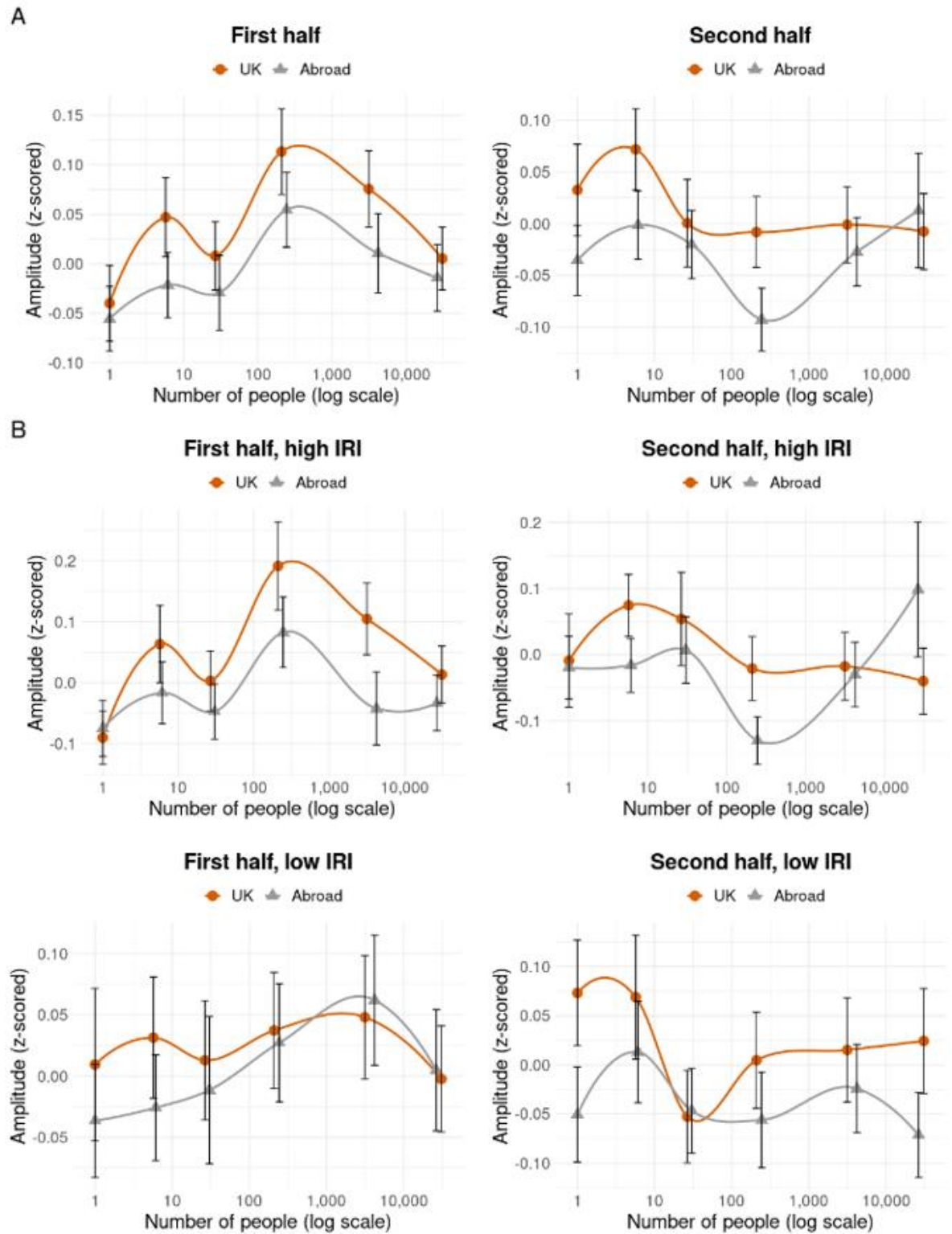
Figure 4. Effect of number of people and location on SCR amplitudes, split by A: first 60 trials and second 60 trials of the task and B: high or low IRI score (median split) and first or second half of the task.

3.2.2. Money (control task 1)

SCR amplitudes during the presentation phase when money was at risk were not significantly predicted by any factor in the model (the number of pounds – linear, quadratic or cubic terms, or trial number).

A model of SCR dispersions during responses to the presentation showed a significant linear relationship with the number of pounds (t = 3.44, p = .0005; figure 5). The best model in this analysis just included this term as well as trial number.
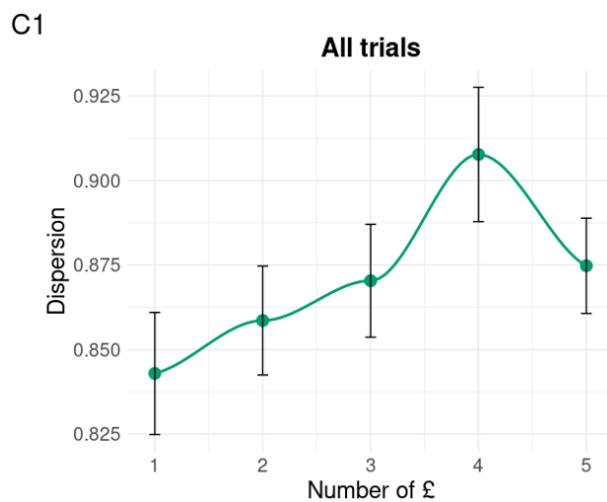
Figure 5. Relationship between the number of pounds at risk and SCR dispersion in control task 1 during the presentation phase.

3.2.3. Credits (control task 2)

The best model of SCR amplitudes for the presentation phase, during the control task with a risk of losing credits, included main effects of trial number and the number of credits (log_{10} scale), with a linear, a quadratic and a cubic term. There were also interactions between trial number and each of these shapes for the number of credits. Random terms were trial number, a linear effect of the number of credits, and the interaction between these.

Results showed that the shape of the relationship between the number of credits and SCR amplitude varied across the course of the task. There were significant interactions between trial number and both the quadratic shape (t = -2.29, p = .022) and cubic shape (t = -2.25, p = .025; figure 6). Results from the same model predicting the SCR dispersion parameters showed the same interaction between the quadratic shape for number of

credits and trial number (t = -2.46, p = .013). However, the interaction between the cubic shape for number of credits and trial number was the exact opposite to the matching term predicting amplitude (t = 2.62, p = .009). Plotting these relationships, split into the first and second half of trials, reveals the amplitudes and dispersions are the inverse of each other (figure 6).
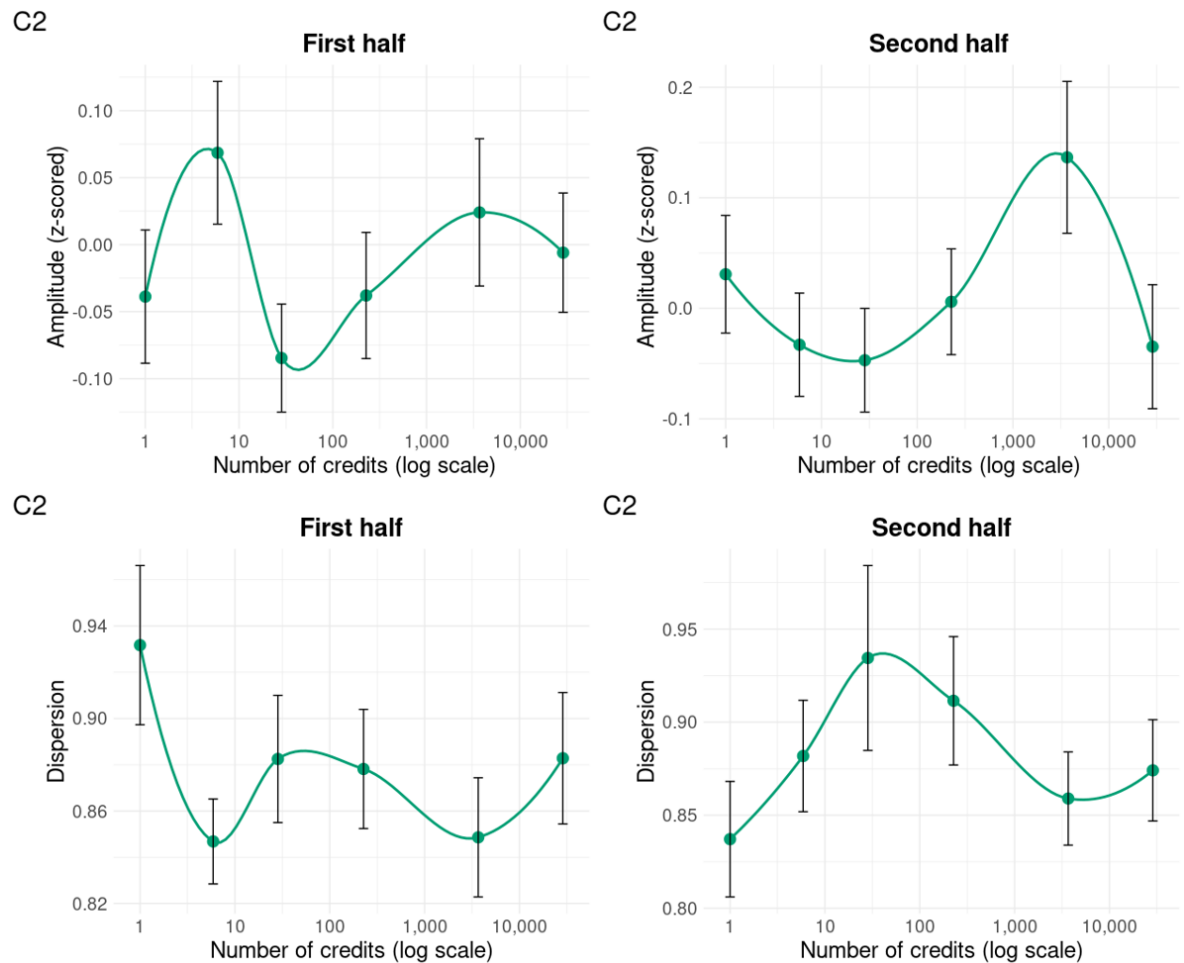
Figure 6. Relationship between SCR amplitude (top) and dispersion (bottom) and the number of credits at risk in control task 2 during the presentation phase.

3.3. Skin conductance results: outcome phase

3.3.1. People (news stories task)

The best model for SCR amplitudes to finding out whether people in the news stories lived or died included main effects of trial number, number of people (log_{10} scale; linear and quadratic), outcome valence (*positive* or *negative*), and location (*UK* or *abroad*). Each main effect was modelled as a fixed and a random term but the only interaction, between number of people and location, just a fixed term. Results revealed a significant effect of valence, with positive outcomes generating larger responses than negative ones (t = -2.16, p = .031; figure 7). The number of people showed an inverse quadratic effect on responses (t = -2.19, p = .029). The effect of abroad was not significant but removing the main effect, interaction, and random terms for this variable significantly decreased the model fit (ΔAIC = 100, $\chi^2(4)$ = 108.16, p = .0001).
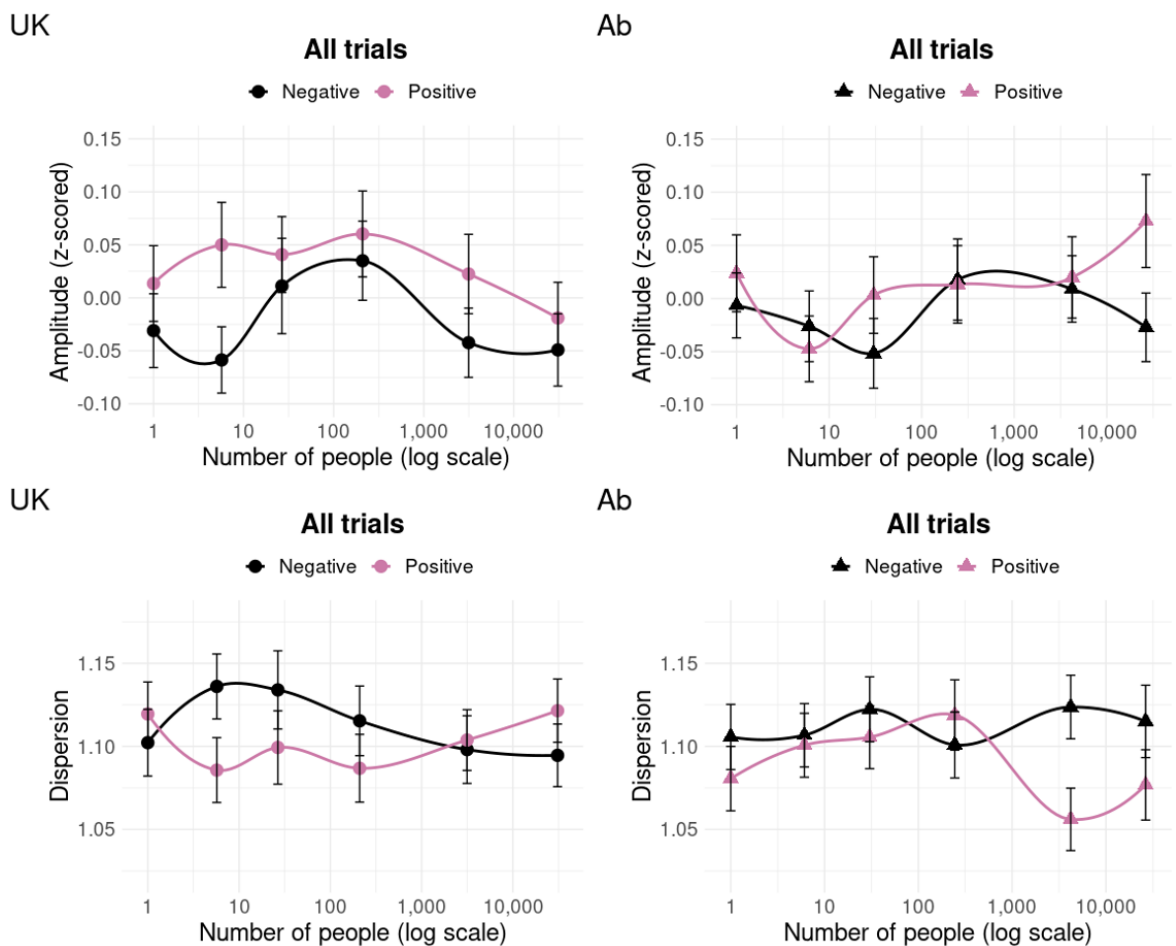
Figure 7. Effect of the number of people and outcome valence on SCR amplitude (top) and dispersion (bottom), for events in the UK (left) and abroad (right) during the outcome phase.

A model for SCR dispersions with the 4 main effects, as fixed and random terms, only showed a main effect of valence. Negative outcomes resulted in longer dispersions than positive outcomes ($t = 1.98$, $p = .047$; figure 7).

3.3.2. Money (control task 1)

Model comparison analysing the SCR amplitudes during responses to outcomes of keeping or losing money produced a model with fixed main effects of trial number, number of pounds (linear and quadratic), and outcome valence. Fixed interactions were trial number with the number of pounds and outcome valence with the number of pounds. Only random terms for trial number and outcome valence were present.

Results showed a main effect of valence ($t = -2.09$, $p = .037$) with negative outcomes generating a greater response than positive outcomes. This effect changed over the task ($t = -2.25$, $p = .025$), at the start positive outcomes lead to larger responses. The effect of the number of pounds also varied across the task with trial number interacting with both the linear relationship ($t = -1.96$, $p = .050$) and quadratic relationship ($t = -4.25$, $p < .0001$). To interpret these interactions, we plotted the relationship between amplitude and number of pounds separately for the first and second half of trials (figure 8). The dispersion of responses was not significantly predicted by any of the variables in the model.

3.3.3. Credits (control task 2)

The best model of SCR amplitudes for the outcome phase revealing whether participants had lost or kept the credits included main effects and all interactions between trial number, number of credits (linear and quadratic), and outcome valence at the fixed level. The random terms were the same except the quadratic element. The only significant effect was the interaction between the quadratic shape for the number of credits and outcome valence ($t = -3.21$, $p = .001$). This showed a positive peak at bin 3 for negative outcomes and a negative dip at bin 3 for positive outcomes (figure 8). No variables significantly predicted the dispersion of the outcome responses.
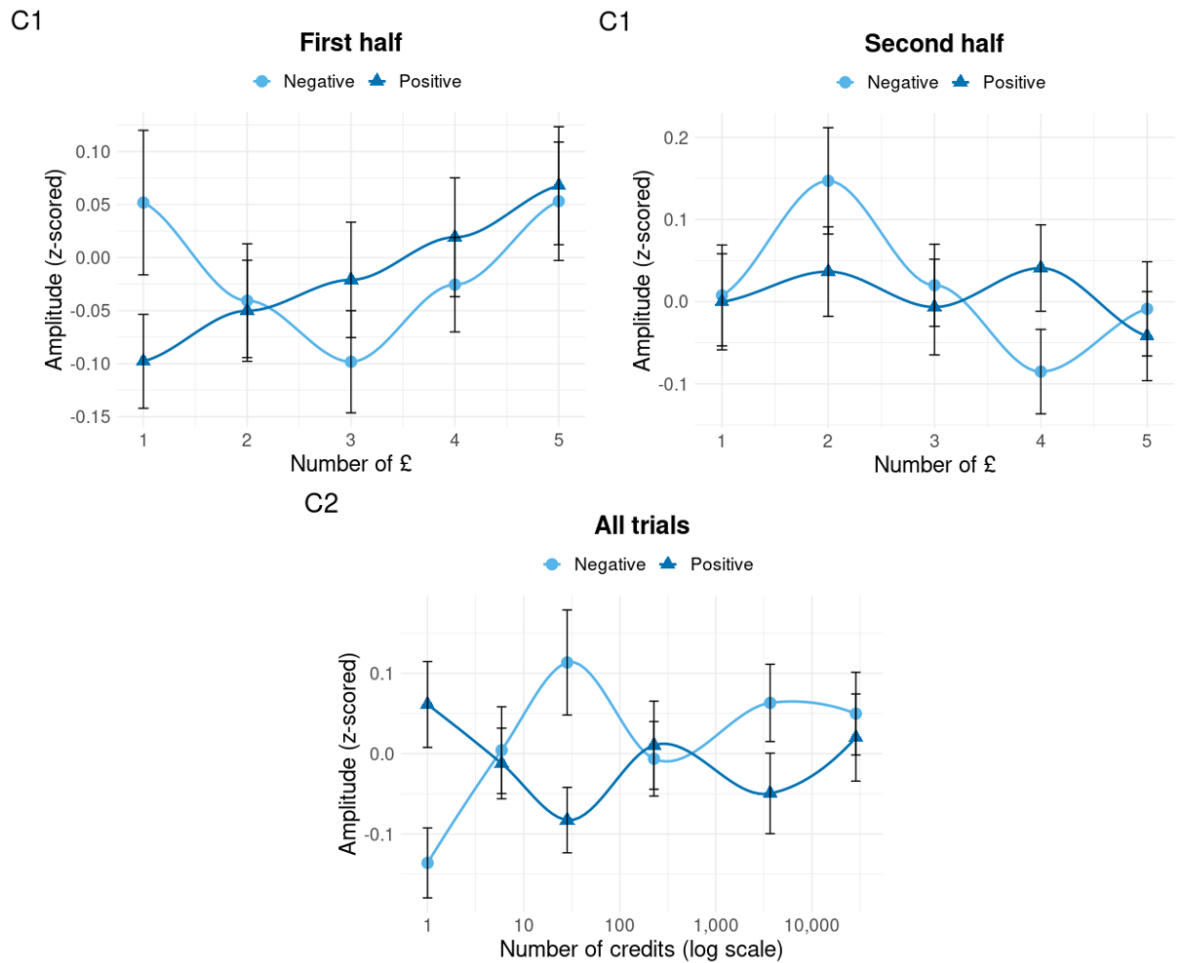
Figure 8. Top: effect of the number of pounds and outcome valence on SCR amplitude during the outcome phase of control task 1 (C1), split into the first and second half of trials. Bottom: effect of the number of credits and outcome valence on SCR amplitude across all trials at the outcome phase of control task 2 (C2).

4. Discussion

Previous research has identified biases, where the reported value of a life departs from the idea that all lives have equal value. Here we show that physiological responses reveal further biases, not captured with a behavioural measure. Responses were greater when lives at risk were in the participants' home country, compared to abroad. Responses also declined as the number of people increased above several hundred. Over time, seeing multiple stories about people dying and at risk shifted the maximal response to stories of even smaller groups – numbers in the single digits or tens of people. Interestingly, responses to financial resources being at risk or lost also showed non-linear relationships with number. We now discuss potential explanations of these effects of location and number on responses

4.1. Location

The news stories task involved a presentation phase showing a number of lives at risk, an anticipation phase, and an outcome phase revealing whether the people survived or died. Half of these events happened in the UK, where all participants lived, and the other half happened abroad. At the presentation phase, the amplitude of SCRs was higher for events in the UK, compared to abroad. This was in contrast to no evidence of a difference between the amount participants said should be spent to save lives in the UK or abroad.

Bias towards one's own geographical location could be due to identity or proximity factors. The people described in the stimuli were explicitly presented as "in" the given country rather than "from" it, to avoid priming identity and as in many cases, it is unlikely that all the people affected in a story were the same nationality.

People who are physically closer are more likely to have a shared or similar identity and a large body of work suggests a role for shared identity in promoting empathy (Azevedo et al., 2013) and prosocial behaviour (Levine et al., 2005). Social closeness has been considered as binary (ingroup vs. outgroup), as in these studies, but also along a continuum (Strombach et al., 2014). In one study, increased donations towards one's ingroup were fully mediated by perceived responsibility to help (Erlandsson et al., 2015). However, responsibility to help could not explain the effect of location in our task, which did not involve donations.

Other studies that have manipulated physical proximity, rather than identity directly, also find an effect on donation behaviours and the perceived impact of donations (Touré-Tillery and Fishbach, 2017). Increased donations following disasters in closer locations have also been linked to counterfactual thinking, the idea that "it could have been me" (Zagefka, 2018). If this counterfactual thinking occurs in relation to a past, resolved threat, it is reasonable to predict that current, ongoing threats are more salient when they are close as "it could affect me or someone I care about".

These explanations for the effect of location on prosocial responses and behaviours could be considered as complimentary or competing with identity-based mechanisms. One direct comparison showed support for identity completely accounting for proximity effects (Levine and Thompson, 2004) but it is difficult to completely rule out multiple mechanisms.

Our paradigm was not designed to distinguish between these potential explanations. If identity was the main motivation, we would have predicted that the increase in responses to stories in the UK would have been stronger for UK citizens and

those who scored higher on identification with the UK. However, these factors did not significantly improve the models of or significantly predict SCR amplitude. While this is a lack of evidence for identity motivations it cannot necessarily be considered as supporting a proximity mechanism.

4.2. Number

In addition to the effect of location, we also tested the effect of number on responses to lives and financial resources at risk and lost. When a number of lives at risk was presented and when whether they survived or died was revealed, SCR amplitudes showed an inverse quadratic relationship with the number of people on a $\log_{10}$ scale. Responses increased as the number of people rose from 1 to tens or hundreds but then declined as the number of people increased further. For SCR amplitudes at the presentation phase, this relationship changed over the course of the task and by the second half, the peak response was to numbers of people in the tens.

By using DCM to extract the SCR parameters, we were able to analyse the dispersion of responses as well as the amplitude. For the presentation phase of both the number of people and the number of pounds at risk in control task 1, dispersion increased with larger numbers. This is in line with a slower, less intense response at high numbers which are suggested to elicit less emotion than smaller groups of people (Dickert et al., 2015).

However, interpreting the decline in SCR amplitudes for responses to large numbers of people as due to a lack of empathy is difficult, given number also had a non-linear relationship with responses during trials involving financial resources. Of the amplitude measures, only positive outcomes of keeping money in the first half of control task 1 showed a linear relationship with number. For negative outcomes, the greatest responses were to small amounts.

One possible explanation for large responses to small amounts of money being lost is that participants knew that they would keep money on 50% of trials and lose money on 50% of trials. They did not know that this proportion would apply to each amount too. Keeping a small amount would become undesirable if interpreted as increasing the chance of losing larger amounts. There are also limits on how relevant results from control task 1 are for the news stories task, as the largest number was £5.

For credits in control task 2, the numbers were more closely matched to the numbers of people, going up to tens of thousands. The decline in responses to large numbers of credits could suggest large numbers generally are incompatible with fast intuitive

processing, as has been suggested in relation to people (Small et al., 2007), if this processing is necessary to produce a SCR. Mechanisms proposed to underlie scope insensitivity in valuing lives such as mental imagery and attention (Dickert et al., 2012) could also be relevant to large numbers of non-social resources, particularly when they are abstract as credits towards chances to win money are.

4.3. Outcome

In addition to number and, for people, location, the third variable we manipulated was whether the outcome was positive or negative. Negative outcomes in the news stories task were the people dying and for the control task losing the money or credits. 'Positive' outcomes were in fact maintaining the status quo – the people surviving but not necessarily free of suffering and the money or credits kept but none gained.

For outcomes affecting people, positive outcomes increased the amplitude of responses but reduced the dispersion of responses, compared to negative outcomes. At the start of control task 1, keeping money increased response amplitude, compared to losing it, but this changed during the task and over all trials, responses to negative outcomes were higher. While SCRs are linked with potential losses (Ring, 2015), previous research using gambling tasks has found greater SCRs to wins than losses (Lole et al., 2012; Wilkes et al., 2017).

4.4. Individual differences in IRI

The individual difference measure of identification with the UK is discussed above and we also measured participants' interpersonal reactivity or empathy. If this trait is related to the impact of our manipulations, it suggests the tendencies measured in the questionnaire are relevant for the mechanism of the effects. At the presentation phase, participants high and low on the IRI were differentially affected by the number of people, particularly at the start of the task, with those with higher empathy having maximal responses to a lower number of people at risk than those lower in empathy. This supports the idea that empathy is related to the decline in responses to high numbers of people at risk. Comparisons between the overall amplitude of responses between participants differing in IRI score is not relevant as the amplitudes were Z-scored and even the raw data can be confounded by irrelevant factors affecting the baseline skin conductance.

4.5. Limitations and future directions

The results presented here make several contributions to the literature, but it is also important to recognise limitations. Measuring SCRs overcomes the issues associated with

self-report but lacks a benefit of another physiological measure, heart rate, due to a lack of forward mechanism from the skin to the brain. Unlike SCR, changes in heart rate can affect cognition (Critchley and Harrison, 2013), and the ability to accurately perceive heartbeats has been linked to empathy (Grynberg and Pollatos, 2015).

Another limitation of SCR as a measure is the inability to distinguish positive from negative arousal. In tasks with clear positive and negative outcomes, such as gambling tasks (Lole et al., 2012; Wilkes et al., 2017), this is less of an issue but it complicates the interpretation of our trials in which outcomes in the control tasks may not have been perceived as independent. A potential solution is to apply fMRI to this task as there are regions of the brain more associated with positive outcomes and others more associated with negative outcomes, although these still overlap (see paper 4).

Considering the task, there is a risk that participants' expectations about the outcome of the event could have affected responses. This could have confounded our results if it was more relevant, for example, to the large numbers of people in the UK, stories which participants may have been more aware of. During the debrief, we asked participants whether there were stories that they remembered so knew the outcome before it appeared. These trials were too few to analyse separately. For half of the participants the outcomes were reversed – the outcome said the people died when in real-life they survived. Averaging across participants should therefore control for the effect of surprise and did not seem to affect whether or not participants believed the task.

Finally, a further issue with the task is the lack of detail or images associated with the stories, which would likely be present in real-world situations, and the large number of trials. However, removing detail was necessary to standardise the trials across conditions and having many trials allowed us to quantify changes in the effect of our variables over time. Our efforts to balance conditions on number, proportion of positive and negative outcomes, and ratings from pilot participants on experienced empathy and emotion provide a novel stimulus set we hope will be useful in future research.

5. Conclusion

Physiological responses to lives at risk and the outcome of whether the people survived showed biases, not revealed in a behavioural measure. Responses were greater when the people were in the same country as the participants, compared to abroad. Responses also showed scope insensitivity. As the number of people increased, responses did too, but only up to a certain point, then they declined for the largest numbers. Scope insensitivity also got worse over time and was more pronounced in participants high in trait empathy.

The neural basis of valuing the lives of others: value depends on the number of people and geographical location

Jo Cutler and Daniel Campbell-Meiklejohn

The Social Decision Lab, the University of Sussex, School of Psychology, Falmer, UK.

1. Introduction

Every day we make decisions that impact other people. Often these have minor consequences, but some choices determine who and how many people live or die. Those with influence over policy and resource allocation make such decisions regularly but everyone with a vote or the power to donate can also have an impact. To understand these decisions, it is crucial to understand how people value the lives of others.

Value is often measured using decisions between different options. However, valuing the lives of others is likely to be determined by the cognitive and emotional response to the prospect of other people dying (Li et al., 2010; Olivola, 2015). Physiological measures such as skin conductance (paper 3) and functional magnetic resonance imaging (fMRI) offer an alternative to decision-making tasks. These techniques allow us to measure responses to lives at risk, people dying and people surviving without the issues of social desirability associated with self-report (Fernandes and Randall, 1992). Contrasting neural responses to lives being lost, compared to people surviving, is one way of quantifying value. We can also measure the impact of factors that may increase or decrease the value of a life.

According to egalitarian normative moral perspectives, there should be no differences between the value of individual lives (Dickert et al., 2012). These egalitarian principles describe how we ought to value lives and are reflected in many modern societies. For example, the Universal Declaration of Human Rights (Article 1) states "all human beings are born free and equal in dignity and rights". However, research shows choices are inconsistent with this equal value (Slovic, 2000) and are prone to biases. Here we present results on the neural correlates of two such biases, relating to the number of people and their location.

One bias affecting the valuation of human lives is the number of people at risk or in need. When responses do not scale linearly with the number of people, each life is not valued equally. This is known as scope insensitivity (Dickert et al., 2015) or psychophysical numbing (Fetherstonhaugh et al., 1997). A common pattern, shown in previous literature, is that the value of each additional life decreases as the number of people increases, creating relative apathy towards large-scale humanitarian catastrophes (Dickert, et al., 2015).

There are also situations where emotional and behavioural responses decrease as the number of people increases. A striking example is the singularity effect, donations to one person in need were higher than to two or more people (Västfjäll et al., 2014). At the opposite end of the spectrum is the 'collapse of compassion' when the number of people suffering gets very large (Dickert et al., 2012). This effect has been explained through an

inability to empathise with, and feel emotional about, large numbers of people suffering (Slovic, 2007). A competing explanation is that we are able to comprehend mass suffering, but are motivated to avoid doing so and regulate our emotions to prevent being overwhelmed (Cameron and Payne, 2011).

A second bias that affects sympathy and valuation of human lives is proximity, both social and geographical. Proximity tends to strengthen responses towards people in need whereas distance tends to decrease it (Molenberghs, 2013; Strombach et al., 2015). Potential mechanisms for this effect include those due to identity factors, such as similarity (Loewenstein and Small, 2007) or being part of an ingroup (Meier, 2011), and physical proximity, such as the threat of being personally affected (Zagefka, 2018).

While these biases have been established in the behavioural domain, research on their neural basis is lacking, particularly for scope insensitivity. Neuroimaging allows us to see similarities and differences between the processing of different stimuli. For example, whether the valuation of people's lives is associated with the same areas as the valuation of non-social resources. Unlike physiological measures such as skin conductance that give a single measure of overall arousal, different regions of the brain may respond to stimuli differently.

To apply these benefits of fMRI to the topic of valuing the lives of others, we measured activation during a novel 'news stories' task. This presented information about different numbers of people at risk of death in the UK and abroad and then whether they died or survived. Matched non-social control trials presented different numbers of credits, which translated to chances to win money, at risk and whether they were lost or kept. Both the social and control trials were in a negative frame with the 'positive' outcome a lack of loss: surviving the risk of death or keeping the credits.

To our knowledge, this is the first study to apply fMRI to scope insensitivity. A previous fMRI study on the effect of identifiability showed increased activity in the nucleus accumbens (NuAcc) when donating to a beneficiary identified with a photo (Genevsky et al., 2013). Identifiability often coincides with the singularity effect but the two are distinct (Lee and Feeley, 2016). Increased NuAcc activity suggests helping an identified person is particularly rewarding. Regions active during rewards for the self are also active when observing others receive rewards (Morelli et al., 2015). However, this positive aspect is less likely to be relevant in our task that uses a negative frame to align with most news reports.

Previous work using fMRI on a range of related topics provide hypotheses for our work. Relevant to our key question, how the lives of others are valued, is the concept of

subjective value. This has been shown to depart from absolute value in the non-social domain and have different value functions in the gain and loss domains (Kahneman and Tversky, 2013). Across gain and loss frames, a meta-analysis linked the subjective value of outcomes to activity in AI, dorsomedial prefrontal cortex (dmPFC), striatum, and thalamus (Bartra et al., 2013). The authors suggest activations could represent salience or arousal.

Other relevant topics include empathy for pain and risk. A meta-analysis on the neural correlates of risk, uncertain outcomes with known probabilities, identified the AI as a key region, across paradigms involving choice or no choice (Feldmanhall et al., 2016). In our task, we defined the probabilities by telling participants that there were equal numbers of positive and negative outcomes for both the social and non-social elements.

Given the focus on other people suffering and negative frame, the most relevant concept from previous fMRI research is vicarious pain or empathy for pain. Experiencing empathy activates a network, including bilateral anterior insula (AI) and anterior cingulate cortex (ACC), that overlaps with regions involved in directly experiencing pain (Lamm et al., 2011). In comparisons between different people suffering, greater activations in AI were shown during empathy for pain for ingroups, compared to outgroups (Hein et al., 2010).

We determined regions of interest based on the fMRI meta-analysis on empathy for pain, as this provided a contrast closest to the current task: responses to others suffering (Lamm et al., 2011). The key regions identified in this meta-analysis for the contrast other pain > other no pain were the ACC and bilateral AI. Increased activity in these regions was reliable across studies and was also related to self-pain. These three regions form the focus of our analysis as they have also been associated with other relevant processes such as risk and value, as described above.

2. Method

2.1. Participants

26 healthy participants (13 female, 13 male, aged 20-37, mean = 26.69, SD = 5.29) took part in the study. All participants were right-handed and UK residents, born in the UK and having lived most of their lives in the UK. One female participant was excluded due to scanner issues, leaving 25 participants for analysis (12 female, 13 male, aged 20-37, mean = 26.56, SD = 5.35). Exclusion criteria in addition to those required for MRI research were dyslexia; dyscalculia; any mental health issues or addictions; having studied psychology,

economics, maths or physics at degree level or above; and being born or living for more than half one's life outside the UK. All participants identified themselves as British, except for one who identified as Irish.

Of the 25 participants in the analysis, 13 were currently students, 11 were in employment and 1 was unemployed. 11 had completed undergraduate degrees, 1 had completed an MSc and for 13 their highest completed qualification was A-levels. Politically, the participants were more representative of the local area than the country, with all but two reporting voting 'remain' in the Brexit referendum (1 did not vote, 1 preferred not to say). When asked who they would vote for if a general election were tomorrow, 12 answered Green, 6 Labour, 3 Liberal Democrat, 2 didn't say, 2 would not vote. Participants were compensated for their time with a payment of £30, which they could give some or all of to charity, and entry into 3 prize draws linked to the tasks (see below). The procedure was approved by the Brighton and Sussex Medical School Research Governance and Ethics Committee [ER/JC620/5].

2.2. Stimuli & measures

2.2.1. News stories stimuli

Stimuli for the news stories were taken from real events in which varying numbers of people were at risk of dying in the UK and abroad. Each was summarised with just the number of people, country, and cause of death. The cause types were natural disasters, accidents, violence, diseases, or long-term issues. Participants were informed before the task that in half the stories the people survived but in the other half they died. The 20 conditions were created with a 5 (**number bin**) x 2 (**location**: *UK* or *abroad*) x 2 (**outcome**: *positive:* alive or *negative*: dead) design (see table 1). There were 10 stories in each condition, meaning 200 in total.

Stories across these conditions were balanced on the number of each cause type and on emotion strength, emotion valence (positive or negative) and experienced empathy for the people involved, based on ratings from independent pilot participants. For these ratings, the number of people and the country were hidden to isolate the cause of death. Ratings were obtained for sentences summarising both the presentation (for example "X people were identified at risk of death from a fire") and the outcome (for example "X people survived breast cancer").

2.2.2. Non-social control stimuli

Stimuli to provide non-social control trials involved a risk of losing credits that translated to the participant's chance to win one of two £50 prizes. Participants started with 1,000,000 credits in their "Pot", but on each trial, some could be lost. We also informed participants there was a 50:50 chance of keeping or losing the credits. The 100 non-social trials were in 10 conditions with a 5 (**number bin**) x 2 (**outcome**: *positive:* keep or *negative*: lose) design (see table 1), with 10 trials per condition. To match the different causes of death, credits also had different causes of being lost, for example "stolen" or "decayed".

Table 1
Design for the stimuli shown in the scanner

Number bin	People				Credits	
	UK		Abroad			
	Alive (+)	Dead (-)	Alive (+)	Dead (-)	Keep (+)	Lose (-)
1	1 person		1 person		1 credit	
2	2 – 10 people		2 – 10 people		2 – 10 credits	
3	11 – 50 people		11 – 50 people		11 – 50 credits	
4	51 – 500 people		51 – 500 people		51 – 500 credits	
5	501 + people		501 + people		501 + credits	

Note. (+): positive outcome, (-): negative outcome.

2.2.3. Task structure

In the main task, the 200 news stories and 100 non-social control tasks were intermixed. This created 30 conditions with a 3 (**trial type**: *credits* or *people* (**location:** *UK* or *abroad*)) x 5 (**number bin**) x 2 (**outcome**: *positive* or *negative*) design. Trial order was determined for each participant pseudo-randomly by dividing the 300 trials into 10 blocks of 30 trials that each had one trial from each of the 30 conditions (see table 1). The order of conditions within each block was then randomised, as was the order of blocks, and this sequence presented with no breaks or change between blocks.

The social news stories trials were presented in the form:

> [number] people
> [country]
> [cause]

The non-social control trials were presented in the form:

> [number] credits
> Your Pot
> [cause]
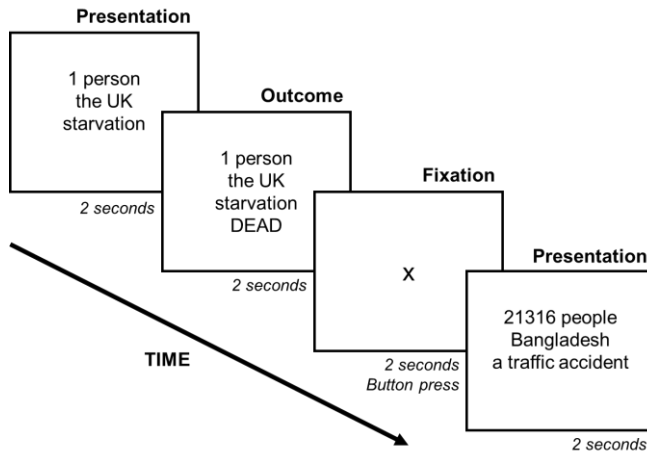
Figure 1. Structure of the task. The fixation cross was either a + or a x and the participant pressed a button to indicate which. The timing was kept constant regardless of the button press.

This **presentation** sentence appeared for 2s. Then the **outcome phase** showed either "ALIVE" or "DEAD" for news stories and "KEEP" or "LOSE" for credits appeared below and all the information remained on the screen for 2s. A **fixation** period of 2s preceded the next trial, meaning each trial took 6s (figure 1).

To ensure and measure concentration during tasks, participants were asked to indicate whether the fixation cross was a + or a × by pressing one of two buttons. The type of cross was randomised on each trial. This data was lost for four participants but monitoring during the scan showed they had responded on most trials.

2.2.4. *Willingness to pay*

To provide a behavioural measure of the value of a life in each condition, we adapted a willingness to pay measure (Frederick and Fischhoff, 1998). Participants were asked to imagine they were a "UK government minister in charge of allocating resources to stop people dying from flooding both in the UK and abroad". Participants then filled in the amount, from £0 - £100,000, to allocate to 12 projects, 6 in the UK and 6 in a named foreign country, to save different numbers of people. The numbers corresponded to the average numbers in each condition of the fMRI stimuli but with an additional bin to split the top bin from the fMRI stimuli in two (approximately 5,000 and approximately 10,000 people).

Instructions read "You must consider the following proposals and allocate funding. For each project, there is a maximum of £100,000 available which you should allocate if the project is very important. You must not spend unnecessary money because the money can also go to other worthy causes. If a project is not worthwhile you can decide to give it £0. However, you cannot use the money you save on one project now for another one of the projects listed, imagine any money you save would be used on different future projects."

2.3.Procedure

Following informed consent and MRI safety checks, participants indicated how much they wanted to know the outcome for a news story representative of each condition. They then received instructions for the task in the scanner, including some example trials and completed most of the trait measures.

In the scanner, stimuli were presented in MATLAB (2018a, Mathworks, Inc., Natick, MA) using Psychophysics Toolbox extensions version 3 (Brainard, 1997) on a screen viewed via a mirror mounted on the head coil. Participants categorised the fixation cross using two buttons on an MR-safe response box. The task lasted 30 minutes so combined with the field maps and structural scan participants were in the scanner for approximately 45 minutes.

After the scan, participants rated all the news stories, including their outcome, in a random order on two 0-100 scales. One measured surprise: "In the scanner, how surprising did you find this?" [Not at all surprising – Very surprising] and the other emotion: "In the scanner, how did you feel about this?" [Very negative – Positive]. Participants then completed final trait measures, including the willingness to pay questions. All measures outside the scanner were completed through Qualtrics Online Survey Software (Qualtrics, Provo, UT).

During the debrief, we asked participants whether there were any news stories for which they had a personal link to either the country or the cause involved, or any that during the scan they recognised and remembered from the news. We also asked whether participants believed that the news stories were real and categorised responses between 1 (did not believe) to 4 (believed) with 2 and 3 representing major and minor doubts respectively. No participants reported not believing the stories at all, 19 believed them completely, 4 had only minor doubts, 1 had major doubts and for 1 participant this was not recorded.

2.4. fMRI data acquisition

Data were acquired using a Siemens Prisma 3T scanner (Siemens Medical Solutions, Erlangen, Germany) fitted with a 32-channel head-coil at the Clinical Imaging Sciences Centre, Brighton and Sussex Medical School. Field maps were collected in the anterior–posterior and posterior–anterior directions to adjust for distortions. Functional T2-weighted echo-planar images were acquired in an interleaved order with multiband factor 4 at a 30° angle to AC-PC. Each volume contained 52 slices with a TR of 1.5s, TE of 37ms, 52°

flip angle, and voxel size of 2.19*2.19*2mm. Finally, whole-brain high-resolution T1-weighted structural images were collected using the MPR sequence with the following parameters: 208 slices, 0.8mm^3 voxel size, 2.4s TR, 2.22ms TE, 8° flip angle, 6.38 minutes total time.

2.5. Analysis

We used several tools from FMRIB's Software Library (FSL, version 6.0) (Jenkinson et al., 2012; Smith et al., 2004) during preprocessing and FEAT for first (Woolrich et al., 2001) and group-level analysis (Woolrich et al., 2004).

2.5.1. Preprocessing

The susceptibility-induced off-resonance field was estimated by applying topup (Andersson et al., 2003), as implemented in FSL, to the two images with reversed phase-encode blips, so distortions in opposite directions. The resulting image was used in preprocessing to unwarp the functional data. We extracted the brain using BET, FSL's Brain Extraction Tool (Smith, 2002). During preprocessing, we applied FMRIB's Image Registration Tools MCFLIRT (Jenkinson et al., 2002) to correct for motion and FLIRT (Jenkinson and Smith, 2001) to register EPI images to an expanded functional image (full search, 6 DOF) and then to the participants' structural image (full search, BBR). Registration of the structural to standard space (Montreal Neurological Institute, MNI152) used FNIRT (Andersson et al., 2010) nonlinear registration (full search, 12 DOF).

We cleaned the unsmoothed data by removing components classified as noise following independent components analysis with Multivariate Exploratory Linear Optimized Decomposition into Independent Components (MELODIC; Beckmann and Smith, 2004). Classification was automated using FMRIB's ICA-based Xnoiseifier (FIX; Griffanti et al., 2014; Salimi-Khorshidi et al., 2014) trained on manually classified components from the same dataset and we checked all classifications. We also generated confound files for motion outliers with FSL's Motion Outliers and used these to remove the volumes from analysis. Smoothing with a 5mm full-width half-maximum Gaussian kernel was applied to the denoised data.

2.5.2. First level

We fitted general linear models (GLMs) for the event-related parameters to the data for each participant. These were optimised for the outcome phase, with onsets at the point the outcome appeared, although this was only 2 seconds after the presentation, so these events cannot be separated. The event duration was 2s. Trials for news stories that

participants reported a link to or remembered were put into a confound regressor. We created a separate confound regressor for trials on which the participant did not respond to the classification of the fixation cross at the end of that trial. Regressors and temporal derivatives were convolved with the default hemodynamic response function in FSL (gamma function, delay: 6s, SD: 3s).

Event files combined trials within each condition: **number bin** (*1-5*) x **trial type & location** (*UK*, *abroad* or *credits*) x **outcome valence** (*positive* or *negative*). 3 versions of each were created: a) with no parametric modulator, b) with surprise ratings as a parametric modulator and c) with emotion ratings as a parametric modulator. For the whole-brain analyses, we used contrast coefficients that created different shapes across the number of people or credits. These used the number bin (1-5), creating an equal distance between each bin.

2.5.3. Regions of interest

To create a-priori regions of interest (ROIs) we used peaks identified in an fMRI meta-analysis (Lamm et al., 2011) for vicarious pain responses. We focused our analysis on the largest 3 clusters: the medial anterior cingulate cortex (ACC: -2, 23, 40); left anterior insula (AI: -40, 22, 0); and right AI (39, 23, -4). For each ROI, we extracted parameter estimates from a 5mm sphere, centred on the peak, for each condition.

We analysed the parameter estimates from the ROIs using lme4 (Bates et al., 2015) and afex (Singmann et al., 2019) in R (R Core Team, 2017) with RStudio (RStudio Team, 2015). Each started with the maximal model (Barr et al., 2013), but these did not converge so random-effect structures and some interactions between fixed terms were removed to balance convergence and model fit, assessed by comparing the Akaike information criterion (AIC; Akaike, 1974). The best model varied for each analysis so is reported alongside the results.

Factors in the model were the average **number** for that bin (of people or credits; transformed with $\log_{10}$), **outcome valence** (*positive* or *negative*) and either a) **trial type** (*people* or *credits*) or b) **location** (*UK* or *abroad*). For the number of people, we included a linear and a quadratic term based on findings from behavioural research that responses to increasing numbers of people may initially rise and then decline again (Dickert et al., 2012).

We analysed answers to the behavioural willingness to pay questions with a similar technique to the ROIs – a mixed model with predictors of **number of people** ($\log_{10}$ scale) and **location** (*UK* or *abroad*).

2.5.4. Whole-brain second level

Group-level models were fitted to parameter estimates from each of the lower-level analyses, using FMRIB's Local Analysis of Mixed Effects (FLAME) 1+2 (Woolrich et al., 2004). We used cluster-corrected thresholding with a voxel threshold of $Z > 3$ and cluster threshold of $p < .05$. to protect against false positives (Eklund et al., 2016). Labels for the regions identified in these results were generated using the Harvard-Oxford cortical and subcortical atlases in FSL.

3. Results

3.1. Willingness to pay

The maximal model included a main effect of number and location and interaction between these factors, both as fixed and random terms. This converged and was made worse by removing any elements. The linear relationship between the total amount given and the number of people (on a $\log_{10}$ scale) was significant ($t = 8.70$, $p < .0001$; figure 2A). However, when calculating the amount per person, responses showed scope insensitivity. An average of £5467.37 was given to save one person but £6.47 for each person in groups of over 10,000 people (these averages exclude two participants who gave the maximum on all questions; figure 2B).
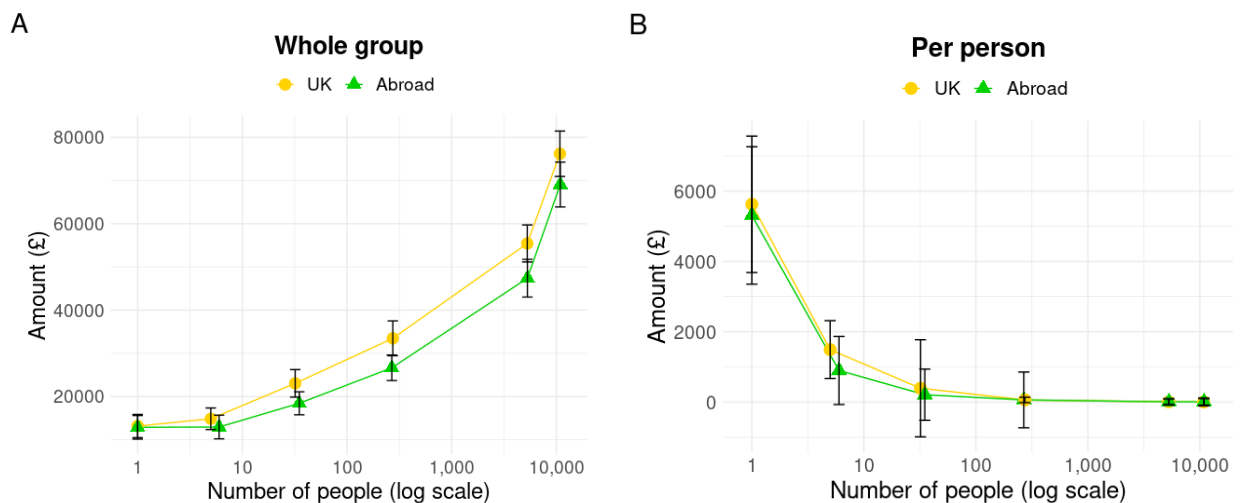
Figure 2. Amounts given on the willingness to pay task to save A: groups of different sizes (*n* = 25) and B: each life in groups of different sizes (excluding when the full amount was given).

Across participants, more was given to save lives in the UK than abroad, but this effect was not significant (t = 1.97, p = .06). Location also did not interact significantly with number (t = 1.26, p = .22). Between participants there were large differences in the effect of location. 12 participants gave more to projects in the UK (up to 7.6 times more), 5 gave more to projects abroad (up to 4.3 times more) and 8 participants gave the same amount between locations.

3.2. People vs. credits

3.2.1. ROIs

A significant main effect of whether the trial was social or non-social was only found in the right AI of the three ROIs statistically analysed (t = 2.25, p = .034). This showed greater activity for outcomes affecting participants' credits than other people. Activity in the right AI was also higher for negative, compared to positive, outcomes when averaging across people and credits (t = 2.22, p = .027). In the ACC, there was a significant interaction between outcome valence (*positive* or *negative*) and whether the trial was social (*people* or *credits; t* = -2.14, p = .033) showing a greater differentiation between positive and negative outcomes for people, compared to credits (figure 3).

Table 2
Results from mixed-models analysis of the 3 ROIs testing the effect of trial type (people or credits), outcome valence (positive or negative), number (log$_{10}$), and interactions between these factors.

	ACC		Left AI		Right AI	
	t	p	t	p	t	p
Intercept	-6.15	<.001*	-1.93	.065	-5.62	<.001*
Number (linear)	3.27	.001*	2.38	.018*	2.70	.007*
Number (quadratic)	-1.70	.101	0.19	.849	-0.04	.970
Credits – people	0.50	.619	-1.37	.182	2.25	.034*
Negative – positive	1.50	.145	1.84	.066	2.22	.027*
Number (linear) x type	3.58	<.001*	1.20	.232	0.79	.431
Number (quadratic) x type	-0.35	.725	-0.13	.893	0.04	.971
Number (linear) x type	1.01	.312	0.45	.656	1.05	.292
Number (quadratic) x type	0.30	.764	0.45	.652	0.27	.789
Type x valence	-2.14	.033*	-1.61	.107	-0.92	.360
N (linear) x type x V	-1.32	.188	-1.27	.206	-0.68	.498
N (quadratic) x type x V	-1.90	.057	-1.28	.201	-0.93	.354

*Note. x: interaction, N: number, L: location, V: valence, blank cells: the variable was not included in the best fitting model, *: p < 0.05*
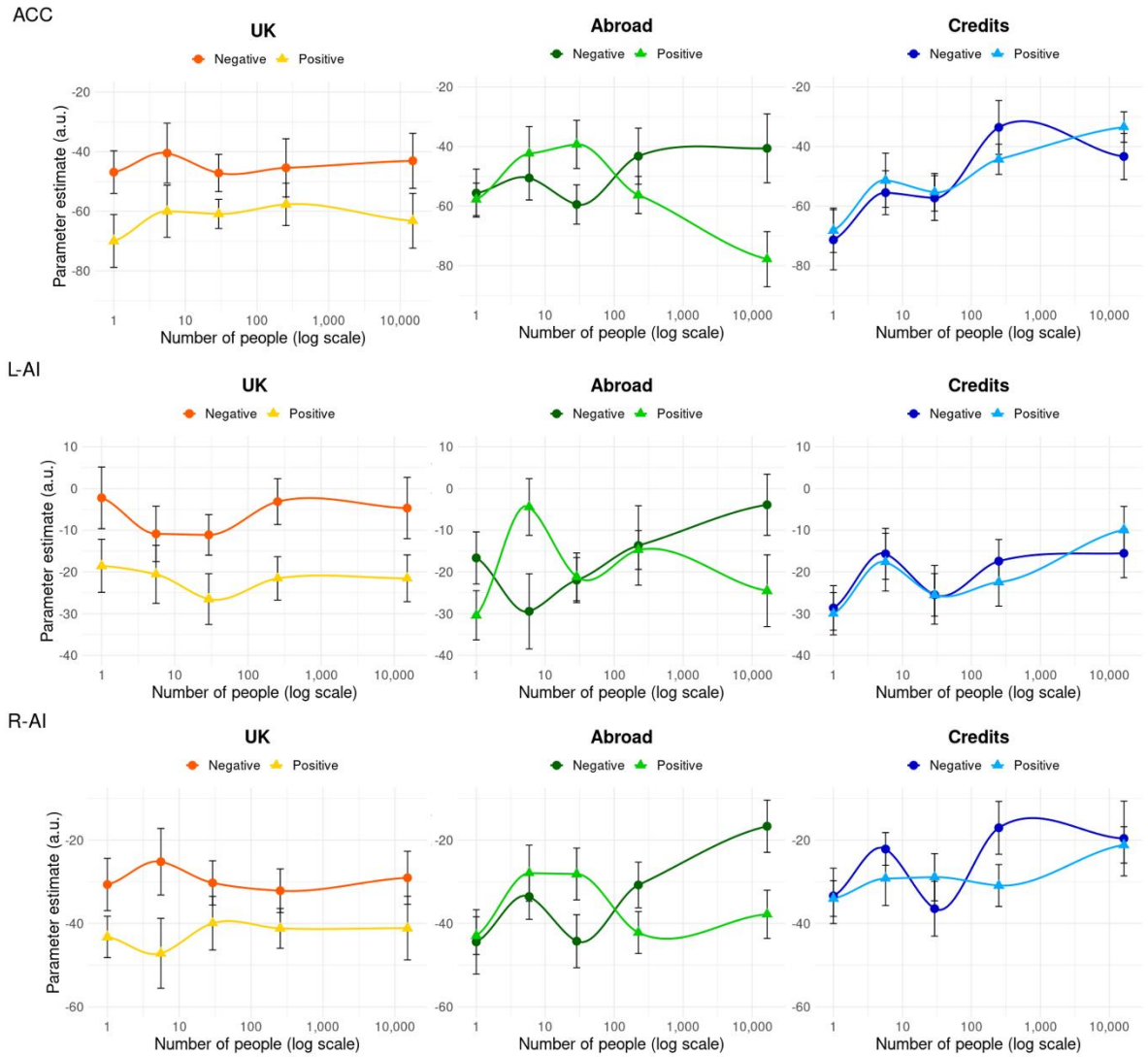
Figure 3. Average estimated activity in arbitrary units (a.u.) for each condition from the 3 main ROIs – ACC: anterior cingulate cortex (top row), L-AI: left anterior insula (middle row) and R-AI: right anterior insula (bottom row).

Regarding number, analysis of the ACC, left AI and right AI, across trials involving people and credits, revealed a significant linear relationship between responses and the ($\log_{10}$) number of people in all three regions (see table 2). However, in the ACC this relationship significantly interacted with whether the trial was social, with the linear effect of number stronger for credits than people ($Z = 3.58$, $p = .0004$).

3.2.2. Whole brain

Results from the ACC, right AI and left AI showed responses to people and credits differed in overall levels of activation, differentiation between positive and negative outcomes and relationship with number. We then looked at these contrasts for the whole brain to see whether other regions showed the same results as the ROIs.

Greater responses to outcomes affecting people, compared to credits, were shown in regions including the precuneus, vmPFC, orbitofrontal cortex (OFC), right hippocampus, and left insula. In contrast, increased activity during outcomes affecting credits were shown in the bilateral frontal and posterior supramarginal gyri, dorsolateral prefrontal cortex (dlPFC), and distinct areas of the precuneus and insula (table 3).

Table 3

Cluster peak details from the whole-brain contrast of people vs. credits

Label	Voxels	Z-score	p	X	Y	Z
People > credits						
Precuneus cortex	9068	6.40	<1.0e-30	-4	-58	8
Frontal medial cortex	3375	6.26	1.4e-30	-2	38	-20
Frontal orbital cortex	2719	6.07	2.36e-26	-42	32	-16
Right hippocampus	628	5.36	2.39e-09	24	-18	-18
Precentral gyrus	332	5.57	8.58e-06	-54	-10	50
Central opercular cortex	199	4.56	0.000745	-36	2	12
Temporal fusiform cortex (posterior)	135	4.34	0.00908	36	-36	-26
Cerebellum vermis VI	124	4.02	0.0145	0	-74	-26
Lateral occipital cortex (superior)	121	5.31	0.0164	-36	-78	40
Precentral gyrus	120	4.30	0.0172	-22	-10	64
Cerebellum right IX	96	4.64	0.0499	4	-56	-44
Credits > people						
Middle frontal gyrus	4381	5.25	1.46e-36	46	14	40
Supramarginal gyrus (posterior)	4348	6.03	2.25e-36	56	-46	48
Precuneus cortex	1903	5.52	1.4e-20	10	-78	46
Cerebellum left crus II	1322	5.44	6.06e-16	-40	-62	-44
Supramarginal gyrus (posterior)	1007	5.12	3.8e-13	-56	-44	46
Frontal pole	407	5.31	8.94e-07	-40	50	6
Paracingulate gyrus	385	4.12	1.73e-06	10	42	18
Cingulate gyrus (anterior)	380	4.83	1.97e-06	6	-14	28
Insular cortex	190	4.39	0.00104	40	12	-12
Temporal occipital fusiform cortex	163	4.25	0.00293	30	-52	-6
Superior frontal gyrus	152	3.88	0.00453	-24	6	58

Note. X, Y & Z: coordinates in MNI space.
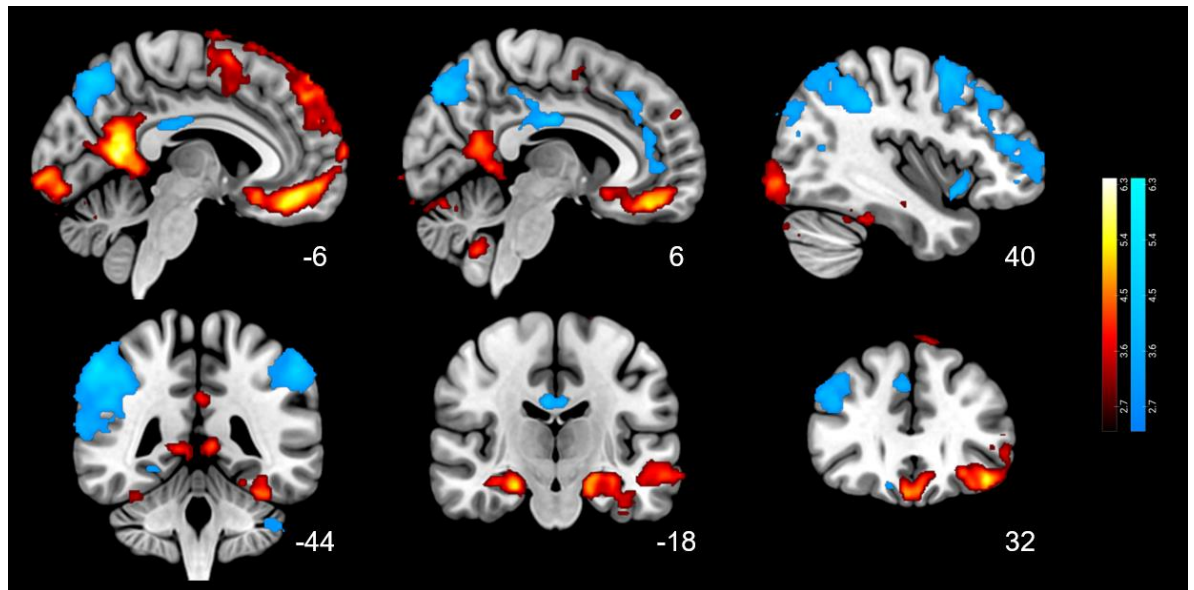
Figure 4. Contrast between people and credits (all numbers and outcomes). Orange: people > credits; blue: credits> people. Images in radiological orientation (right = left), coordinates in MNI space (mm), threshold Z > 3.

For trials involving people, several regions differentiated between positive and negative outcomes (discussed below) but this was not the case for credit trials. The interaction between trial type (*people* or *credits*) and outcome valence (*positive* or *negative*) only showed a small cluster in the occipital fusiform gyrus (-32, -82, -14) where positive and negative outcomes for people were differentiated more than for credits.

A positive correlation between activity and number bin (1-5) was found for credits in ACC, OFC, frontal gyri and occipital areas. Analysing positive and negative outcomes separately for credits showed similar results and no regions showed an interaction between number and valence. These results from the whole-brain analysis support the choice of ROIs as key areas involved during the task. The specific area of the ACC covered by the ROI is representative of a broader area where activity correlates with the number of credits but is insensitive to outcome valence.

In contrast, responses to outcomes involving people do not show this pattern. Few regions showed a positive correlation with the number of people. These were mostly in occipital areas, likely reflecting the visual aspect of larger numbers. The exception was activity in bilateral insula which correlated with number when people died abroad. In ACC and dlPFC, activity increased with number significantly more for credits than people (table 4 & figure 5).

Table 4

Cluster peak details from the whole-brain analysis of positive correlation with number bin

Label	Voxels	Z-score	p	X	Y	Z
Credits						
Paracingulate gyrus	2405	4.82	1.15e-24	0	42	28
Occipital pole	614	4.26	2.27e-09	22	-100	-8
Occipital pole	436	4.03	2.98e-07	-12	-104	-10
Supramarginal gyrus (posterior)	233	3.83	0.000181	-58	-52	42
Frontal pole	188	3.95	0.000948	-46	42	16
Inferior frontal gyrus, pars opercularis	169	3.95	0.00198	-54	18	4
Superior frontal gyrus	142	3.95	0.00594	12	18	66
Cerebellum right crus I	141	4.19	0.0062	34	-62	-30
Frontal orbital cortex	114	4.09	0.0199	-38	20	-18
Frontal orbital cortex	105	4.04	0.0299	34	20	-12
Cerebellum left crus I	104	3.82	0.0313	-38	-62	-26
Occipital pole	103	3.95	0.0328	-34	-92	0
Left thalamus	101	3.89	0.036	-2	-20	12
People						
Occipital pole	2720	5.01	2.42e-27	26	-92	-12
Lateral occipital cortex (superior)	192	4.10	0.000686	-24	-70	40
Inferior temporal gyrus (temporooccipital)	95	3.81	0.0433	48	-52	-14
Lateral occipital cortex (superior)	95	3.82	0.0433	32	-76	42
Credits > people						
Paracingulate gyrus	562	8.25	5.68e-09	0	38	30
Frontal pole	145	3.96	0.00447	-26	46	38
Frontal pole	128	3.80	0.00933	30	54	18
Abroad negative						
Occipital fusiform gyrus	254	3.95	0.000131	20	-84	-4
Central opercular cortex	167	3.92	0.0029	-38	-14	16
Lateral occipital cortex (superior)	165	4.10	0.00313	-30	-76	24
Insular cortex	143	4.09	0.00744	38	-2	12
Precentral gyrus	112	4.00	0.0271	62	12	12

Note. X, Y & Z: coordinates in MNI space, credits > people: clusters showing a stronger positive correlation with number bin (1-5) for credits than people. Results for credits, people and credits > people all combining positive and negative outcomes.
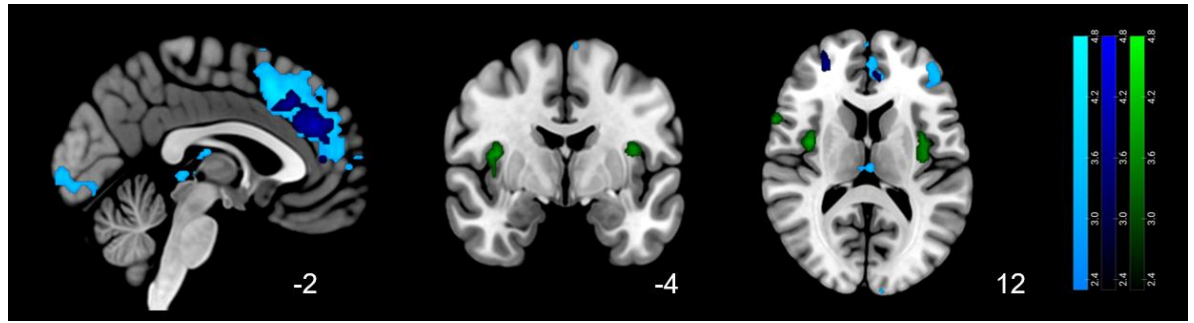
Figure 5. Regions showing a positive correlation with number bin (1-5). Light blue: credits (both positive & negative outcomes); dark blue: credits correlation with number > people correlation with number; green: negative outcomes abroad. Images in radiological orientation (right = left), coordinates in MNI space (mm), threshold Z > 3.

3.3. UK vs. abroad

3.3.1. ROIs

When analysis was limited to trials involving people, across the UK and abroad, negative outcomes were associated with greater activity than when the people survived in all three ROIs (table 5). Mixed-models analysis of parameters in the 3 main ROIs did not show a significant main effect of location (*UK* or *abroad*) in any of the regions. However, in both left and right AI positive and negative outcomes in the UK were differentiated more than those abroad (left $t = 2.25$, $p = .025$, right $t = 2.04$, $p = .042$).

In the ACC and right AI, a three-way interaction was shown between location (*UK* or *abroad*), outcome valence (*positive* or *negative*) and number. As shown in the graphs in figure 3, responses to outcomes in the UK were relatively flat across the number of people, regardless of valence. In contrast, for outcomes abroad, responses to positive outcomes generally declined as the number increased whereas responses to negative outcomes increased over number. This pattern resulted in a point at approximately 100 people abroad where outcomes started to be differentiated in a similar way to all numbers of people in the UK.

Table 5

Results from mixed-models analysis of the 3 ROIs testing the effect of location (UK or abroad), outcome valence (positive or negative), number (log_{10}), and interactions between these factors.

	ACC		Left AI		Right AI	
	t	p	t	p	t	p
Intercept	6.13	<.001*	-1.68	.106	-6.29	<.001*
Number (linear)	0.24	.807			1.65	.100
Number (quadratic)	1.56	.121				
UK – abroad	-0.34	.737	1.35	.178	0.41	.681
Negative – positive	2.18	.039*	2.53	.018*	2.64	.014*
Number (linear) x location	0.82	.410			-1.31	.191
Number (quadratic) x location	0.63	.532				
Number (linear) x valence	1.96	.050			1.49	.138
Number (quadratic) x valence	1.82	.069				
Location x valence	1.84	.066	2.25	.025*	2.04	.042*
N (linear) x L x V	-2.24	.025*			-2.10	.036*
N (quadratic) x L x V	-0.72	.470				

Note. x: interaction, N: number, L: location, V: valence, blank cells: the variable was not included in the best fitting model, *: p < 0.05

Table 6

Cluster peak details from the whole-brain contrast of UK vs. abroad

Label	Voxels	Z-score	p	X	Y	Z
UK > abroad						
Inferior frontal gyrus, pars opercularis	179	4.19	0.0019	-56	20	26
Abroad > UK						
Lateral occipital cortex (inferior)	2338	5.77	3.13e-23	-38	-86	-8
Occipital pole	986	5.40	1.14e-12	28	-98	-10
Cingulate gyrus (posterior)	538	4.14	5.96e-08	8	-48	6
Parahippocampal gyrus	149	4.02	0.006	32	-34	-16
Right amygdala	133	4.05	0.0114	16	-6	-14

Note. X, Y & Z: coordinates in MNI space.

3.3.2. Whole brain

Results from the ROIs suggest that the effect of location (*UK* or *abroad*) was more relevant for interactions with outcome valence and number but we also ran an exploratory whole-brain contrast between outcomes and the UK and abroad, averaging across valence and number. This revealed greater activity in the inferior frontal gyrus (IFG) for UK > abroad. For the contrast abroad > UK, activity was in the posterior cingulate gyrus, right hippocampus, right amygdala and occipital regions.

Responses to people dying in the UK were greater than when the people survived in bilateral insula in the whole-brain analysis, in addition to the ROIs and a more dorsal, posterior region of the ACC. Additional regions revealed in this contrast included the right amygdala, left TPJ and occipital areas (table 7, figure 7). No regions showed increased activity for positive, compared to negative, outcomes in the UK. The same contrast for outcomes abroad did not show any areas of significance at our whole-brain threshold in either direction. An interaction between the effect of valence in the UK vs. abroad was shown in the intracalcarine cortex (24, -64, 6). No regions showed the three-way interaction found in the right AI and ACC ROIs between valence, location and number.

Interestingly, the increase in activity associated with die, compared to live, outcomes in the UK was a result that particularly differed when accounting for the post-scan ratings of how surprising and emotional the stories and their outcomes were. Including surprise as a parametric modulator reduced some of the regions of significance whereas accounting for emotion left no significant regions at all (see table 7 and figure 7).

As the contrasts creating a linear relationship between activity and the number bin for people did not seem to capture how this variable is processed, we also created contrasts to detect activity that correlated with an *inverted 'V'* shape, increasing from bin 1 (1 person) to bin 2 (2-10 people) but then declining across bins 3, 4 and 5. This was based on the graphs from the ROIs and also results from skin conductance measures on a similar news stories task (paper 3). A relationship would imply a maximum responsivity to numbers near 2-10.

Correlations with this inverted V pattern was found for outcomes both in the UK and abroad in ACC, but in separate, neighbouring regions (figure 8 and table 8). Activations for events in the UK were closer to the corpus collosum in the anterior cingulate gyrus whereas activations for events abroad were more anterior and larger. This contrast for outcomes in the UK also showed activity in the temporoparietal junction (TPJ). The inverse contrast abroad, creating a 'V' rather than an inverted 'V', revealed a cluster in the left superior lateral occipital cortex (-26, -72, 44).

Repeating this analysis with the peak at bin 3 (increase between bins 1 to 3 then decline) showed a similar cluster for the UK in the planum temporale (58, -30, 22). No activations showed significant associations with the 'inverted V' abroad but the 'V' (decline to bin 3 then increase) showed clusters in the left pSTS (-52, -36, 2), right superior lateral occipital cortex (32, -68, 28) and left temporal pole (-52, 8, -16). Overall, these results show that different regions respond maximally to different numbers of people.
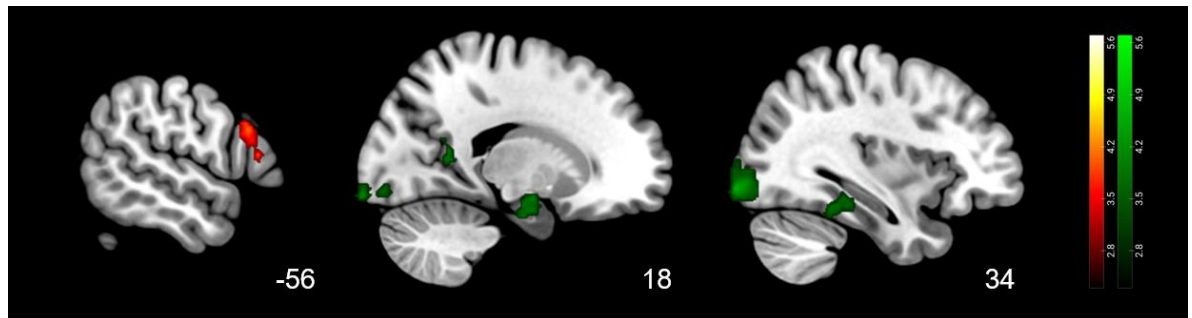
Figure 6. Contrast between the UK and abroad, including both positive and negative outcomes. Orange: UK > abroad; green: abroad > UK. Coordinates in MNI space (mm), threshold Z > 3.

Table 7

Cluster peak details from the whole-brain contrast between negative and positive outcomes in the UK

Label	Voxels	Z-score	*p*	X	Y	Z	
UK negative > UK positive							
Occipital fusiform gyrus	1052	4.12	3.17e-14	-18	-88	-20	†
Occipital fusiform gyrus	510	4.05	2.04e-08	18	-78	-18	
Superior frontal gyrus	407	4.08	4.17e-07	6	12	62	†
Inferior temporal gyrus (temporooccipital)	405	4.16	4.17e-07	-48	-64	-20	†+
Cingulate gyrus (anterior)	324	4.07	5.66e-06	2	20	20	†
Frontal operculum cortex	310	4.03	9e-06	-34	24	6	
Frontal operculum cortex	257	4.02	5.56e-05	48	16	-4	†
Brain-stem	197	3.86	0.00051	-4	-22	-16	†
Right amygdala	143	3.95	0.00459	18	-6	-14	
White matter / right thalamus	133	3.85	0.00709	10	-2	0	
Lingual gyrus	104	3.74	0.0266	2	-76	6	+
Angular gyrus	100	3.88	0.0321	-50	-52	24	

Note. X, Y & Z: coordinates in MNI space. †: cluster remains significant when accounting for surprise (none remained significant when accounting for emotion) - additional brain-stem cluster in this analysis at 4, -30, 0. +: significant interaction between frame and location – additional intracalcarine cortex cluster for this interaction at 24, -64, 6.
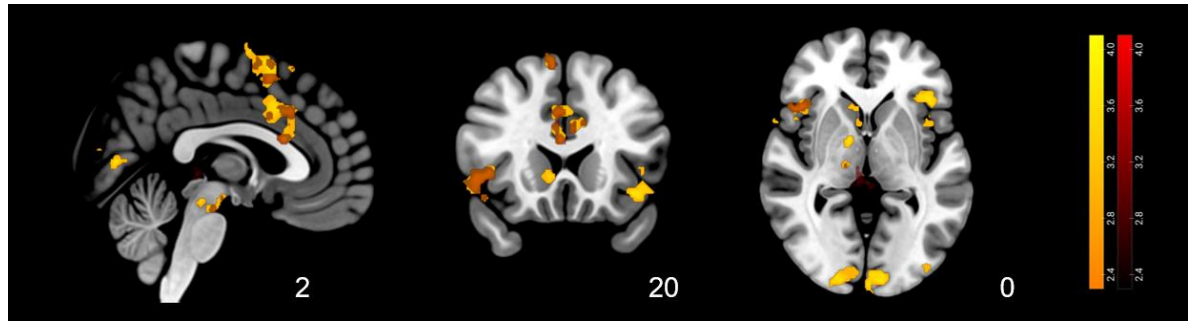
Figure 7. Contrast between the negative (die) and positive (live) outcomes in the UK. Orange/yellow: UK negative > UK positive with no parametric modulator; red: UK negative > UK positive after including surprise ratings as a parametric modulator. Images in radiological orientation (right = left), coordinates in MNI space (mm), threshold Z > 3.

Table 8

Cluster peak details from the whole-brain contrast modelling an 'inverted 'V' relationship between activity and the number of people

Label	Voxels	Z-score	p	X	Y	Z
UK inverted 'V'						
Planum temporale	541	4.09	5.96e-08	62	-30	20
Cingulate gyrus (anterior)	151	3.85	0.00655	2	26	20
Cuneal cortex	144	3.82	0.00859	4	-84	24
Lingual gyrus	105	4.21	0.0422	-12	-64	-8
Abroad inverted 'V'						
Paracingulate gyrus	783	4.04	2.38e-10	6	46	0
Frontal pole	160	3.98	0.00512	30	52	26
Paracingulate gyrus	120	4.03	0.0243	4	26	36

Note. X, Y & Z: coordinates in MNI space.
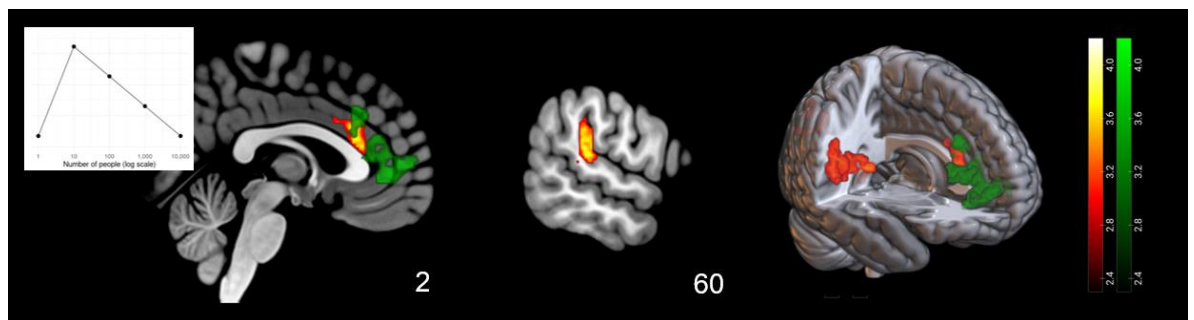
Figure 8. Regions where activity showed an inverted 'V' relationship (peaking at bin 2, see inset) with the number of people. Orange: outcomes in the UK; green: outcomes abroad. Images in radiological orientation (right = left), coordinates in MNI space (mm), threshold Z > 3.

4. Discussion

Using novel stimuli adapted from real news stories, we identify the neural correlates of responses to other people's lives being at risk and finding out whether they survived or died. A human life is processed differently to non-social outcomes. This is shown by a departure from the linear increase in neural activation with the amount of risk we see in a non-social context. Instead, we find that processing of risk to human life is less sensitive to the number at risk, and more sensitive to what happened to people. However, these attributes of valuing the lives of others are not the same for lives in the same country and lives abroad.

Behaviourally, there was not strong evidence that participants differentiated people in the same country from people abroad. Participants also increased the amount they would spend to save increasing numbers of people. However, our fMRI results show insensitivity to the number of people and that lives in the UK are processed differently to those abroad in three important ways. First, the effect of death, compared to survival, is greater in the UK than abroad. Second, the effect of number differs between people in the UK and abroad and finally, the relationship with number correlates with activity in separate regions.

Here we discuss differences between experimental conditions in the activation patterns shown in three ROIs (ACC, left AI and right AI) taken from a meta-analysis on empathy for pain (Lamm et al., 2011) and whole-brain analyses.

4.1. People vs. credits

4.1.1. Main effect

Scope insensitivity to numbers occurs for valuation of both economic and social outcomes (Kahneman and Tversky, 2013; Slovic, 2007). For non-social or financial outcomes, a common finding is that each additional unit adds value but not as much as the previous one. However, previous work suggests that valuing the lives of other people may be affected by additional, emotional processes that can result in decreases in value as the number of lives increases (Dickert et al., 2015). One of our contributions is further insight into how the valuation of people differs from non-social outcomes, by contrasting people's lives with credits, a financial resource for participants.

One key difference between responses to outcomes affecting people and those affecting credits was the location of activation when collapsing across different numbers and combining positive with negative outcomes. Greater activity for people > credits was

shown in areas associated with valuation including the vmPFC and OFC (Bartra et al., 2013). Activity in vmPFC as well as the precuneus, also revealed in this contrast, were linked to empathy for pain communicated through abstract cues, compared to images (Lamm et al., 2011), and are associated with inferring the mental states of others (Frith and Frith, 2006). Finally, several further regions more active during social, compared to non-social trials, are linked to episodic memory including the precuneus, hippocampus, and an area of the right cerebellum (Jeong et al., 2015). Given that the people trials were taken from real news stories in the past, it is possible that participants were remembering or trying to remember the event.

In contrast, trials involving credits showed increased activity in regions associated with working memory and attention in the frontal cortex (Majerus et al., 2016). The specific region of the superior frontal gyrus has also previously been linked with arithmetic and calculation (Fehr et al., 2007). The right AI ROI also showed increased activity for credits compared to people, but in the whole-brain analysis, the left insula showed the reverse.

4.1.2. Number

In addition to contrasting all trials involving people and all trials involving credits, we compared the effect of number between these trial types. Multiple regions had activity that positively correlated with the number of credits when put on a scale of 1-5 (similar to taking the $\log_{10}$ of the average number in that bin + 1). The ACC showed this pattern in both the ROI and whole-brain analysis and linearity was greater for credits than people here and into the dmPFC. While the ROIs were selected from research on empathy for pain, the ACC has also been linked to experiencing pain, losses and errors (Vassena et al., 2017). Positive correlations between activity and the number of credits were also shown in OFC, an area linked to value (Bartra et al., 2013).

Combining all trials about people gives the greatest power to detect a positive correlation with number but the only regions found in this contrast were in the occipital cortex. While there is some evidence that emotion visual stimuli affect activity in visual areas (Vassena et al., 2017), it is impossible to exclude the possibility that this was simply due to larger numbers having more digits so being longer on the screen. Interestingly, the same areas of ACC and dmPFC that showed a positive correlation with the number of credits showed activity during outcomes affecting people that varied in an inverted V shape with the number of people.

4.1.3. Outcome

The final distinction between people and credit trials was a greater differentiation between negative and positive outcomes when they affected people, compared to credits, in the ACC ROI. Analysis separating people trials into those in the UK and those abroad showed this was driven by the UK, so this result is discussed below.

4.2. UK vs. abroad

4.2.1. Main effect

An overall difference between outcomes affecting people in the UK and people abroad was not found in any of our ROIs. Whole-brain analysis of this contrast showed increased activity in the IFG for UK > abroad. This region has been associated with comprehending language and reading (Andin et al., 2015; Citron et al., 2014). Regions including the right hippocampus and right amygdala were shown in the inverse contrast (abroad > UK). The amygdala has typically been associated with negative emotions but is also linked with reward and a range of social processes (Bickart et al., 2014). The increase in activity here during outcomes abroad, compared to the UK, may have been driven by the fact that this region differentiated positive from negative outcomes in the UK, with a deactivation for positive outcomes.

4.2.2. Number

In the four combinations of location (*UK* or *abroad*) and outcome valence (*positive* or *negative*), only people dying abroad elicited activity that positively correlated with number (scale 1-5), outside occipital regions. As this was in bilateral insula, ROIs that we graphed parameter estimates for, we can see that the increase in activity with number is only after bin 2 (2-10 people) in the left and after bin 3 (11-50) people in the right.

Based on the ROI graphs and the relationship between skin conductance amplitudes and numbers of people in paper 3, we ran a whole-brain analysis to detect regions where activity showed an 'inverted V' shape – rising from bin 1 to bin 2 then declining. For outcomes in the UK, activity in right TPJ correlated with this shape. This region is linked to social processing, particularly theory of mind (Decety and Lamm, 2007; Frith and Frith, 2003; Schurz et al., 2014).

As predicted from the ROIs, the ACC also showed an inverted V pattern but in separate, neighbouring regions for the UK and abroad. The region showing an inverted V relationship with number of people in the UK was in the ACC gyrus (ACCg). The region for abroad was more anterior and larger, covering ACC sulcus (ACCs) and ACCg. For abroad,

this overlapped with the region showing a positive correlation between activity and number bin for credits.

One interpretation of these differing activations is based on the idea that the ACCg is more specialised for social processing, whereas the ACCs is more domain general and responds to stimuli involving the self, other and non-biological agents (Apps et al., 2016). Overlap in the less specialised ACCs for the relationship with number, albeit different shapes, is one way that responses to events abroad were more like responses to credits than responses to events in the UK were.

4.2.3. Outcome

A final distinction between responses to outcomes affecting people in the UK vs. abroad, and another way in which events abroad were more similar to credits, was that positive and negative outcomes were only distinguished when they affected people in the UK. The contrast UK negative > UK positive showed a network of regions linked to social and emotional processing including bilateral insula, amygdala and left TPJ (see above). If a measure of the value of a life is the difference between responses to people living or dying, these results suggest this was greater for lives in the same country as the participant.

4.3. Limitations

As outlined above, combining our novel news stories task with fMRI gives insight into the neural correlates of valuing the lives of others and biases that affect this value. However, there are some limitations that must be recognised.

Like most fMRI studies, our sample size was small, limiting the power to generalise results. However, given the sample was politically liberal, the levels of scope insensitivity and bias towards outcomes in one's own country are more likely to be an underestimation than an overestimation of these biases in the general population.

Another issue is whether participants' expectations about the outcomes affected responses. To account for this, after the scan, we asked participants to rate how surprising and how emotional they found the outcomes. As shown in table 7 and figure 7 for the contrast between negative and positive outcomes in the UK, accounting for surprise did not fully explain the results. Instead, in line with the interpretation that these differences were due to emotional processes, accounting for emotion did mean no activations met the threshold for significance. No regions showed activity that correlated with surprise or emotion ratings across all trials.

5. Conclusion

Responses in the brain when lives are lost are distinct from responses to the loss of financial resources. While the value of credits with financial value increases linearly with the amount, neural correlates of the value of people show non-linear relationships with number. Within people's lives, those in the UK are valued differently to lives abroad. Whether people abroad died or survived did not affect responses overall. In contrast, a network of emotion and social processing regions differentiated deaths in the UK from situations in which the people survived.

Responsibility and valence determine interest in outcomes for others

Jo Cutler and Daniel Campbell-Meiklejohn

The Social Decision Lab, the University of Sussex, School of Psychology, Falmer, UK.

1. Introduction

Altruism is crucial for positive interactions between people and cohesive societies. Altruistic behaviours range from the low cost, such as holding open a door for someone, to the extraordinary, like donating a kidney to a stranger (Brethel-Haurwitz et al., 2018; Marsh et al., 2014). In addition to benefits for recipients, altruistic behaviours have been linked to increased wellbeing (Aknin et al., 2013a; Dunn et al., 2014, 2008). Anonymous donations, to other people or charities, exclude strategic motivations to be prosocial to induce reciprocity. Even in these contexts, altruistic decisions correlated with activation in reward-related brain regions (Cutler and Campbell-Meiklejohn, 2019; Moll et al., 2006).

Previous work has identified two broad motivations for altruistic behaviours: pure and impure (Andreoni, 1990, 1989), also termed outcome-oriented and action-oriented (Kuss et al., 2013). We use the terms outcome and action orientation here. Outcome-oriented altruism is motivated by the positive outcome or public good, independent of one's own contribution. In contrast, action-oriented altruism is motivated by positive emotions from the act of giving. This has been described as warm glow giving in the past (Andreoni, 1990, 1989), but recent uses of the term warm glow do not necessarily exclude outcome-orientations.

These motivations are not necessarily in competition and likely work together, even within a single decision. However, there are important implications of which is most prominent for understanding and encouraging prosocial behaviour. For example, charities often tell donors what has been achieved with their donation. An outcome-oriented donor would be most strongly reinforced by the change in wellbeing for the recipients. Whether this is from their specific donation, or donations from other donors, should not make a difference. In contrast, an action-oriented donor would prefer to hear about the impact of their personal donation.

Understanding motivations for giving also offers insight into why prosocial decisions show biases, deviations from rationality and the moral rule that all lives are equal (Dickert et al., 2012). Some examples of these biases are increased giving to singular, identified beneficiaries (Genevsky et al., 2013; Västfjäll et al., 2014) and to beneficiaries who are more similar to the donor (Loewenstein and Small, 2007). Often these factors work in opposition to having the greatest impact. Impact is more likely to be obtained through giving to causes that are evidence-based (statistical) and help people in foreign countries. Perhaps these biases could be explained as giving in the way that makes the donor feel good (action-

oriented altruism), rather than the way that has the biggest impact for the recipients (outcome-oriented altruism).

To study whether the distinction between outcome and action orientations is useful for understanding altruism, it is necessary to separate these motivations experimentally. Most work on this topic has come from studies of crowding out: whether participants reduce their donations if someone else provides the public good (Ottoni-Wilhelm et al., 2017). For example, paradigms may give participants the opportunity to contribute to an appeal from their own earnings, but the amount they gave would be deducted off a donation from the lab. As the total donation remained the same, the outcome for the beneficiaries is kept constant and the only possible motivation to give is warm glow from the action of giving. While these paradigms provided evidence for action-oriented motivations, through incomplete crowding out, they do not isolate, so cannot measure, individual differences in outcome-oriented motivations.

Several studies have developed measures of both outcome and action orientations by using functional magnetic resonance imaging (fMRI). In the first, participants made voluntary charitable donations and also observed tax-like involuntary transfers from their money to the charity. The ventral striatum was active in both contexts and more active during voluntary donations, supporting the existence of both outcome-oriented and action-oriented giving (Harbaugh et al., 2007). Between-subjects, whether activation was greater during transfers to the charity or money for the participant predicted real-world giving. This measure of 'pure altruism' was also linked to a general benevolence trait that increases with age (Hubbard et al., 2016) and suggested to strengthen following gratitude practice (Karns et al., 2017).

In a different paradigm, participants also made costly charitable donation decisions in the scanner but on some trials, donations were discarded. As the action-oriented reward of being generous is unaffected by whether a donation is discarded, a reward prediction error signal in the nucleus accumbens is evidence for other-oriented motivations. This was present, but only for the most generous participants, again supporting a relationship between other orientation and giving (Kuss et al., 2013).

These fMRI studies demonstrate the insight that can be gained in this area but rely on expensive methods. To further study how action and outcome orientations relate to other traits and behaviours, a behavioural measure of these motivations is required. To do this, we focused on a key difference between them: interest in outcomes for other people.

Interest in outcomes for us is a normal psychological feature for most people. This is expressed through an optimism bias: updating beliefs more in response to positive than negative information (Sharot, 2011). Participants also preferred, and even paid, to find out information about positive outcomes (monetary wins) and avoid information about negative outcomes (losses) for themselves, even though the outcome was fixed. Patterns of brain activity further supported the idea that positive information has intrinsic value (Charpentier et al., 2018).

There is also evidence for a vicarious optimism bias and this was linked to concern for other people (Kappes et al., 2018). Greater vicarious optimism bias was found for likeable and identifiable strangers, factors that also predict concern, and was linked to altruistic behaviour. These results support the basis for the current study: that concern for others translate into a bias towards information about positive, over negative, outcomes.

The method used by Kappes et al. (2018) measured updating of beliefs about positive or negative life events. The fact that none of these events were caused by the participant aligns them with an outcome-oriented concern for others. In the present study, we adapt the tasks developed by Charpentier et al. (2018) to create conditions in which positive and negative financial outcomes affect either the participant or a third person. In our 'visual perception task', these outcomes affect either the participant or someone else and are caused either by the participant or a third person. This allows us to compare interest in positive and negative outcomes for the self and others in contexts where altruistic concern could be driven by action-oriented or outcome-oriented interest.

We present four versions of this paradigm and a fifth experiment using charitable donation decisions made by either the participant or a third person. Results show interest in outcomes, and the level of bias towards positive outcomes, depend on who is affected and who is responsible. We also adapt the 'probability task' from Charpentier et al. (2018), in which positive or negative outcomes occur with different probabilities and apply this to outcomes for someone else as well as the self. In this task, no one is responsible for the outcomes. Patterns of interest across the conditions are similar for self and other suggesting a concern for the wellbeing of others, independent of being responsible for it.

2. Experiments 1-4

Experiments 1-4 consisted of two parts: an online questionnaire (20 minutes, see below) completed before a single lab session (40 minutes). Experiment 5 was run entirely online and used different tasks so is described separately below. All experiments were approved by the Sussex University Sciences & Technology Cross-Schools Research Ethics Committee.

The lab sessions across experiments 1-4 involved two tasks, each measuring how often participants found out about outcomes in different conditions. The '*probability task*' did not manipulate who was responsible but tested the effect of probability on the frequency of finding out about outcomes affecting self and other. Details of the probability task did not change between experiments, so analysis and results are across all participants to increase power.

The '*visual perception task*' manipulated responsibility for the outcomes through performance on a simple task (see below). The trials also differed in who the outcome affected and positive or negative frame. How these frames were created, and other details of the task, differ between experiments 1-4.

2.1. General method

2.1.1. Participants

Across experiments 1-4, a total of 226 participants were tested. For the probability task, data from 205 participants is included in the analysis, as data was lost due to technical issues or excluded for 21 people. The details for the participants included in each analysis of the visual perception task are given below for each experiment.

2.1.2. Procedure

Participants for experiments 1-4 were recruited on campus or through social networks. Some took part for course credit and all participants were reimbursed for their time with the amount depending on the tasks. Adverts for the study said a payment of up to £7.50 but the maximum amount was higher for some participants, up to approximately £12.

Individual difference measures, except the charitable giving preference questions (see below) were completed before participants attended the lab, through a Qualtrics Online Survey Software (Qualtrics, Provo, UT) questionnaire.

Participants then attended the lab session in pairs and were randomly assigned to be Player 1 and Player 2. They arrived at the same time, were given instructions together, and asked to introduce themselves but then had no further contact and completed the tasks in separate rooms.

Verbal instructions were provided for the two tasks before the participants completed the visual perception task first, then the probability task and the charitable giving preferences questions (see below). Both tasks and these questions were presented using MATLAB (2017b, Mathworks, Inc., Natick, MA) using Psychophysics Toolbox extensions version 3 (Brainard, 1997).

2.1.3. Probability task

We adapted the task created by Charpentier et al. (2018), which only concerned outcomes for the participant. In the current versions, outcomes affected either the participant's payment or the payment of another player, who the participant did not meet but was described as in need. Each trial presented a **probability** (*10, 20, 30, 40, 50, 60, 70, 80* or *90%*) of the given outcome occurring. These trials were either in a *positive* (win / no change) or *negative* (no change / loss) **frame**, with the percentage referring to how likely a change in payment (win or loss) was. The conditions of **probability**, **frame** (*positive* or *negative)* and who was **affected** (*for me* or *for someone else*) created a 9x2x2 design. Participants could not influence the outcome in any way.

There was one block for each of the four frame-recipient combination blocks, with each probability presented four times. This resulted in 36 trials per block and 144 trials in total. On every other trial, the participant could press to find out whether the outcome had occurred so whether the payment had changed (increased or decreased, depending on the frame).

Trials were quasi-randomised so each probability-frame-recipient combination occurred twice on trials were the outcome could be found out and twice when it could not be. This resulted in a measure of finding out between 0 and 2 for each condition. Trials with no opportunity to find out took 2s and trials with an opportunity to find out took 6s, giving an average trial length of 4s plus a 1s fixation between each trial.

2.1.4. Visual perception task

To generate responsibility for outcomes in experiments 1-4, we used correct and incorrect answers to a visual perception task adapted from one originally designed to measure metacognition (Fleming and Lau, 2014). In this task, participants indicate which of

two circles on the screen contains more dots. The stimuli were displayed for 0.7s on each trial then participants had 3s to answer. A fixation cross was displayed for 1s between trials. A one-up two-down method staircased task difficulty through adjustments to the difference in the number of dots. This maintained average performance at approximately 70%.

The visual perception task involved four players, the two participants who completed this task during the lab session and two other players. Player 3 was described as in financial need and their payment depended on the performance of Players 1 & 2, the participants. Player 4 had already completed the task, playing to affect the payment of the two participants (see figure 1). Players 3 & 4 were a researcher. In the case of Player 4, they genuinely completed the task in advance, with the responses, including reaction times, recorded and used in the experiment. All payments and donations were made so there was no deception, and this was stressed to participants during the instructions.
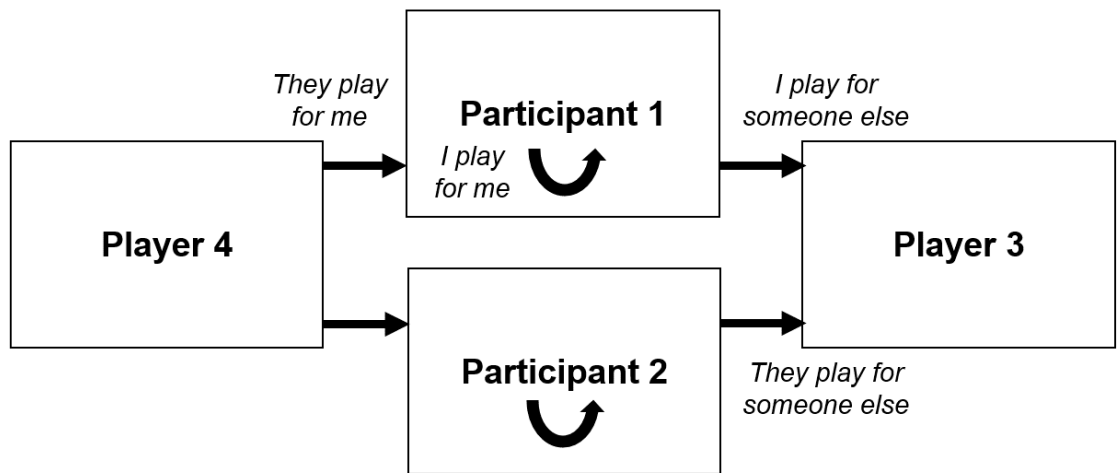
Figure 1. Relationship between players. Arrows indicate an effect on that person's payment, curved arrows indicate playing for oneself. Labels of how the arrows relate to the conditions are given for participant 1, for participant 2 these would be mirrored.

The combination of these players created 2 of the 3 main conditions, fully-crossing who was **playing** (*I play* or *they play*) with whose payment was **affected** by the outcome (*for me* or *for someone else*). Finally, the **frame** of trials was either *positive* or *negative*, creating a 2x2x2 design (**playing** x **affected** x **frame**). The nature of the positive and negative frame changed between experiment 2 and experiment 3, so is described for each below.

Participants then had the opportunity to find out the outcome of trials in these 8 conditions, although the detail of this differed between experiments (see below). Opportunities to find out were always on every other round to prevent tracking

of each players' total, as this could promote a more competitive approach. Finding out always required an active button press and not pressing resulted in not finding out.

Whether participants found out or not did not affect the payments at all or the length of the task. If they found out, the outcome was always shown and if they didn't find out, a fixation cross was displayed for the same amount of time. It was stressed to participants that whether they found out or not would not affect any of the payments or how quickly they finished the study.

Unfortunately, the nature of the connection between the computers being used by the two participants was unreliable, leading to the loss of data from several participants in each experiment.

2.1.5. Questionnaire measures

Interpersonal Reactivity Index (IRI) – participants completed the 28-item version of the IRI measure. This includes 7 items of each subscale; Perspective Taking, Fantasy, Empathic Concern and Personal Distress (Davis, 1983). For individual differences analysis, we averaged across the subscales. This measure allowed us to test whether differences in interest between conditions were linked to other-focused emotions.

Charitable giving – following the end of the lab tasks, participants completed a final set of novel questions on donating to charities. For each of 7 causes, participants indicated the percentage of their winnings from the lab part they wished to donate to the charity. The charities were not named organisations but differed in the information given about each. At the time of making these decisions, participants did not know how much they had won but were told to think of it as £2.50 on average as it could be up to £5. This was to ensure that the amount they had won did not affect these choices.

Participants were told all charities save lives through stopping preventable diseases. The first charity formed a baseline measure with no further information about the organisation. The next 6 choices were presented in a randomised order for each participant. One organisation was 5% towards a £1,000 fundraising target, another was at 95% towards an identical target. A separate organisation had a matched funding appeal so participant donations would be doubled. A further pair differed in the proportion of donations spent on projects (100% compared to 85% projects, 7% admin and 8% fundraising). The final organisation was "already receiving a donation from the lab. Anything you contribute will be given to the charity, but the lab will give that much less so the overall donation will always be the same amount". For the analysis here, we just averaged the donations made across all organisations.

2.1.6. Analysis

We used generalised linear mixed models (GLMM) with lme4 (Bates et al., 2015) in R (R Core Team, 2017) with RStudio (RStudio Team, 2015) for analysis of all tasks. A binomial distribution was modelled to account for the binary outcome of whether participants found out the outcome or not on each trial. Models were fitted by maximum likelihood estimation.

Analysis started with the maximal random-effects model (Barr et al., 2013) but this was always too complex so we simplified the models, including by using the Analysis of Factorial EXperiments (afex) package (Singmann et al., 2019) to remove correlations between random effects of binary factors. However, the pattern of results was the same as in the non-converging maximal model, suggesting the lack of convergence may have been a false positive and that fixed effects remain significant when accounting for all possible random effects.

Converging models were compared based on the Akaike information criterion (AIC; Akaike, 1974) to identify the model that best fit the data. We report results from this model in each analysis. For the probability task, which did not differ across experiments 1-4, the data were combined and analysed as a single set. For the visual perception task, this process was run on data from each experiment (1-4) separately for the main effects. Parameters for the fixed effects from all models were converted to odds ratios and their confidence intervals calculated.

Individual differences analysis was run on data from experiments 1-4 together. This was to increase power to detect such effects and find results consistent across the versions. This model included trial number plus interactions between frame (*positive* or *negative*), who was playing (*I play* or *they play*), who was affected (*for me* or *for someone else*) and the trait of interest (average donation / IRI average).

In addition to the variables of interest that form the conditions, trial number is included as a predictor in the models. For a blocked design, there is a risk that this captures some of the variance due to condition differences, although this is minimised by the randomisation of blocks for each participant. We chose this approach as it is the more conservative, compared to only modelling block order, which has a much lower resolution, or trial within each block, which does not account for changes in interest over the full task.

For data visualisation and average interest in each condition, we calculated the proportion of times participants found out in each condition and took the mean and within-subject standard error of this measure. These values are presented as percentages. For the

visual perception task, this was out of 9 opportunities to find out in each condition and for the probability task, 2 opportunities in each condition. For reporting the models, Z-scores and *p* values were calculated for each parameter by afex, which uses lmerTest from lme4.

2.2. Probability task results

The best model for data on the probability task across experiments 1-4 included main effects for all the variables of interest: frame, probability, and who was affected, as well as trial number. It also included interactions between a) frame with who was affected and b) probability with frame – the key interaction of interest. The model had a fixed and a random term for each of these factors.

Results showed all the fixed terms were significant predictors of whether participants found out, except for the main effect of probability and the interaction between who was affected and frame (table 1). However, removing this interaction significantly decreased the model fit (ΔAIC for removing fixed and random term = 58, $\chi^2(1)$ = 59.52, *p* < .001).
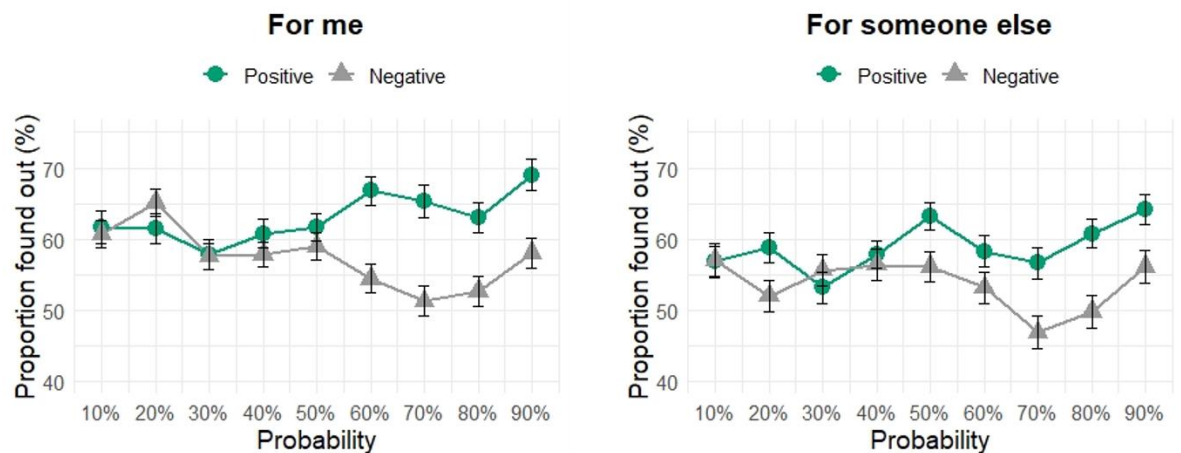

Figure 2. Effect of probability on the proportion of positive and negative outcomes participants found out about when outcomes affected the participant (left) or someone else (right).

Table 1

Model parameters for the probability task, experiments 1-4

	Odds ratio	Lower CI	Upper CI	|Z|	*p*
Intercept	2.03	1.46	2.82	4.21	<.0001***
Trial number	0.25	0.20	0.32	10.89	<.0001***
Positive – negative	0.82	0.75	0.90	4.15	<.0001***
For me – for someone else	0.84	0.75	0.94	3.15	.002**
Probability	1.02	0.83	1.27	0.22	.826
Affected * frame	1.04	0.96	1.13	1.01	.310
Probability * frame	0.52	0.37	0.74	3.66	.0003***

Note. CI: confidence interval, |Z|: absolute value of Z-score.
*: p < .05, **: p < .01, ***: p < .001

Participants found out the outcome more when it could be positive (average 60.95% of trials), than when it could be negative (55.53%, mean difference – M.D. 5.42%, |Z| = 4.15, *p* < .0001). Interest was also higher when the outcome affected the participant (*for me* 60.23%), than when they were *for someone else* (56.25%, M.D. = 3.98%, |Z| = 3.15, *p* = .002). Crucially, the effect of probability depended on frame, with an increase in interest as probability increased for positive outcomes but a decrease for negative outcomes (figure 2 & table 2).

Table 2

Descriptive statistics for the probability task, experiments 1-4

	10%	20%	30%	40%	50%	60%	70%	80%	90%	all
Mean %										
For me										
+	61.71	61.46	57.80	60.73	61.71	66.83	65.37	62.93	69.02	63.06
-	60.73	65.12	57.56	57.80	59.02	54.39	51.22	52.68	58.05	57.40
+ > -	*0.98*	*-3.66*	*0.24*	*2.93*	*2.68*	*12.44*	*14.15*	*10.24*	*10.98*	*5.66*
For someone else										
+	56.83	58.78	53.17	57.80	63.17	58.29	56.59	60.73	64.15	58.83
-	57.07	51.95	55.61	56.34	56.10	53.17	46.83	49.76	56.10	53.66
+ > -	*-0.24*	*6.83*	*-2.44*	*1.46*	*7.07*	*5.12*	*9.76*	*10.98*	*8.05*	*5.18*
Within-subject standard error										
For me										
+	2.29	2.20	2.12	2.04	1.96	2.01	2.31	2.11	2.14	
-	1.95	1.96	1.82	1.79	1.94	2.00	2.13	2.06	2.17	
For someone else										
+	2.18	2.16	2.24	1.87	1.96	2.17	2.16	2.00	2.10	
-	2.31	2.18	2.13	2.17	2.09	2.21	2.30	2.22	2.28	

Note. +: positive frame, -: negative frame, + > -: positive frame subtract negative (italicised), all: average across percentages

At the lowest probabilities, levels of interest were not different between positive and negative outcomes (average of 0.28% difference between frames across probabilities 10, 20 and 30%) but at the highest probabilities, participants found out more about positive outcomes (average difference 10.69% at probabilities 70, 80 and 90%). Interest declined over the course of the task ($|Z| = 10.89$, $p < .0001$)

3. Experiment 1

3.1. Method

3.1.1.   Participants

59 people were tested but the data from 8 were lost, leaving 51 participants (39 female, 10 male, 2 undisclosed) aged 18-42 (M = 25.55, S.D. = 6.03).

3.1.2.   Visual perception task

In experiment 1, the positive and negative frame was created through the trials either being to win money or to avoid losing money for the recipient. If the player got the answer right on that trial, the positive outcome (win / no loss) occurred but if it was wrong, the negative outcome (no win / loss) occurred. Thanks to the staircasing, positive outcomes occurred with approximately 70% probability. Each change in payment was an increase or decrease of £0.10, but participants were not aware of the amount.

Participants completed a block of 18 trials for each of these 8 conditions. In *I play* conditions, the participant responded to a round of the visual perception question. When one of the other players was playing, they were told to wait while that player completed a visual perception round. After every other trial, regardless of who had just played, participants were asked whether they wanted to know whether the round was answered correctly or incorrectly, so what the outcome for the payment was. If they wanted to know, they pressed a key within 3s. Trials with an opportunity to find out took 8s and trials without an opportunity to find out took 4s giving an average trial length of 6s, plus a 1s fixation cross between each.

3.2. Results

The best converging model for experiment 1 (table 3) included main effects of trial number, who was playing, and who the round affected, as well as two interactions: a) who was playing with who was affected and b) who was affected with trial number. Each of these

factors had a fixed and a random term. Adding whether the outcome was positive or negative did not improve the model and was a non-significant predictor when included.

Results showed significant main effects of trial number ($|Z| = 2.08$, *p* = .038) and who was playing (*I play* or *they play*) (M.D. = 11.11%, $|Z| = 2.76$, *p* = .006). Participants found out more on trials for which they were responsible for the outcome (on average 61.93% of trials) than when someone else was responsible (50.82%). Overall interest declined over the task. The main effect of who the outcome affected was not significant (see table 4 *p* = .065) but this factor interacted with who was playing ($|Z| = 2.57$, *p* = .010). As shown in figure 3 and table 4, the decreased interest when *they play,* compared to when *I play*, was greater when outcomes were *for someone else* (decrease of 16.34%), than when they were *for me* (5.88% decrease, M.D. = 10.46%)*.* The fixed term for the interaction between trial and who was affected was not a significant predictor (*p* = .328) but improved the model through variance explained by the random term (ΔAIC for removing fixed and random term = 9.1, $\chi^2(2) = 13.07$, *p* = .001).
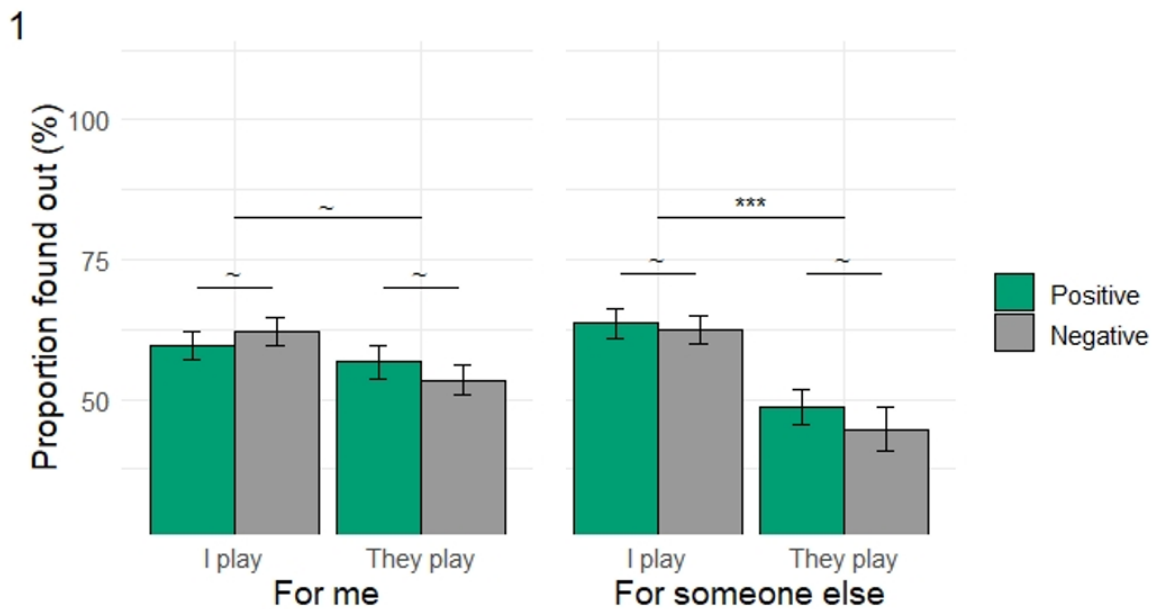

Figure 3. Average proportion of times (%) participants found out the outcome in each condition in experiment 1. Lower lines indicate the level of significance for the effect of frame within that condition, upper lines indicate the level of significance for the effect of who is responsible. ~: p > .05, *: p < .05, **: p < .01, ***: p < .001

Table 3
Model parameters for experiment 1

	Odds ratio	Lower CI	Upper CI	|Z|	*p*
Intercept	1.85	0.87	3.97	1.59	.111
Trial number	0.57	0.34	0.97	2.08	.038*
I play – they play	0.68	0.52	0.89	2.76	.006**
For me – for someone else	0.83	0.68	1.01	1.85	.065
Playing * affected	0.84	0.74	0.96	2.57	.010*
Trial number * affected	1.27	0.79	2.04	0.98	.328

Note. CI: confidence interval, |Z|: absolute value of Z-score.
*: p < .05, **: p < .01, ***: p < .001

Table 4
Descriptive statistics for experiment 1

	For me			For someone else			For me & for someone else		
Plays:	I	They	I & T	I	They	I & T	I	They	I & T
Mean %									
+	59.69	56.64	58.17	63.62	48.58	56.10	61.66	52.61	57.14
-	62.09	53.38	57.73	62.31	44.66	53.49	62.20	49.02	55.61
+ > -	*-2.40~*	*3.27~*	*0.44~*	*1.31~*	*3.92~*	*2.61~*	*-0.54~*	*3.59~*	*1.53~*
+ & -	**60.89**	**55.01**	**57.95**	**62.96**	**46.62**	**54.79**	**61.93**	**50.82**	**56.37**
Within-subject standard error									
+	2.46	2.87		2.73	3.27				
-	2.59	2.62		2.53	3.96				

Note. +: positive frame, -: negative frame, + > -: positive frame subtract negative (italicised), + & -: average of positive and negative frame (bold font), I & T: average of when I play and when they play, ~: no significant / *: significant difference between positive and negative frames, when analysis restricted to that condition.

4. Experiment 2

Experiment 1 found greater interest in outcomes the participant was responsible for. This difference was greater when the outcome affected someone else whereas responsibility had a smaller effect on interest when outcomes affected the participant themselves. However, there was no evidence that participants differentiated between the positive and the negative frames.

While combining the interest in finding out, across frames, showed effects of who the outcome affected and an interaction between who was affected and who was responsible, the lack of positivity bias causes an issue for the interpretation of these results. The experiment was designed such that a key contrast would the difference between wanting to know in the positive and negative frames, for each combination of player and

recipient (for example *I play for me*). Comparing these difference scores between conditions, or the interactions between frame and other factors, would control for extraneous differences between conditions.

However, when looking at main effects of who was playing and who was affected, the results are more susceptible to influence by factors other than concern about the outcome. In particular, when participants are deciding whether they want to find out on trials in the *I play* conditions, they have just completed the visual perception task, so their attention is on the screen. In contrast, in the *they play* conditions, the participant has just waited several seconds so may have looked away and missed the opportunity to find out. Finding out about one's performance on the visual perception task during the task may also have utility for improving performance on the following rounds. Experiment 2 changed the design to remove these potential confounds.

4.1. Method

4.1.1. Participants

51 new participants were tested but the data were lost for 13 due to technical difficulties leaving 38 participants in the analysis (32 female, 6 male) aged 18-29 (M = 20.50, S.D. = 2.84).

4.1.2. Visual perception task

The only difference in this task compared to experiment 1 was the separation of completing the visual perception task from the opportunity to find out the outcome. The visual perception task, staircasing and timings were exactly the same as experiment 1 but in experiment 2, participants first completed all the rounds of this task for the 4 conditions on which they were playing (*for me* and *for someone else*, each in *positive* and *negative* frames). Each of the 4 conditions had 18 trials, making a total of 72 visual perception trials. They were informed the other participant was doing the same simultaneously.

Each participant's performance, correct or incorrect, for each trial were then shuffled in order, maintaining the number of each. Both participants then were presented with 144 trials forming all 8 conditions, as in experiment 1. These conditions combined the recipient (*me* or *someone else*), frame (*positive* or *negative*) and at this point also the player manipulation (*I play* or *they play*). On every other trial, participants could press a key to find out whether the round was answered correctly or incorrectly so what the outcome for the payment was. Like in experiment 1, these were presented in 8 blocks of 18 trials, 9 with an opportunity to find out.

4.2. Results

Comparing converging models led to selection of a model including fixed main effects for all three key variables: who was responsible (*I play* or *they play*); who was affected (*for me* or for *someone else*); and frame (*positive* or *negative*), plus two fixed interactions: a) who was responsible with who was affected and b) who was affected with frame. The final predictor in the model was trial number (table 5). Random terms were included for all of these factors with the exception of the interaction between who was affected and frame, as this term explained so little variance, it led to a singular fit.

Table 5
Model parameters for experiment 2

	Odds ratio	Lower CI	Upper CI	|Z|	*p*
Intercept	8.94	4.92	16.22	7.20	<.0001***
Trial number	0.56	0.30	1.06	1.79	.073
Positive – negative	0.78	0.59	1.04	1.72	.086
I play – they play	0.67	0.53	0.86	3.25	.001**
For me – for someone else	0.75	0.59	0.96	2.28	.024*
Playing * affected	0.76	0.59	0.98	2.10	.036*
Affected * frame	1.38	1.20	1.59	4.57	<.0001***

Note. CI: confidence interval, |Z|: absolute value of Z-score.
*: p < .05, **: p < .01, ***: p < .001
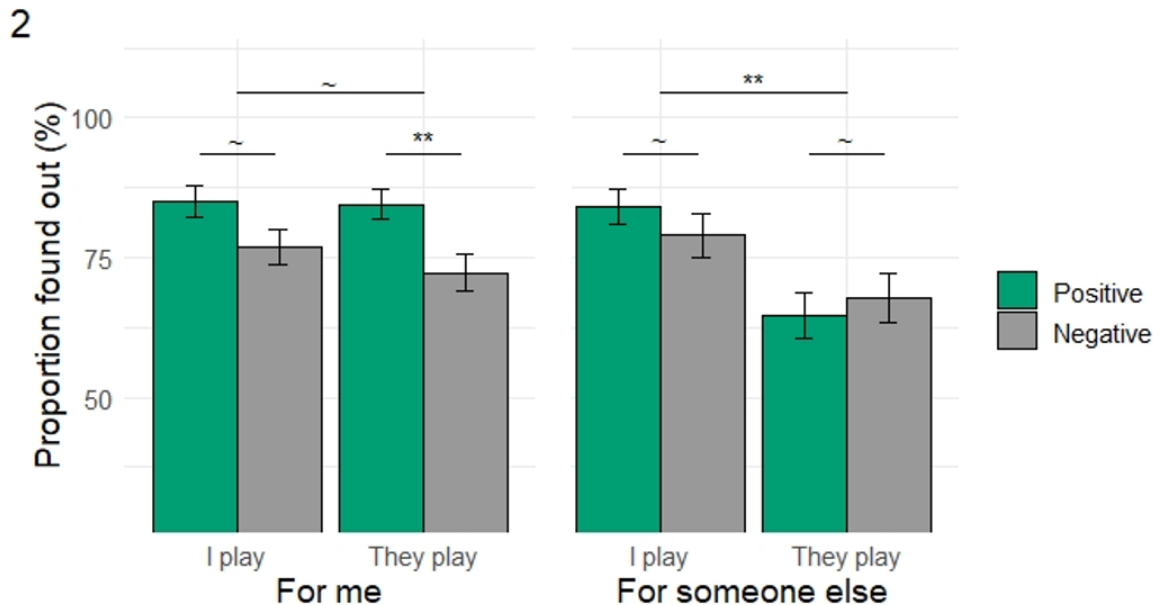
Figure 4. Average proportion of times (%) participants found out the outcome in each condition in experiment 2. Lower lines indicate the level of significance for the effect of frame within that condition, upper lines indicate the level of significance for the effect of who is responsible. ~: p > .05, *: p < .05, **: p < .01, ***: p < .001

Table 6
Descriptive statistics for experiment 2

Plays:	For me			For someone else			For me & for someone else		
	I	They	I & T	I	They	I & T	I	They	I & T
Mean %									
+	85.09	84.50	84.80	84.21	64.62	74.42	84.65	74.56	79.61
-	76.90	72.22	74.56	78.95	67.84	73.39	77.92	70.03	73.98
+ > -	*8.19~*	*12.28**	*10.23**	*5.26~*	*-3.22~*	*1.02~*	*6.73~*	*4.53~*	*5.63~*
+ & -	**80.99**	**78.36**	**79.68**	**81.58**	**66.23**	**73.90**	**81.29**	**72.30**	**76.79**
Within-subject standard error									
+	2.73	2.64		3.18	4.12				
-	3.01	3.33		3.88	4.33				

Note. +: positive frame, -: negative frame, + > -: positive frame subtract negative (italicised), + & -: average of positive and negative frame (bold font), I & T: average of when I play and when they play, ~: no significant / *: significant difference between positive and negative frames, when analysis restricted to that condition.

Results revealed all the fixed effects were significant predictors, except for the main effect of frame and trial number. Participants wanted to find out more when they were responsible for the outcome (*I play*: 81.29%, *they play*: 72.30%, |Z| = 3.25, *p* = .001) and when it affected them (*for me*: 79.68%, *for someone else*: 73.90%, |Z| = 2.28, *p* = .022). These factors also interacted, such that the decrease in interest when *they play,* compared to when *I play*, was greater when outcomes affected *someone else* (decrease of 15.35%), than when they affected the participant (2.63% decrease, M.D. = 12.72%, (|Z| = 2.10, *p* = .036). There was also a significant interaction between frame and who was affected (|Z| = 4.57, *p* < .0001). This showed that the increase in interest when information was likely to be positive, over negative, was greater when the outcomes affected the participant (10.23% increase), compared to when they affected *someone else* (1.02% increase, 9.21% difference; table 6).

5. Experiment 3

Experiment 2 was designed to control for potential confounds in experiment 1 that could have caused a strong effect of who was playing. If these confounds were present, they may have dominated the task and overshadowed the frame manipulation. Results from experiment 2 support the results from experiment 1 that participants found out more about outcomes that affect them and that who is affected interacted with who was responsible. Experiment 2 also reveals a main effect of who was responsible and importantly, an interaction between frame and who was affected. Participants showed more of a bias towards positive outcomes when they were affected than when someone else was affected.

However, there remains an alternative explanation for increased interest in the outcomes for the *I play* conditions other than being responsible for positive impact. Finding out in these conditions also gives participants insight into how good they are at the task which may be reinforcing in its own right, rather than in relation to gaining money for oneself and someone else. Experiment 3 therefore further adapts the design to remove this possibility.

5.1. Method

5.1.1. Participants

72 new participants took part in experiment 3 but for 13 the data were not recorded. This left 59 participants included in analysis (48 female, 8 male, 1 other, 2 undisclosed) aged 18-46 (M = 20.89, S.D. = 4.18).

5.1.2. Visual perception task

Experiment 3 maintained the change from experiment 1 to 2 in separating in time the visual perception task from the questions on whether participants wanted to find out the outcome. This experiment further eliminated the possibility that participants were finding out to learn about their performance by changing how the positive and negative frames were created. In experiments 1 and 2 these were at the point of playing, through trials to win and others to avoid losing. In experiment 3, the frame was created by telling participants whether that round had been answered correctly (*positive*) or incorrectly (*negative*).

Correct answers lead to increases in payment and incorrect answers lead to decreases in payment. However, not all trials "counted" towards the actual payment - some answers were discounted and did not change the payment. This provided the opportunity for participants to find out whether the trial counted or not, rather than whether it was correct or incorrect. Like the previous experiments, this was only an option on every other trial to prevent tracking of participants own payment, compared to the other players, which we thought may induce a competitive approach, rather than wanting everyone to get the maximum.

As the frame was now only introduced at the point of deciding whether to find out, the blocks for the visual perception tasks only differed in who the recipient of the payment was (*me* or *them*). Participants completed 72 trials of the task, to match experiment 2, meaning 2 blocks of 18 trials, so 36 trials in total, for each recipient.

During the second section, where participants chose whether to find out if the trial counted, the trials were presented in a quasi-random order. The order ensured that more rounds on which they had lost appeared on trials which had an opportunity to find out, than rounds on which they had won. This was because the staircasing, to maintain performance around 70%, meant that participants won on average 70% of the rounds. Full randomisation during the second section would have meant too few loss trials had an opportunity to find out on for the contrasts between positive and negative to be balanced.

Whether trials counted or not was determined for each block of 18 trials in experiment 3. This was randomly determined with a 70% probability of the block counting, to match the 70% chance of positive outcomes, from the staircasing, in experiments 1 & 2.

5.2. Results

In addition to the factors included in the analysis of experiments 1 and 2, analysis of experiment 3 included whether trials in that block counted. The best model involved fixed terms for 5 main effects: trial number, frame (*positive* or *negative*), who was playing (*I play* or *they play*), who was affected (*for me* or *for someone else*), and whether the block counted towards the payment (*counts* or *doesn't count*). The model also included 4 interactions: a) trial number with frame, b) trial number with whether the block counts, c) who was playing with who was affected, and d) who was affected with frame (table 7). Random terms were included for all these predictors except for the interaction between who was affected and frame, as this didn't account for any variance and produced a singular fit.

Table 7
Model parameters for experiment 3

	Odds ratio	Lower CI	Upper CI	|Z|	*p*
Intercept	1.24	0.77	2.01	0.87	.383
Trial number	0.34	0.20	0.58	3.95	<.0001***
Positive – negative	0.71	0.57	0.90	2.90	.004**
Counts – doesn't count	0.94	0.82	1.08	0.91	.362
I play – they play	0.82	0.71	0.94	2.90	.004**
For me – for someone else	0.74	0.65	0.85	4.30	<.0001***
Trial number * frame	0.69	0.53	0.91	2.68	.007**
Trial number * counts	1.35	0.84	2.18	1.22	.221
Playing * affected	0.78	0.68	0.88	3.83	.0001***
Affected * frame	1.12	1.03	1.22	2.51	.012*

*Note. CI: confidence interval, |Z|: absolute value of Z-score. *: p < .05, **: p < .01, ***: p < .001*
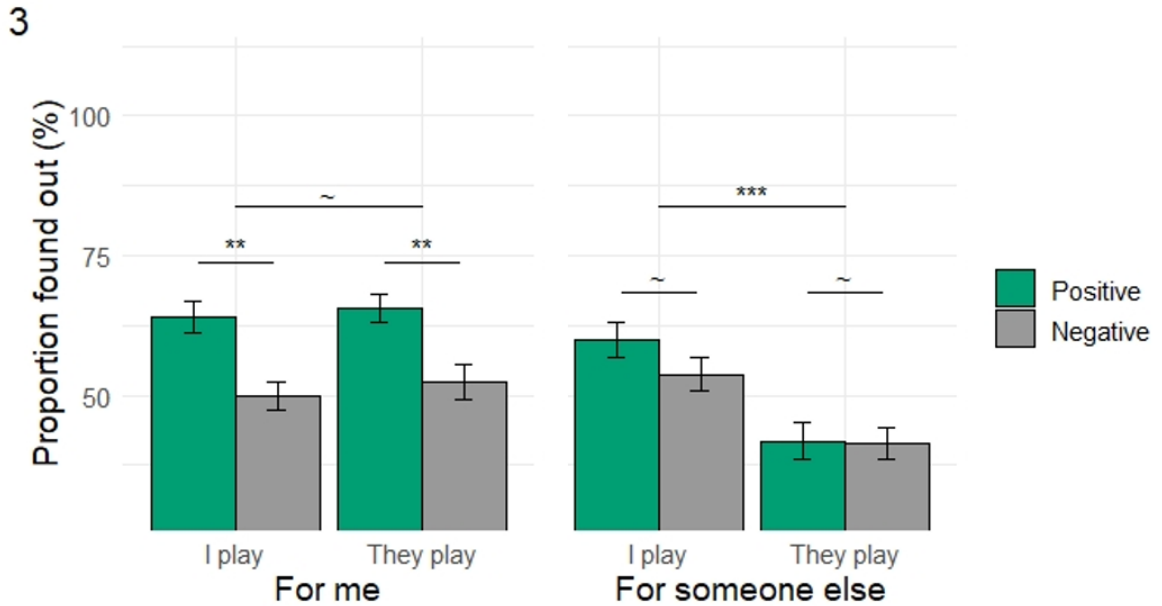
Figure 5. Average proportion of times (%) participants found out the outcome in each condition in experiment 3. Lower lines indicate the level of significance for the effect of frame within that condition, upper lines indicate the level of significance for the effect of who is responsible. ~: p > .05, *: p < .05, **: p < .01, ***: p < .001

Results show that participants wanted to know more often in the *positive* frame (average of 57.78% of trials) than the *negative* frame (49.32%, M.D. = 8.46%, |Z| = 2.90, *p* = .004; table 8). There was also greater interest when *I play* (56.84%), compared to when *they play* (50.26%, M.D. = 6.58%, |Z| = 2.90, *p* = .004), and when the outcome was *for me* (57.89%), compared to *for someone else* (49.21%, M.D. = 8.68%, |Z| = 4.30, *p* < .0001). Finally, the effect of trial number was significant (|Z| = 3.95, *p* < .0001). The fixed terms of whether the trial counted and the interaction between this factor and trial number were both non-significant predictors (*p* = .362 & *p* = .221 respectively). However, these factors improved the model through the random terms and removing them completely significantly decreased the model fit (ΔAIC = 10.8, $\chi^2(4)$ = 18.82, *p* = .0009).

Significant interactions revealed that increased interest in positive outcomes was greater when they affected the participant (increase of 13.65%), than when they affected *someone else* (increase of 3.27%, M.D. = 10.38%, |Z| = 2.51, *p* = .012). As in experiments 1 and 2, the decrease in interest when *they play,* compared to when *I play*, was also greater when outcomes affected *someone else* (decrease of 15.32%), than when they affected the participant, which actually showed an *increase* in interest of 2.16% (M.D. = 17.48%, |Z| = 3.83, *p* = .0001).

Table 8
Descriptive statistics for experiment 3

	For me			For someone else			For me & for someone else		
Plays:	I	They	I & T	I	They	I & T	I	They	I & T
Mean %									
+	63.86	65.57	64.71	59.97	41.73	50.85	61.91	53.65	57.78
-	49.76	52.37	51.07	53.78	41.38	47.58	51.77	46.87	49.32
+ > -	*14.10**	*13.20**	*13.65**	*6.19~*	*0.35~*	*3.27~*	*10.14**	*6.78~*	*8.46**
+ & -	**56.81**	**58.97**	**57.89**	**56.88**	**41.55**	**49.21**	**56.84**	**50.26**	**53.55**
Within-subject standard error									
+	2.84	2.66		3.15	3.33				
-	2.56	3.19		3.13	2.89				

Note. +: positive frame, -: negative frame, + > -: positive frame subtract negative (italicised), + & -: average of positive and negative frame (bold font), I & T: average of when I play and when they play, ~: no significant / *: significant difference between positive and negative frames, when analysis restricted to that condition.

6. Experiment 4

Experiment 3 was the first to show an overall bias towards positive information across who was responsible and who benefitted. Levels of this bias and overall interest were greater when the outcome affected the participant than when it affected someone else. In line with experiments 1 and 2, the drop in interest when *they play,* compared to when *I play* was greater for outcomes affected *someone else*, than outcomes *for me.*

Determining whether the outcome counted or not for each block, rather than each trial, introduces the potential for learning about the blocks. This could mean optimism bias is relevant in addition to the bias towards positive information. For example, if an outcome is positive but doesn't count, it is optimistic to check whether that changes on the next trial. However, this combination of factors can make the results difficult to interpret. Complications for interpretation are particularly relevant if different participants noticed this trend at different rates, as suggested by an improvement in model fit when including the random term for the interaction between whether the trial counts and trial number. Experiment 4 excludes the possibility of learning by randomising whether the trial counts on every trial.

6.1. Method

6.1.1. Participants

44 new participants took part in experiment 4 but the data from 11 were lost leaving data from 33 participants for analysis (22 female, 11 male) aged 19-42 (M =25.13, S.D. = 5.10).

6.1.2. Visual perception task

Experiment 4 was identical to experiment 3 except for whether the trial counted or not was randomised on each trial, rather than each block. This was still done with a 70% chance of the trial counting. The visual perception part and the opportunities to find out remained separate and the number and length of trials were both the same as experiment 3.

6.2. Results

Model comparison for experiment 4 resulted in a model including the main effects of trial number, frame, who was responsible, and who was affected, plus the interaction between who was responsible and who was affected (table 9). All of these were modelled with both fixed and random terms.

Table 9
Model parameters for experiment 4

	Odds ratio	Lower CI	Upper CI	|Z|	*p*
Intercept	1.73	0.66	4.52	1.11	.265
Trial number	0.22	0.11	0.45	4.22	<.0001***
Positive – negative	0.96	0.75	1.23	0.29	.769
I play – they play	0.67	0.50	0.88	2.82	.005**
For me – for someone else	0.52	0.37	0.72	3.87	.0001***
Playing * affected	0.80	0.66	0.96	2.39	.017*

*Note. CI: confidence interval, |Z|: absolute value of Z-score. *: p < .05, **: p < .01, ***: p < .001*
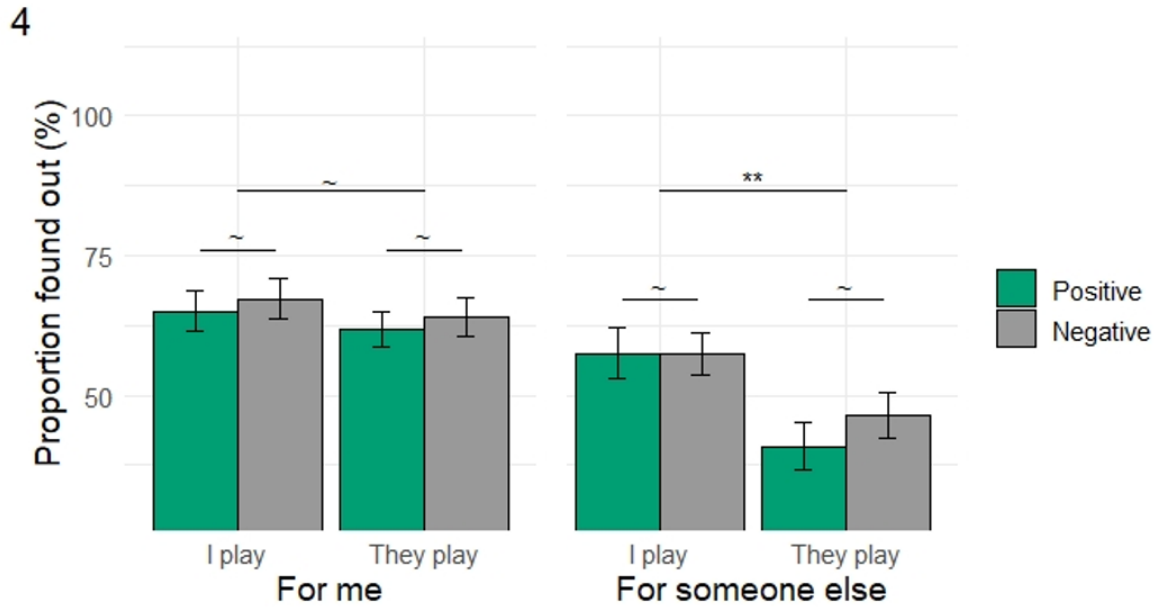
4

Figure 6. Average proportion of times (%) participants found out the outcome in each condition in experiment 4. Lower lines indicate the level of significance for the effect of frame within that condition, upper lines indicate the level of significance for the effect of who is responsible. ~: p > .05, *: p < .05, **: p < .01, ***: p < .001

Significant results were found for all the fixed terms, except for frame ($p = .769$). As in the previous experiments, participants found out more often when the outcome was *for me* (64.56% on average) than *for someone else* (50.56%, M.D. = 14.00%, |Z| = 3.87, $p = .0001$). Interest was also greater when the participant was playing (61.77%), compared to when *they play* (53.35%, M.D. = 8.42%, |Z| = 2.82, $p = .005$; table 10). However, this detrimental impact on interest when *they play*, compared to when *I play*, was smaller when outcomes were *for me* (decrease of 3.12%) than when they were *for someone else* (decrease of 13.72%, M.D. = 10.60%, |Z| = 2.39, $p = .017$). As in the other experiments, interest declined over the task (|Z| = 4.22, $p < .0001$).

Frame was not a significant predictor in the model but removing its fixed and random terms significantly decreased the model fit, suggesting the random term contributed to explaining variance (ΔAIC = 29.9, $\chi^2(2) = 33.88$, $p = < .0001$).

Table 10

Descriptive statistics for experiment 4

Plays:	For me			For someone else			For me & for someone else		
	I	They	I & T	I	They	I & T	I	They	I & T
Mean %									
+	65.04	61.92	63.48	57.55	40.91	49.23	61.29	51.42	56.35
-	67.19	64.07	65.63	57.30	46.49	51.89	62.25	55.28	58.76
+ > -	*-2.15~*	*-2.15~*	*-2.15~*	*0.25~*	*-5.58~*	*-2.67~*	*-0.95~*	*-3.86~*	*-2.41~*
+ & -	**66.12**	**63.00**	**64.56**	**57.42**	**43.70**	**50.56**	**61.77**	**53.35**	**57.56**
Within-subject standard error									
+	3.66	3.16		4.44	4.32				
-	3.61	3.47		3.80	4.09				

*Note. +: positive frame, -: negative frame, + > -: positive frame subtract negative (italicised), + & -: average of positive and negative frame (bold font), I & T: average of when I play and when they play, ~: no significant / *: significant difference between positive and negative frames, when analysis restricted to that condition.*

Table 11

Model parameter |Z| and associated significance across experiments 1-5

	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5
Intercept	1.59	7.20***	0.87	1.11	3.69***
Trial number	2.08*	1.79	3.95***	4.22***	1.91
Positive – negative		1.72	2.90**	0.29	4.00***
I play – they play	2.76**	3.25**	2.90**	2.82**	1.60
For me – for someone else	1.85	2.28*	4.30***	3.87***	N/A
Playing * affected	2.57*	2.10*		2.39*	N/A
Affected * frame		4.57***	3.83***		N/A
Trial number * affected	0.98		2.51*		N/A
Frame * responsible					4.66***
Trial number * responsible					1.28

Note. |Z|: absolute value of Z-score. N/A: not relevant for the study design, blank: not included in the best fitting model.

*: p < .05, **: p < .01, ***: p < .001

7. Individual differences: experiments 1-4

The final aspect of the analysis of experiments 1-4 was to see whether any of the main effects described above varied with how generous or how high on the IRI the participant scored. We ran this analysis on all the data from experiments 1-4 together and in table 12 present the full model with all interactions, regardless of their contribution to the model, for comparison.

7.1. Generosity

Averaging across the 7 questions on which participants could donate some of their winnings to charity provides an overall measure of generosity for each participant. The main effect of this trait did not predict interest in outcomes overall ($p = .264$). However, the measure interacted with several of the conditions of interest, as we would predict if the task is relevant for understanding altruism.

More generous participants differentiated between who was affected less ($|Z| = 4.35$, $p < .0001$). The decrease in interest when outcomes were *for someone else*, compared to *for me*, was smaller than for more selfish participants. Interestingly, more generous participants also showed a *greater* effect of who was responsible ($|Z| = 4.02$, $p < .0001$). The increase in interest when *I play*, compared to when *they play*, was larger for more generous participants. Finally, levels of generosity showed a three-way interaction with the frame (*positive* or *negative*) and who was responsible (*I play* or *they play*; ($|Z| = 2.61$, $p = .009$). This showed that the level of differentiation between frames, depending on who was responsible, was larger for more generous participants. Less generous participants did not show more or less bias towards positive information depending on who was responsible.

7.2. Interpersonal reactivity index

Like generosity, average scores on the IRI questionnaire did not predict overall levels of interest ($p = .844$). Like generosity, participants higher on IRI showed less distinction between outcomes for themselves and outcomes for others ($|Z| = 3.50$, $p = .005$). Average IRI scores also showed an interaction with frame ($|Z| = 2.20$, $p = .028$) such that participants scoring higher on the IRI showed less bias towards positive information than participants lower on the IRI. This seemed to be driven by increased interest in negative outcomes for high IRI participants, particularly based on the personal distress subscale.

Table 12

Model parameters for individual differences models with data from experiments 1-4

	Odds ratio	|Z|	*p*	Odds ratio	|Z|	*p*
	Generosity			*IRI average*		
Intercept	2.01	4.74	<.0001***	1.92	4.39	<.0001***
Trial number	0.52	11.59	<.0001***	0.50	12.53	<.0001***
Positive – negative	0.83	8.10	<.0001***	0.84	7.82	<.0001***
I play – they play	0.77	10.91	<.0001***	0.76	11.82	<.0001***
For me – for s. else	0.79	10.12	<.0001***	0.79	10.37	<.0001***
Trait	0.95	1.12	.264	0.91	0.20	.844
Frame x Playing	1.02	0.84	.401	1.02	0.72	.471
Frame x Affected	1.07	3.00	.003**	1.07	3.01	.003**
Playing x Affected	0.82	8.64	<.0001***	0.81	9.19	<.0001***
Frame x Trait	1.00	0.57	.567	1.18	2.20	.028*
Playing x Trait	0.97	4.02	<.0001***	0.94	0.77	.443
Affected x Trait	1.03	4.35	<.0001***	1.30	3.50	.0005***
F x P x A	1.03	1.36	.173	1.04	1.91	.056
F x P x T	0.98	2.61	.009**	1.07	0.93	.353
F x A x T	0.99	1.49	.136	1.10	1.30	.195
P x A x T	0.99	1.90	.057	0.88	1.72	.085
F x P x A x T	1.00	0.29	.773	0.92	1.09	.273

*Note. s. else: someone else, F: frame, P: playing, A: affected, T: trait – charitable giving on the left / IRI average on the right, x: interaction, |Z|: absolute value of Z-score, bold font: terms including the trait. *: p < .05, **: p < .01, ***: p < .001*

8. Experiment 5

Taken together, experiments 1-4 show people are more interested in outcomes that affect them and there is some evidence of bias towards positive outcomes when this is the case. For outcomes affecting others, interest was still relatively high (over 40% in all experiments). Who was responsible had a greater effect on interest in outcomes for someone else, with increased interest when the participant was responsible (table 11).

Evidence of a bias towards positive outcomes was mixed. A main effect of frame was only found in experiment 3 but in both experiment 2 and 3, there was more of a preference for positive outcomes *for me* that those outcomes *for someone else*. Results from the probability task show a similar pattern of distinguishing between frames more at higher probabilities, both for outcomes *for me* and *for someone else.*

An issue with relating the results from any of the above results to action or outcome-oriented altruism is that they were not altruistic decisions, but outcomes based on performance on a task without a moral component. Experiment 5 applies the principle of bias towards positive outcomes caused by the participant or a third person to charitable donation decisions.

Asking participants to decide whether to donate or not naturally introduces variability between individuals in how many times they chose each option so how many trials are available in each condition. For example, if a participant never donates, there cannot be a positive condition because there cannot be any questions about whether their donation had a positive impact. We therefore ran this experiment on Amazon Mechanical Turk through TurkPrime and using Qualtrics Online Survey Software (Qualtrics, Provo, UT) to test enough participants so that there were adequate numbers with choices both to donate and not donate.

8.1. Method

8.1.1. Participants

370 participants (196 female, 169 male, 3 identified as other, 3 undisclosed) aged 18-77 (M =36.04, S.D. = 11.92) completed the online study with satisfactory answers to attention checks. Participants were required to be in America although there are methods to make an IP address look like the participant is in America when they are not. Data were collected in the evening local time on a Friday and a Sunday (approximately half the sample

on each day). Participants received a baseline payment of $3 and a bonus of $2, minus the amount they had donated to the cause on a randomly chosen trial.

8.1.2. Proportion dominance task

Experiment 5 did not use the visual perception or probability tasks.

25 questions asked participants how much, if anything, up to a maximum of $2 they wanted to donate to charity. All questions were in the format "X people are in need of Y" and varied the number of people in need (X) and the outcome they need (Y). The numbers of people were grouped in 5 bins: 1 person, 2-10 people, 11-50 people, 51-500 people and 501 – 1000 people. The outcomes were in 5 categories: vaccinations (against 5 different diseases), surgery (for 5 different joints), food & water (e.g. rehydration salts), medication (against 5 different diseases) and maternal health (e.g. a caesarean). Participants were told that the people in all cases could die if they do not get the help they need. Numbers of people and outcomes were pseudo-randomly paired such that one number from each bin was with one cause from each category. 5 question sets were created in this way and participants were randomly allocated to a set.

After participants had completed all the proportion dominance donation questions, they read that the researchers running the study were also donating additional money to charity. Whether the participant had donated / not donated to that charity in the first part acted as a "vote" for / against that charity which would increase / decrease the amount that the researchers gave to that cause.

The slight dissociation between the moral decision of the participant and the money for the charity was necessary to avoid competition between money for the participant and money for the charity. If the measure had been whether their own donation would be applied or not, participants may have wanted to find out because of their interest in getting a larger bonus if the money was not donated.

In line with the previous versions of the task, participants were told 70% of these votes would count and the measure of interest was whether they wanted to know whether it counted or not. 28 trials were presented across a maximum of 4 conditions varying across the decision made: to donate (positive) or not donate (negative) and whether it was the participants own decisions (I decide) or another person's decisions (they decide).

These were fully crossed with 7 trials in each condition when possible but dependent on the participants' decisions. If they always donated or never donated, a condition was lost. If they donated on n of the 25 proportion dominance trials and n < 7, the

positive self condition had n trials and the negative self condition had 14 – n trials. The reverse was true if they didn't donate on less than 7 trials. In the other conditions there were exactly 7 positive and 7 negative trials. No information about the number of people in need or outcome they need was presented in this task, each trial just stated for example "You donated so voted to increase the donation. Click "find out" to see whether the donation will actually be increased or not".

Questions were presented in blocks for each condition, like the versions used in the experiments reported above. Also, like experiments 1-4, participants had to actively choose to find out by clicking "find out" and if they did not do so within 6s the screen moved on and they did not find out. This longer time interval was used so that slow internet connections did not mean the interval became too short to read the information and because there were no screens at the start of blocks indicating a new condition. If the participant did not find out, "..." was displayed instead of whether the donation changed or not so that not finding out did not speed up finishing the study.

8.1.3. Questionnaire measures

Following the main tasks, participants reported the causes they usually donate to, the extent to which "I am motivated to give when I hear an emotional story about a person or animal in need" on a sliding scale and, separately, the extent to which "I am motivated to give when I hear statistics about how many people are in need". Belief in the study was rated on a 4-point scale and participants could specify any aspects of the study they did not believe.

8.1.4. Analysis

We followed the same process of analysis for experiment 5 as the visual perception task in experiments 1-4, but without who was affected as outcomes were always for the charity.

8.2. Results

The best model for experiment 5 involved main effects of trial number, frame (*positive* or *negative*), and who was responsible (*I decide* or *they decide*) plus interactions between a) who was responsible with frame and b) who was responsible with trial number (table 12). Each of these factors had a fixed and a random term.

Table 12

Model parameters for experiment 5

	Odds ratio	Lower CI	Upper CI	|Z|	*p*
Intercept	1.73	1.29	2.31	3.69	.0002***
Trial number	0.90	0.80	1.00	1.91	.057
Positive – negative	0.81	0.73	0.90	4.00	.0001***
I decide – they decide	0.93	0.86	1.02	1.60	.111
Responsible * frame	1.25	1.14	1.37	4.66	<.0001***
Trial number * responsible	1.09	0.95	1.25	1.28	.202

Note. CI: confidence interval, |Z|: absolute value of Z-score.

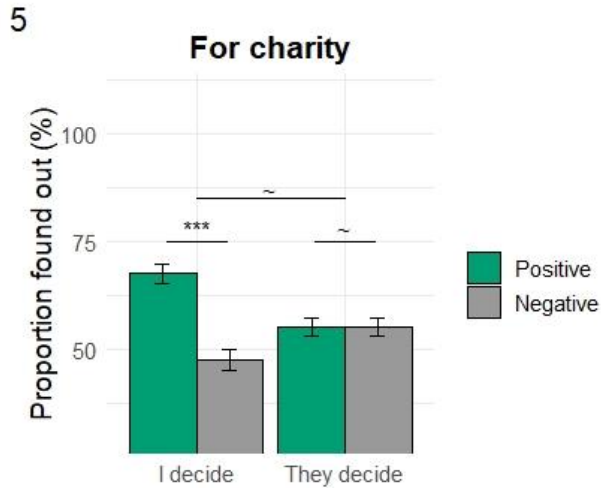**: p < .05, **: p < .01, ***: p < .001*

Figure 7. Average proportion of times (%) participants found out the outcome in each condition in experiment 5. Lower lines indicate the level of significance for the effect of frame within that condition, upper lines indicate the level of significance for the effect of who is responsible. ~: p > .05, *: p < .05, **: p < .01, ***: p < .001

Table 13

Descriptive statistics for experiment 5

Decides:	I	They	I & they	I	They
Mean %				Within-subject SE	
+	67.80	55.44	61.62	1.12	0.90
-	47.75	55.48	51.62	1.04	0.94
+ > -	*20.05**	*-0.04~*	*10.00**		
+ & -	**57.58**	**55.46**	**56.62**		

Note. SE: standard error, +: positive frame, -: negative frame, + > -: positive frame subtract negative (italicised), + & -: average of positive and negative frame (bold font), ~: no significant / *: significant difference between positive and negative frames, when analysis restricted to that condition.

Results revealed a main effect of frame such that participants found out more when the outcome could be *positive* (average of 61.62%), compared to when it could be *negative* (51.62%, M.D. = 10.00%, |Z| = 4.00, p = .0001). Importantly, frame also interacted with who was responsible (|Z| = 4.66, p < .0001). When someone else made the decision, there was no difference between positive and negative frames (0.04% more interest in negative outcomes; table 13). However, when the participant was responsible, they found out about outcomes on average 20.05% more. The fixed term for the interaction between trial number and who was responsible was not a significant predictor but removing the random and fixed terms for this interaction significantly weakened the model fit (ΔAIC = 6.4, $\chi^2(2)$ = 10.36, p = .006).

9. Discussion

Here we present 5 experiments that use different variations of tasks measuring how often participants find out about positively or negatively framed outcomes that affect themselves or another person. We also manipulated the probability of the outcome occurring (probability task) or who was responsible for the outcome: the participant or a third person (visual perception tasks and experiment 5). These conditions allow us to test for a) the presence of a bias towards finding out information likely to be positive when it affects another person and b) the role of responsibility in determining levels of this bias and general interest, in the context of separating action from oriented altruism.

In the probability task, a similar pattern of results was found for outcomes *for me* and *for someone else* - interest increased for positive and decreased for negative outcomes as likelihood of the outcome increased. However, in the versions of the task where a bias towards positive information was found for outcomes affecting the self, it was not present or decreased for outcomes affecting other people. Overall interest was also lower for outcomes affecting other people, particularly when a third person, not the participant, was responsible. This effect of responsibility was not found for outcomes affecting the participant themselves – interest remained high regardless of who was responsible.

9.1. Differences across experiments

The pattern of results described above was found in experiments 2, 3 and 5 but not in experiments 1 and 4 for the visual perception task. This could indicate unreliable results but there are also explanations based on differences between the versions of the task, as we were developing and improving it across the experiments. For the original design in experiment 1, a gain frame was created by trials where the player could win money and

contrasted with a loss frame through trials where the player had to get it right to avoid losing money. However, no bias towards outcomes in the positive frame was found for any of the conditions.

One possible explanation for the lack of positivity bias in experiment 1 is that participants were finding out whether they were right or wrong immediately after the relevant round of the visual perception task. Interest in this context could have been determined by how ambiguous that round of the perception task had been. This could explain the main effect of responsibility for the outcome and this effect may have led to the frame as positive or negative being overlooked. Experiment 2 separated the phase of completing the rounds of the visual perception task from the opportunity to find out.

Experiment 2 shows the expected pattern of results as described above. However, a limitation is that finding out about the result of the trial in conditions where the player themselves were playing also gave information about how good they were at the task. Only having the opportunity to find out on every other trial does minimise participants' ability to keep track of performance. This confound also could not explain the differences between conditions in the levels of bias towards positive frames. However, we cannot confidently rule out that it contributes to the main effect of responsibility for reasons other than those linked to altruistic concern.

Studies 3 and 4 further separated the performance by changing how the gain and loss frames were created. At the point of deciding whether to find out about the outcome, participants were now told whether they had got the trial correct or incorrect but were asked whether they wanted to know if it was one of the 70% of trials which actually counted towards the final payments. Keeping this to every other trial prevented tallying of payments to avoid a more competitive mind-set of which player had gained the most money overall.

9.2. Bias towards positive information

The first aim of the research was to test whether people prefer to find out about outcomes likely to be positive, compared to negative, when they affect other people. This builds on existing literature establishing this bias in relation to outcomes affecting oneself (Charpentier et al., 2018). Related to this bias is an optimism bias, an overestimation of positive outcomes and increased updating of beliefs in a positive, compared to a negative direction (Sharot, 2011). This optimism bias also applies to other people, depending on concern for them (Kappes et al., 2018). Our research is the first to extend this work to measure interest in outcomes affecting others. In the probability task, we find that the

increasing distinction between positive and negative frames, as the outcome gets more likely, is the same for other people as it is for one's own gains and losses.

9.3. Action and outcome-oriented motivations

The second research question was the role of responsibility in determining levels of bias towards positive outcomes and general interest. This was motivated by the aim of developing a behavioural task to separate action from outcome-oriented altruism (Kuss et al., 2013). In the probability task, no one was responsible for the outcomes, they were determined by the computer. That participants differentiated between positive and negative outcomes suggests they were interested in outcomes for other people, independent of responsibility, to some extent.

However, in the visual perception tasks and donation decision task, responsibility had a significant effect on interest in outcomes for others. This suggests an additional value of being responsible for outcomes, in line with action-oriented motivations. The presence of both action and outcome-oriented motivations in our results supports previous findings from fMRI measures (Harbaugh et al., 2007; Kuss et al., 2013). In turn, this suggests our tasks are measuring processes relevant for altruism.

Further evidence of the relevance of these new tasks for understanding altruism comes from our individual differences results. Participants who were more generous on the charitable donation questions in the experiments and those higher in empathy showed less of a decrease in interest for outcomes affecting others, compared to themselves, than those lower on these traits. However, for more generous participants, being responsible for the outcome had a greater effect. If the importance of responsibility reflects action-oriented motivations, this result contrasts with the fMRI studies, which found outcome orientations were more relevant for more generous participants.

9.4. Limitations and future directions

In addition to the limitations of early versions of the task that later versions were designed to overcome, there are further improvements that could be made. We are currently collecting more data on the version of the task presented here in experiment 4 to check whether the lack of evidence for a bias towards positive information is robust or due to low quality data.

Data collection is also currently ongoing for a further experiment which changes the decision to find out from active to forced. In all of the experiments presented here, to find out the outcome required participants to press a keyboard key (experiments 1-4) or

onscreen button (experiment 5). If they did nothing, they did not find out and instead saw a fixation cross for the length of time the outcome was displayed for to keep the length of the experiment the same, regardless of choices.

We chose this design for two reasons. First, it most accurately represents real-world situations which often require an action or effort to obtain more information. Second, as finding out was not costly, we were concerned about a ceiling effect if the choice were forced between find out or not as finding out provides more variety in the task. However, there is evidence of a relationship between positive frames and action, compared to negative frames and inaction and there is a risk this accounts for increased interest in the positive conditions. Results from the final experiment where participants are forced to press one key to find out or a different key to not find out will provide insight on this issue.

Another issue to note is that the change in how the frames were created causes a difference in the payoff matrix probabilities between 1 & 2 (can win / might lose) and 3 & 4 (won or lost). However, there are no quantitative comparisons between the experiments and changing the payoff matrix allowed us to maintain the 70% probability of an outcome. Probability was shown to be an important factor in the results from our probability task.

Finally, only experiment 5 studies interest in outcomes caused by moral decisions. Removing the moral aspect in experiments 1-4 was the only way we found to create balanced conditions between positive and negative conditions. However, in the visual perception task, participants chose to put effort into completing the task correctly to earn more money for other people. This reflects real-world situations of investing effort to benefit others (Lockwood et al., 2017).

10. Conclusion

Participants showed a bias towards positive information affecting other people when outcomes were randomly determined. This bias increased as the likelihood of the outcome increased, just like outcomes affecting the participant themselves. This provides evidence of outcome orientations for interest in other people. However, when outcomes were either caused by the participant or another person, and affected the participant or another person, interest dropped if the participant was not involved at all. Increased interest when the participant was responsible for outcomes affecting others suggests action-orientation is also important for interest in the welfare of other people.

General discussion

In this thesis, I have presented a literature review then five research chapters, in which I applied neuroimaging, physiological and behavioural methods to the research questions: when and why do we value the lives of others? In addition to different methods, the papers consider different aspects of these questions and use different tasks. In this section I will bring together the findings from all the papers in relation to the research questions, before considering the contributions and limitations of the work as a whole.

1. Summary

1.1. Research questions

As outlined in the literature review, there are many reasons that other people can have value to us. The first key distinction is whether other people only provide value through how they can increase our own resources, such as money. This is separate to valuing other people in their own right. Such a distinction is captured in the comparison between strategic and altruistic decisions in our fMRI meta-analysis (paper 1). Before answering the main research questions of when and why we value the lives of others, it is first important to establish whether people do in fact value the lives of others, independent of outcomes for themselves.

1.1.1. Do we value the lives of others, independent of strategic motives?

As outlined in the literature review, if it is the case that people value the lives of others, or in other words care about their wellbeing, we would predict that people would:

i. Get distressed when bad things happen to others
ii. Feel good when good things happen to others
iii. Take actions to improve outcomes for others
iv. Be interested in what happens to others
v. Show similar biases in valuing outcomes for self and other (on the basis that people value their own life and wellbeing)

Adding to existing evidence covered in the literature review, the studies in this thesis support the idea that people do value the lives of others. I will now outline this evidence for each of the criteria in turn.

People are distressed when bad things happen to others

The idea of empathic distress when others are suffering is certainly not new (Batson, 2014). The findings here extend evidence of empathic distress through physiological distress responses, in the brain and body, to brief written descriptions of other people at risk and dying. In the body, skin conductance responses (SCRs) are a proxy for arousal and the findings show this responded to lives being at risk and the outcomes. In the brain, we saw responses to the deaths of other people in the anterior insula (AI) and anterior cingulate cortex (ACC), regions that respond to experiencing pain and observing more vivid depictions of other suffering (Lamm et al., 2011).

A trait measure used in the studies is the Interpersonal Reactivity Index (Davis, 1983), often considered synonymous with empathy. Several of the items in this questionnaire, particularly on the personal distress subscale, measure self-reported distress responses to negative events affecting other people. The fact that participants rate themselves as above minimal on these items complement our findings of physiological and neural responses in the tasks. Patterns of SCRs and fMRI results also differed between those high and low on this trait.

Finally, in the behavioural task measuring interest in outcomes for others, there was some evidence that negative information was avoided, compared to positive information, suggesting negative utility. This pattern did depend on whether the participant was responsible for the outcome, as discussed below.

People feel good when good things happen to others

Like empathic distress, the experience of vicarious reward, a positive feeling when good things happen to other people (Mobbs et al., 2009), is supported by the studies in this thesis in different ways. The news stories task was designed in a negative frame, such that the best outcome was simply surviving the risk of death. Despite this, findings from fMRI and SCR, suggest that people surviving, compared to dying, was less distressing or arousing in some contexts. This shows participants were differentiating the 'positive' outcomes from the negative ones. Participants from the fMRI scope insensitivity study also rated how emotional they found the stories and clearly differentiated between positive and negative, rating outcomes where people lived as above neutral.

In the fMRI meta-analysis, participants in the original studies were able to create positive outcomes for others by giving them money. Results show activation during these altruistic decisions across reward-related regions in the brain. We can describe this as the

neural basis of 'warm glow' experienced when helping others. Positive feelings and reward activity during altruistic decisions could be attributed to a range of factors, depending on the motivations for helping, as described below. Finally, as mentioned above, the interest in positive outcomes for other people, shown by finding out more often than for negative outcomes, suggests that good things happening to others has a positive value.

People take actions to improve outcomes for others

Experiencing empathic distress and vicarious reward provides motivations to minimise suffering and maximise positive outcomes for others. Helping behaviours are not sufficient evidence that people want to improve outcomes for others, as there are alternative explanations as discussed below. However, actions to improve outcomes for others are arguably necessary to make the claim that people value the lives of others.

Helping can be non-costly but our findings support the idea that prosocial motivations are powerful enough to promote effort and sacrifices to help others. The existence of an altruistic condition in the fMRI meta-analysis was only possible because participants in the studies we included made costly donations to other people and charities. Participants across the other studies were given opportunities to make donations as part of the individual difference measures and many did donate from their payment for participation.

People are interested in what happens to others

If positive outcomes for others are rewarding and negative outcomes are distressing, people should be interested in what happens to other people, to the extent to which they think the news will be positive. We showed evidence for this in results from paper 5 on interest in outcomes affecting others, although the level of interest depended on responsibility for the outcome.

In the news stories task, the positive outcome was really a baseline of avoiding the risk of death. The fact that responses, in brain activity and physiological responses, were different between live and die outcomes suggests that participants were paying attention to the outcome and interested to some extent. Similarly, the phrasing of the stories as "at risk of death" would only generate an arousal or negative emotional response if the outcome was of interest.

People show similar biases in valuing outcomes for self and other

While the focus of papers 3, 4 and 5 was on outcomes for other people, our control conditions measured responses to outcomes for the self that varied on the same factors as the social stimuli. For example, in the scope insensitivity tasks, trials presented different amounts of money or credits that could be lost, to match the numbers of people. Positive outcomes were keeping these resources while in negative trials they were lost. In paper 5, the condition of who was affected (me or someone else) was fully crossed with the positive and negative manipulation and either who was responsible (visual perception task) or the probability of the outcome (probability task).

Results showed mixed evidence for whether the biases of scope insensitivity and towards positive information were the same for self and other outcomes. In the probability task in paper 5, the change in interest over different probabilities was similar for outcomes affecting the participant and someone else, although interest in 'for me' trials was generally higher. However, in the visual perception task, who was responsible for outcomes did not matter when they affected the participant but did matter when they affected someone else. In the fMRI study, the pattern of responses to credit trials had some similarities with lives abroad, but lives in the UK were more distinct.

1.1.2. Why do we value the lives of others?

As I have outlined above, the results in this thesis contribute to the evidence in the literature review that other people have value. This section considers the possible sources of that value. These align with the different categories of motivations for prosocial behaviour as separated in the literature review.

Other people can increase our own extrinsic gains

The focus of this thesis and the section above is the value of other people's lives, independent of whether they contribute to our own extrinsic resources such as money. However, this is a key element of prosocial behaviours that involve reciprocation or cooperation (Camerer, 2003; Camerer and Fehr, 2003). Extrinsic reward as motivation for prosocial behaviour is only considered in the fMRI meta-analysis in paper 1 through the strategic group. Results show strategic prosocial decisions correlate with activity across reward-related regions. The fact that this was greater than during altruistic decisions shows the additional or stronger reward of strategic decisions that can enhance both intrinsic and extrinsic rewards. We purposefully excluded the possibility of strategic motivation in all the other studies through making behaviours as anonymous as is possible in a lab experiment.

It is distressing when other people are suffering

In contexts where the strategic motivation to help is not available, help is commonly in response to the suffering of others. As outlined above, results from the news stories tasks show negative emotional responses in the brain and body when others are at risk of death and die.

Evidence that people experience distress when others are suffering could be considered as both evidence of valuing other people and a reason that other people have value to us. If the wellbeing of others affects our mood, then their lives are important and relevant so valuable. For example, considering the people we value the most, it is hard to distinguish cause and effect: does their wellbeing affect our mood the most because we value them the most, or do we value them because their wellbeing affects our mood? The answer to this question may just be in the semantics of how value is defined.

We can personally make a positive difference

The opportunity to reduce the suffering of another person is an opportunity to have a positive impact on something and this could be rewarding, independent of the change in the other person's wellbeing (Aknin et al., 2013b). This is the key difference between action and outcome-oriented altruism as described in the literature review (Kuss et al., 2013). Altruistic decisions in the fMRI meta-analysis could be driven by either or both motivations, as the participant was always directly responsible and actively chose to donate. Two of the eligible studies (Harbaugh et al., 2007; Kuss et al., 2013) had designs able to separate these motivations and showed an additional value of being responsible for the altruistic outcome.

The behavioural studies measuring interest in positive and negative outcomes for self and other, caused by either the participant or a third person, were specifically designed to separate these motivations behaviourally. Results from some versions show greater interest and bias towards positive information when the participant was responsible for the outcome. This supports action orientation and the idea that other people have value through providing an opportunity to make a positive difference.

It is rewarding when other people are better off

In contrast to action-orientated altruism, where value comes from personally making a difference, outcome-oriented altruists should also value positive outcomes for other people that they are not responsible for. In the news stories tasks, the participants were never responsible for the outcomes, but we still found a reduction in negative responses to the positive (or less negative) outcomes. In paper 5, interest in positive

outcomes caused by a third person was still above 40% on average. In the donation decision version of the task (experiment 5), participants were more interested in outcomes caused by others than negative outcomes they caused. In the probability task, the increasing interest in positive, over negative, outcomes at higher probabilities was found for others as well as the self.

In summary, other people have value as they can contribute to increasing our own resources, provide an opportunity to make a positive difference and affect our mood, both positively and negatively. In the real world, these are likely to work together in many cases and be indistinguishable from each other based on behaviour alone. A key contribution of this thesis is to combine novel tasks with measures of physiology, neural responses and interest in outcomes to separate and measure these components of the value of other people.

1.1.3. *When do we value the lives of others?*

Situational factors

As touched on above, the different ways that other people are valuable depend to some extent on the context, as certain motivations are only possible in certain contexts (indicated in parenthesis below). There are also contextual factors that can increase or decrease the value we attribute to a life. These differences are labelled biases because they diverge from the belief that most people hold at a moral, cognitive level that all lives are equal (Slovic et al., 2011). Several such factors are considered in this thesis and results suggest that the value of another person increases when:

i) *They can return the favour (strategic motivation)* – demonstrated by strategic prosocial decisions correlating with increased activity in reward-related regions (paper 1).

ii) *When we can personally make a difference (action-oriented motivation)* – supported by increased interest in and bias towards positive over negative outcomes when personally responsible for the outcome (paper 5).

iii) *When they are in the same country* – shown through increased overall arousal, and differentiation of live from die outcomes, for lives in the UK, compared to abroad (papers 3 & 4).

iv) *They are in a medium sized group* – suggested by non-linear relationships between responses in the brain and body and the number of people at risk and affected (papers 3 & 4).

v) *When the news is likely to be positive* – supported by greater interest in outcomes for other people that could be positive, compared to those in a negative frame (paper 5). Note this is specific to our measure of wanting to find out and other measures, such as donations, can be increased through negative frames (Erlandsson et al., 2018).

Individual differences and interactions with situational factors

In addition to the situational factors studied in this thesis, there are also individual differences that affect whether and how much someone values the lives of others. Some key examples of these traits are described in the literature review. In the studies presented here, we focused on the relationship between individual differences in interpersonal reactivity or empathy (Davis, 1983) and the situational factors of interest.

The relationships between interpersonal reactivity and the effect of the variables on physiological responses to news stories and interest in outcomes affecting other people suggest that processes linked to empathy may be relevant for these effects. An ongoing question in relation to empathy is whether it promotes prosocial behaviour in a fair and more utilitarian way or only towards certain people or groups (Bloom, 2014). Our results show that participants higher on this trait showed levels of interest in outcomes affecting others that were closer to the level of interest in outcomes affecting themselves. As the others were strangers, this could be interpreted as a general increase in prosocial emotions. However, in the news stories task, the maximal response of high empathy participants was to smaller groups of people than those lower on the scale. This greater scope insensitivity suggests biases may be stronger for more empathic people.

1.2. Contributions of the thesis

The above summary of the research findings brings together the conclusions of the 4 main empirical papers, in relation to the research questions of when and why we value the lives of others. The fifth paper, paper 2, is a methodological paper. Overall, the thesis contributes to the related literature and methodology, as well as having practical implications for prosocial behaviour in the real world. In the following section I will expand on each of these in turn.

1.2.1. Methodological contributions

Technique to account for variable coverage in fMRI meta-analysis

As part of the fMRI meta-analysis on prosocial decision making (paper 1), we also developed a new technique to account for the issue of variable coverage, a problem affecting all studies using this method. This technique is described and demonstrated in paper 2 and the code, alongside example data, is freely available at doi.org/10.25377/sussex.c.4223411. A recent review highlighted the importance of our adjustment in detecting activity in the subgenual ACC during moral decision-making tasks (Zahn et al., 2019). Our technique is also being used by other researchers currently working on an fMRI meta-analysis project.

Novel tasks and stimuli

To achieve the theoretical contributions described below, we designed several new paradigms across the studies in this thesis. Developing the news stories task involved creating a large novel set of stimuli, all from real news stories, that are balanced across conditions, based on ratings from pilot participants and the type of risk (natural disaster, disease etc.). This task also had to meet requirements for SCR measures – maximising risk and uncertainty – and fMRI, both of which require many trials within subjects. In sharing these stimuli, alongside the paper when published, we hope that other researchers may utilise them.

Another aspect of developing the tasks was to create paradigms suitable for within-subject designs. Between-subject designs are less powerful, particularly to measure individual differences in the effects. The task we developed for the behavioural study on finding out about outcomes for others allowed within-subject measures of both action and outcome orientations.

Dynamic causal modelling of SCRs in a social task

As part of the analysis of our SCR data, we applied dynamic causal modelling to separate responses with an interstimulus interval shorter than the time for responses to return to baseline (Bach et al., 2010). To our knowledge, this was the first application of such a method to responses during a social task. Given the novel task, with advice from the software creators, we decided on a method to fit different models to the responses, blind to condition, and compare these to identify the best timings. Our results benefitted from being able to analyse sudomotor burst amplitude and dispersion separately, as these showed different relationships with our variables of interest. We hope that this can guide other

researchers wanting to apply this advanced technique to SCRs during similarly new paradigms.

1.2.2. Theoretical contributions

The literature review at the start of this thesis and discussion section of each paper highlight the contribution of specific findings and how they link to the rest of the field. Here, I focus on the theoretical contributions of the findings as a whole, as described in the summary of research findings in the previous section.

Moving from a psychological to a biological understanding

There is a considerable amount of existing research on the topics covered in this thesis, as described in the literature review. The first paper, an fMRI meta-analysis, was only possible thanks to previous work and the generosity of authors in sharing their data. While this was a summary of neuroimaging findings on the topic of prosocial decision making, for other topics, the studies in this thesis are the first to apply physiological and neuroimaging techniques.

Previous work on scope insensitivity and physical proximity, related to prosocial behaviour or valuing other people, has only used behavioural measures (Dickert et al., 2015, 2012; Fetherstonhaugh et al., 1997; Friedrich et al., 1999; Slovic, 2010; Slovic et al., 2011; Small et al., 2007; Touré-Tillery and Fishbach, 2017; Västfjäll et al., 2014; Zagefka, 2018). These often rely on self-report and between-subject designs, which have several limitations. By integrating techniques from neuroscience to study the brain and body with within-subject experimental designs, the current studies contribute unique insight.

Issues with self-report of giving socially desirable answers or being unaware of one's own biases are particularly relevant given the subject matter (Fernandes and Randall, 1992). This is demonstrated by our findings that participants in the scope insensitivity studies do not report differences between the value of a life in the UK compared to abroad. In self-report, participants also increased their willingness to pay to save an increasing number of lives, albeit not on a linear scale. In contrast, results from the neuroimaging and physiological measures showed a collapse in responses at the highest numbers of people and greater responses to events in the UK, compared to abroad.

Moving from the economic to the social domain

Several of the concepts in this thesis have roots in economics and here I have applied them to the social concept of valuing other people. These concepts include scope

insensitivity; economic games, as used by studies in the fMRI meta-analysis; and the concept of valuation or value-based decision-making in general. To measure these concepts in relation to people has involved the creation of multiple new tasks and stimuli sets, as outlined above in the section on methodological contributions. The results from these tasks also contribute towards an ongoing theoretical shift away from the neoclassical economic view that people are selfish and only motivated by extrinsic rewards for themselves (Adamus, 2017).

Talking about other people having value to us could be interpreted as supporting the selfish view of human nature, rather than the existence of altruistic motivations. However, I argue this is a false dichotomy, as an action will only ever be taken, particularly reliably and repeatedly, if it has some value. Given that prosocial behaviours are common in the real world (UNICEF, 2018) and in experimental settings (Engel, 2011), being prosocial must have value.

In understanding prosocial behaviour, the question is not whether valuing others is selfless but what the source or nature of the value is. The results presented here show evidence for many different sources of value, some of which are more self-focused and some of which focus more on the other person. Traditionally, a key distinction has been drawn between other-focused empathic motivations for helping and selfish relief of one's own distress at witnessing suffering (Batson et al., 1987; Schroeder et al., 1988). However, insights from neuroscience, including our fMRI results, challenge this clear distinction. Empathy for suffering is likely processed in the same brain regions as experiencing pain or distress (Lamm et al., 2011), although distinctions within these regions have been identified in some studies (Lamm et al., 2019).

The idea that altruistic motivations are not necessarily undermined if we experience reward or warm glow from being prosocial is reflected by people's perceptions of others (Barasch et al., 2014). Given the likely role of reinforcement learning in prosocial behaviour (Gęsiarz and Crockett, 2015), it is possible that warm glow generalised from helping close others (Telzer et al., 2010), who could have reciprocated, to more distant others.

1.2.3. Practical implications

Prosocial behaviour and altruism outside the lab is vital for positive social functioning and has the power to save and improve lives (UNICEF, 2018). Research on these topics therefore has important implications for maximising the benefits of prosocial behaviour. This may be particularly useful for fundraising organisation appealing to potential donors. Each of the conclusions on when people value the lives of others most

could suggest a way to increase prosocial behaviour. However, these should be interpreted with caution until future work can test factors in more realistic settings. Taking each situational factor from above provides suggestions to increase charitable giving:

i) *Providing extrinsic rewards (strategic motivation)* – recipients of charitable donations are unlikely to 'return the favour' to the donor but the organisations themselves can provide extrinsic incentives, such as membership benefits or small gifts. However, there is a risk that extrinsic motivations 'crowd out' or replace intrinsic ones (Ottoni-Wilhelm et al., 2017), rather than complementing or adding to them. This could in fact decrease altruism.

ii) *Opportunities to personally make a difference (action-oriented motivation)* – at the appeal stage, requests for donations can be framed as what that potential donor can personally achieve, for example by donating a certain amount. Making an impact has been shown to increase positive feelings associated with altruism (Aknin et al., 2013b). Feedback on what was achieved by a donation can also be tailored to the impact of the individual donor, rather than donors in general (even if this is in principle - for example, every donor who gave the same amount receives the same message).

iii) *Physical and social closeness with recipients* – our results show increased arousal responses to events affecting people in the UK, compared to abroad. This does not necessarily mean that charities should focus on appeals for physically close recipients, particularly as a larger personal impact (see above) may be possible for causes abroad. We cannot differentiate between explanations based on social or physical proximity but based on studies covered in the literature review, both may play a role. In general, it may be beneficial for appeals to emphasise similarities and shared identities between donors and recipients (Azevedo et al., 2013; Håkansson and Montgomery, 2003; Mobbs et al., 2009). Crucially, our news stories task measured passive responses, not donations, so more work is needed to measure how these responses relate to behaviour.

iv) *A number of people that maximises emotion* – in line with previous behavioural work, we find a collapse of compassion in responses to large numbers of people in need and suffering. Charities should be aware of the possibility of this happening in response to fundraising appeals that include statistics about many people. Like the effect of proximity (see above), our task only measured passive responses and it is

unclear whether the peaks of responses to numbers in the tens or hundreds will be the same across contexts. Instead, perhaps the main implication of these results, and those on proximity, is to consider how to maximise emotional reactions without overwhelming potential donors.

v)   *Information likely to be positive can increase interest* – often potential donors have choices whether to engage with appeals before they see the appeal content. For example, opening an email or envelope from the charity. This requires a desire to want to find out more information. Our results suggest that this is increased by the belief that information is positive, particularly if it relates to a donation decision made by the individual.

In addition to these situational factors, our findings on trait measures could also have practical implications. However, there is an issue of how charities know what traits their donors have, as the information held by charities is often almost exclusively demographic. Future work is needed to translate research on individual differences into specific insights for charities. More relevant at this stage are the interactions between situational and trait factors. These demonstrate the importance of recognising differences between how individual donors will respond to appeals that manipulate the factors listed above.

I have focused this section on implications for fundraising organisations and therefore the beneficiaries they support, but increased prosocial behaviour also benefits the person being prosocial (Aknin et al., 2013a; Crocker et al., 2017; Dunn et al., 2008). The different ways in which we value the lives of others can be conceptualised as different goals of prosocial behaviour, for example making a positive difference or increasing the wellbeing of others. Acting in line with one's goals is rewarding and improves wellbeing (Zaki and Mitchell, 2016). Increased reward from being prosocial in the way that achieves our goals is likely to mean prosocial behaviour is repeated (Gęsiarz and Crockett, 2015), creating a positive feedback loop for all involved.

2. Limitations and future directions

Directions for future research to maximise the practical applications are mentioned above. There are also improvements to, and extensions of, the work for academic and theoretical reasons. Specific limitations and future directions of each study are provided in the relevant discussion section for that chapter.

2.1. Lab environment

All the studies in this thesis except one online experiment were conducted in a lab setting, some inside the fMRI scanner or involving attaching wires to the participants' hand. While these physiological and neural measures move the experience further from the real world, they also overcome some of the issues associated with social desirability in experimental settings as highlighted above.

While the studies incorporated into the meta-analysis all used an artificial economic game in a laboratory setting, the prevalence of these games is thanks to their relevance to real-world settings. In many instances outside of the lab, more can be gained or achieved through working together than any party could achieve alone. This could be through simultaneous, bidirectional generosity such as in the prisoners' dilemma or more linear interpersonal decision-making, where one party may use generosity to elicit reciprocity in the second player to make a decision.

2.2. Measuring prosocial behaviour

The practical implications of the work in this thesis described above relate to prosocial behaviours, donations in particular. However, most of the main tasks in the studies did not measure prosocial behaviour. This could be considered a criticism as more work is needed to test whether the identified biases, such as scope insensitivity and interest in positive outcomes, translate into prosocial behaviours. The current research provides the basis to take this research forward, while giving insights that may not have been possible with behavioural measures, as demonstrated by the difference between willingness to pay and physiological or neural response patterns.

2.3. Behaviours over time

Another aspect of prosocial behaviour not covered by this thesis is the role of learning processes in changing behaviour over time. Paper 3 measured changes over time in physiological responses and the impact of number on these but this was still within a single session. Learning processes are involved in prosocial decisions, as with other decision-making processes (Gęsiarz and Crockett, 2015). This involves the experienced utility of previous decisions determining prediction errors for upcoming decisions (Lin et al., 2012). For moral behaviours specifically, being prosocial initially can lead to decreases in prosocial behaviour at the next opportunity (Merritt et al., 2010). Studying learning processes benefits from the application of computation models and this would be a beneficial direction for future research.

3. Conclusion

Making sacrifices to help others is evidence that their lives and wellbeing have value to us. The studies in this thesis further our understanding of a) how this happens through biological processes, b) the possible reasons for this value c) the factors which affect how much we value a life. By understanding differences between motivations to help other people, and how these vary between individuals and contexts, the research supports ways of helping that provide the greatest rewards for both the giver and the receiver. Results show that people do value the lives of others, but this value can be biased by factors including the number of people at risk, where they are, and whether the participant is responsible for helping them.

References

Acikalin, M.Y., Gorgolewski, K.J., Poldrack, R.A., 2017. A coordinate-based meta-analysis of overlaps in regional specialization and functional connectivity across subjective value and default mode networks. Front. Neurosci. 11, 1. https://doi.org/10.3389/fnins.2017.00001

Adamus, M., 2017. Reasons for doing good: Behavioural explanations of prosociality in economics. Econ. Sociol. 10, 122–134. https://doi.org/10.14254/2071-789X.2017/10-1/9

Agnoli, S., Pittarello, A., Hysenbelli, D., Rubaltelli, E., 2015. "Give, but Give until It Hurts": The Modulatory Role of Trait Emotional Intelligence on the Motivation to Help. PLoS One 10, e0130704. https://doi.org/10.1371/journal.pone.0130704

Aguiar, F., Brañas-Garza, P., Miller, L.M., 2008. Moral distance in dictator games. Judgm. Decis. Mak. 3, 344–354.

Akaike, H., 1974. A New Look at the Statistical Model Identification, in: Selected Papers of Hirotugu Akaike. pp. 215–222. https://doi.org/10.1007/978-1-4612-1694-0_16

Aknin, L.B., Dunn, E.W., Helliwell, J.F., Biswas-Diener, R., Nyende, P., Barrington-Leigh, C.P., Burns, J., Kemeza, I., Ashton-James, C.E., Norton, M.I., 2013a. Prosocial spending and well-being: Cross-cultural evidence for a psychological universal. J. Pers. Soc. Psychol. 104, 635–652. https://doi.org/10.1037/a0031578

Aknin, L.B., Dunn, E.W., Whillans, A. V., Grant, A.M., Norton, M.I., 2013b. Making a difference matters: Impact unlocks the emotional benefits of prosocial spending. J. Econ. Behav. Organ. 88, 90–95. https://doi.org/10.1016/j.jebo.2013.01.008

Albajes-Eizagirre, A., Radua, J., 2018. What do results from coordinate-based meta-analyses tell us? Neuroimage. https://doi.org/10.1016/j.neuroimage.2018.04.065

Alcalá-López, D., Smallwood, J., Jefferies, E., Van Overwalle, F., Vogeley, K., Mars, R.B., Turetsky, B.I., Laird, A.R., Fox, P.T., Eickhoff, S.B., Bzdok, D., 2017. Computing the social brain connectome across systems and states. Cereb. Cortex 1–26. https://doi.org/10.1093/cercor/bhx121

Amodio, D.M., Frith, C.D., 2006. Meeting of minds: The medial frontal cortex and social cognition. Nat. Rev. 7, 268–277. https://doi.org/10.1038/nrn1884

Andersson, J.L.R., Jenkinson, M., Smith, S., 2010. Non-linear registration, aka spatial normalisation.

Andersson, J.L.R., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. Neuroimage 20, 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7

Andin, J., Fransson, P., Rönnberg, J., Rudner, M., 2015. Brain & Language Phonology and arithmetic in the language – calculation network. Brain Lang. 143, 97–105. https://doi.org/10.1016/j.bandl.2015.02.004

Andreoni, J., 1990. Impure altruism and donations to public goods: A theory of warm-glow giving. Econ. J. 100, 464–477.

Andreoni, J., 1989. Giving with impure altruism: Applications to charity and ricardian equivalence. J. Polit. Econ. 97, 1447. https://doi.org/10.1086/261662

Andreoni, J., Rao, J.M., 2011. The power of asking: How communication affects selfishness,

empathy, and altruism. J. Public Econ. 95, 513–520.
https://doi.org/10.1016/j.jpubeco.2010.12.008

Andreoni, J., Rao, J.M., Trachtman, H., 2017. Avoiding the Ask: A Field Experiment on
Altruism, Empathy, and Charitable Giving. J. Polit. Econ. 125, 625–653.
https://doi.org/10.1086/691703

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.
https://doi.org/10.1016/j.neuron.2016.04.018

Ariely, D., Berns, G.S., 2010. Neuromarketing: the hope and hype of neuroimaging in
business. Nat. Rev. Neurosci. 11, 284–292. https://doi.org/10.1038/nrn2795

Ashar, Y.K., Andrews-Hanna, J.R., Dimidjian, S., Wager, T.D., 2017. Empathic Care and
Distress: Predictive Brain Markers and Dissociable Brain Systems. Neuron 94, 1263-
1273.e4. https://doi.org/10.1016/j.neuron.2017.05.014

Azevedo, R.T., Beukelaer, S. De, Jones, I., Safra, L., Tsakiris, M., 2019. When the lens is too
wide : the visual dehumanization of refugees and its political consequences. PsyArXiv
1–30. https://doi.org/10.31234/OSF.IO/Y69SQ

Azevedo, R.T., Macaluso, E., Avenanti, A., Santangelo, V., Cazzato, V., Aglioti, S.M., 2013.
Their pain is not our pain: Brain and autonomic correlates of empathic resonance
with the pain of same and different race individuals. Hum. Brain Mapp. 34, 3168–
3181. https://doi.org/10.1002/hbm.22133

Bach, D.R., 2014. A head-to-head comparison of SCRalyze and Ledalab, two model-based
methods for skin conductance analysis. Biol. Psychol. 103, 63–68.
https://doi.org/10.1016/j.biopsycho.2014.08.006

Bach, D.R., Castegnetti, G., Korn, C.W., Gerster, S., Melinscak, F., Moser, T., 2018.
Psychophysiological modeling: Current state and future directions. Psychophysiology
55, e13214. https://doi.org/10.1111/psyp.13209

Bach, D.R., Daunizeau, J., Friston, K.J., Dolan, R.J., 2010. Dynamic causal modelling of
anticipatory skin conductance responses. Biol. Psychol. 85, 163–170.
https://doi.org/10.1016/j.biopsycho.2010.06.007

Bach, D.R., Friston, K.J., 2013. Model-based analysis of skin conductance responses:
Towards causal models in psychophysiology. Psychophysiology 50, 15–22.
https://doi.org/10.1111/j.1469-8986.2012.01483.x

Bach, D.R., Seymour, B., Dolan, R.J., 2009. Neural activity associated with the passive
prediction of ambiguity and risk for aversive events. J. Neurosci. 29, 1648–1656.
https://doi.org/10.1523/JNEUROSCI.4578-08.2009

Balconi, M., Canavesio, Y., 2013. Prosocial attitudes and empathic behavior in emotional
positive versus negative situations: Brain response (ERPs) and source localization
(LORETA) analysis. Cogn. Process. 14, 63–72. https://doi.org/10.1007/s10339-012-
0525-1

Balleine, B.W., Delgado, M.R., Hikosaka, O., 2007. The Role of the Dorsal Striatum in
Reward and Decision-Making. J. Neurosci. 27, 8161–8165.
https://doi.org/10.1523/JNEUROSCI.1554-07.2007

Barasch, A., Levine, E.E., Berman, J.Z., Small, D. a, 2014. Selfish or selfless? On the signal
value of emotion in altruistic behavior. J. Pers. Soc. Psychol. 107, 393–413.
https://doi.org/10.1037/a0037207

Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. J. Mem. Lang. 68, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bartels, D.M., 2006. Proportion dominance: The generality and variability of favoring relative savings over absolute savings. Organ. Behav. Hum. Decis. Process. 100, 76–95. https://doi.org/10.1016/j.obhdp.2005.10.004

Bartra, O., McGuire, J.T., Kable, J.W., 2013. The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage 76, 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Bastian, B., Costello, K., Loughnan, S., Hodson, G., 2011. When Closing the Human-Animal Divide Expands Moral Concern: The Importance of Framing. Soc. Psychol. Personal. Sci. 3, 421–429. https://doi.org/10.1177/1948550611425106

Bastin, C., Harrison, B.J., Davey, C., Moll, J., Whittle, S., 2016. Feelings of shame , embarrassment and guilt and their neural correlates : A Systematic Review. Neurosci. Biobehav. Rev. https://doi.org/10.1016/j.neubiorev.2016.09.019

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67. https://doi.org/10.18637/jss.v067.i01

Batson, C.D., 2014. The Altruism Question, The Altruism Question. https://doi.org/10.4324/9781315808048

Batson, C.D., 2009. These Things Called Empathy: Eight Related but Distinct Phenomena, in: Decety, J., Ickes, W. (Eds.), The Social Neuroscience of Empathy. MIT Press, Cambridge, MA, pp. 3–15. https://doi.org/10.1007/s11229-015-0771-8

Batson, C.D., Batson, J.G., Slingsby, J.K., Harrell, K.L., Peekna, H.M., Todd, R.M., 1991. Empathic joy and the empathy-altruism hypothesis. J. Pers. Soc. Psychol. 61, 413–426. https://doi.org/10.1037/0022-3514.61.3.413

Batson, C.D., Fultz, J., Schoenrade, P.A., 1987. Distress and Empathy: Two Qualitatively Distinct Vicarious Emotions with Different Motivational Consequences. J. Pers. 55, 19–39. https://doi.org/10.1111/j.1467-6494.1987.tb00426.x

Batson, C.D., Klein, T.R., Highberger, L., Shaw, L.L., 1995. Immorality from empathy-induced altruism: When compassion and justice conflict. J. Pers. Soc. Psychol. 68, 1042–1054. https://doi.org/10.1037/0022-3514.68.6.1042

Beckmann, C., Smith, S., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans. Med. Imaging.

Bekkers, R., Wiepking, P., 2010. A Literature Review of Empirical Studies of Philanthropy: Eight Mechanisms That Drive Charitable Giving, Nonprofit and Voluntary Sector Quarterly. https://doi.org/10.1177/0899764010380927

Bekkers, R., Wiepking, P., 2007. Generosity and philantropy: a literature review. Sci. Generosity 1–68. https://doi.org/10.2139/ssrn.1015507

Bellucci, G., Chernyak, S. V., Goodyear, K., Eickhoff, S.B., Krueger, F., 2017. Neural signatures of trust in reciprocity: A coordinate-based meta-analysis. Hum. Brain Mapp. 38, 1233–1248. https://doi.org/10.1002/hbm.23451

Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games Econ. Behav. 10, 122–142. https://doi.org/10.1006/game.1995.1027

Bernhardt, B.C., Singer, T., 2012. The Neural Basis of Empathy. Annu. Rev. Neurosci. 35, 1–

23. https://doi.org/10.1146/annurev-neuro-062111-150536

Bickart, K.C., Dickerson, B.C., Feldman Barrett, L., 2014. The amygdala as a hub in brain networks that support social life. Neuropsychologia 63, 235–248. https://doi.org/10.1016/j.neuropsychologia.2014.08.013

Bloom, P., 2017. Empathy and Its Discontents. Trends Cogn. Sci. 21, 24–31. https://doi.org/10.1016/j.tics.2016.11.004

Bloom, P., 2014. Against empathy [WWW Document]. URL http://bostonreview.net/forum/paul-bloom-against-empathy (accessed 1.6.16).

Boenigk, S., Mayr, M.L., 2015. The Happiness of Giving: Evidence from the German Socioeconomic Panel That Happier People Are More Generous. J. Happiness Stud. https://doi.org/10.1007/s10902-015-9672-2

Böhm, R., Theelen, M.M.P., Rusch, H., Van Lange, P.A.M., 2018. Costs, needs, and integration efforts shape helping behavior toward refugees. Proc. Natl. Acad. Sci. 115, 7284–7289. https://doi.org/10.1073/pnas.1805601115

Bolton, E., Ockenfels, A., 1995. ERC : A Theory of Equity , Reciprocity , and Competition. Am. Econ. Rev. 90, 166–193. https://doi.org/10.2307/117286

Bornemann, B., Kok, B.E., Böckler, A., Singer, T., 2016. Helping from the heart: Voluntary upregulation of heart rate variability predicts altruistic behavior. Biol. Psychol. 119, 54–63. https://doi.org/10.1016/j.biopsycho.2016.07.004

Bowman, W., 2006. Should donors care about overhead costs? Do they care? Nonprofit Volunt. Sect. Q. https://doi.org/10.1177/0899764006287219

Brainard, D.H., 1997. The Psychophysics Toolbox. Spat. Vis. 433–436.

Brethel-Haurwitz, K.M., Cardinale, E.M., Vekaria, K.M., Robertson, E.L., Walitt, B., VanMeter, J.W., Marsh, A.A., 2018. Extraordinary Altruists Exhibit Enhanced Self–Other Overlap in Neural Responses to Distress. Psychol. Sci. 29, 1631–1641. https://doi.org/10.1177/0956797618779590

Buchanan, T.W., Preston, S.D., 2014. Stress leads to prosocial action in immediate need situations. Front. Behav. Neurosci. 8, 1–6. https://doi.org/10.3389/fnbeh.2014.00005

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

Buffone, A.E.K., Poulin, M., DeLury, S., Ministero, L., Morrisson, C., Scalco, M., 2017. Don't walk in her shoes! Different forms of perspective taking affect stress physiology. J. Exp. Soc. Psychol. 72, 161–168. https://doi.org/10.1016/j.jesp.2017.04.001

Burton, A.C., Nakamura, K., Roesch, M.R., 2015. From ventral-medial to dorsal-lateral striatum: Neural correlates of reward-guided decision-making. Neurobiol. Learn. Mem. https://doi.org/10.1016/j.nlm.2014.05.003

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: Why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365–76. https://doi.org/10.1038/nrn3475

Camerer, C.F., 2003. Behavioural studies of strategic thinking in games. Trends Cogn. Sci. 7, 225–231. https://doi.org/10.1016/S1364-6613(03)00094-9

Camerer, C.F., Fehr, E., 2003. Measuring Social Norms and Preferences Using Experimental

Games. Found. Hum. Soc. Econ. Exp. Ethnogr. Evid. from fifteen Small-scale Soc. 1–40.

Cameron, C.D., Fredrickson, B.L., 2015. Mindfulness Facets Predict Helping Behavior and Distinct Helping-Related Emotions. Mindfulness (N. Y). 6, 1211–1218. https://doi.org/10.1007/s12671-014-0383-2

Cameron, C.D., Hutcherson, C.A., Ferguson, A.M., Scheffer, J.A., Hadjiandreou, E., Inzlicht, M., 2019. Empathy is hard work: People choose to avoid empathy because of its cognitive costs. J. Exp. Psychol. Gen. https://doi.org/10.13140/RG.2.2.25908.04486

Cameron, C.D., Payne, B.K., 2011. Escaping affect: How motivated emotion regulation creates insensitivity to mass suffering. J. Pers. Soc. Psychol. 100, 1–15. https://doi.org/10.1037/a0021643

Campbell-Meiklejohn, D., Simonsen, A., Frith, C.D., Daw, N.D., 2016. Independent neural computation of value from the confidence of others. J. Neurosci. 37, 4490–15. https://doi.org/10.1523/JNEUROSCI.4490-15.2016

Capraro, V., 2017. Do the right thing: Preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality. SSRN Electron. J. 13, 99–111. https://doi.org/10.2139/ssrn.2965067

Capraro, V., Rand, D.G., 2017. Do the Right Thing: Experimental Evidence that Preferences for Moral Behavior, Rather Than Equity or Efficiency per se, Drive Human Prosociality. Ssrn 13, 1–24. https://doi.org/10.2139/ssrn.2965067

Capraro, V., Vanzo, A., 2019. The power of moral words: Loaded language generates framing effects in the extreme dictator game 1–16.

Carlson, R.W., Crockett, M.J., 2018. The lateral prefrontal cortex and moral goal pursuit. Curr. Opin. Psychol. 24, 77–82. https://doi.org/10.1016/j.copsyc.2018.09.007

Carlson, R.W., Zaki, J., n.d. Belief in altruistic motives predicts prosocial actions and inferences. PsyArXiv 2019, 139–144.

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

Caviola, L., Faulmüllert, N., Everett, J.A.C.C., Savulescu, J., Kahane, G., Faulmüller, N., Everett, J.A.C.C., Savulescu, J., 2014. The evaluability bias in charitable giving: Saving administration costs or saving lives? Judgm. Decis. Mak. 9, 303–315. https://doi.org/None

Chang, C.T., Lee, Y.K., 2010. Effects of message framing, vividness congruency and statistical framing on responses to charity advertising. Int. J. Advert. 29, 195–220. https://doi.org/10.2501/S0265048710201129

Chang, L.J., Sanfey, A.G., 2013. Great expectations: Neural computations underlying the use of social norms in decision-making. Soc. Cogn. Affect. Neurosci. 8, 277–284. https://doi.org/10.1093/scan/nsr094

Charities Aid Foundation, 2019. UK Giving 2019.

Charpentier, C.J., Bromberg-Martin, E.S., Sharot, T., 2018. Valuation of knowledge and ignorance in mesolimbic reward circuitry. Proc. Natl. Acad. Sci. 115, E7255–E7264. https://doi.org/10.1073/pnas.1800547115

Chaudhuri, A., 2011. Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. Exp. Econ. 14, 47–83. https://doi.org/10.1007/s10683-010-9257-1

Chen, S., Friesdorf, R., Jordan, C.H., 2019. State and trait narcissism predict everyday helping. Self Identity 00, 1–17. https://doi.org/10.1080/15298868.2019.1598892

Chernyak, N., Kushnir, T., 2018. The influence of understanding and having choice on children's prosocial behavior. Curr. Opin. Psychol. 20, 107–110. https://doi.org/10.1016/j.copsyc.2017.07.043

Chiu Loke, I., Evans, A.D., Lee, K., 2011. The neural correlates of reasoning about prosocial-helping decisions: An event-related brain potentials study. Brain Res. 1369, 140–148. https://doi.org/10.1016/j.brainres.2010.10.109

Christopoulos, G.I., King-Casas, B., 2015. With you or against you: Social orientation dependent learning signals guide actions made for others. Neuroimage 33, 395–401. https://doi.org/10.1038/nbt.3121.ChIP-nexus

Christov-Moore, L., Iacoboni, M., 2016. Self-other resonance, its control and prosocial inclinations: Brain-behavior relationships. Hum. Brain Mapp. 37, 1544–1558. https://doi.org/10.1002/hbm.23119

Christov-Moore, L., Sugiyama, T., Grigaityte, K., Iacoboni, M., 2016. Increasing generosity by disrupting prefrontal cortex. Soc. Neurosci. 9, 1–8. https://doi.org/10.1080/17470919.2016.1154105

Citron, F.M.M., Gray, M.A., Critchley, H.D., Weekes, B.S., Ferstl, E.C., 2014. Emotional valence and arousal affect reading in an interactive way: Neuroimaging evidence for an approach-withdrawal framework. Neuropsychologia 56, 79–89. https://doi.org/10.1016/j.neuropsychologia.2014.01.002

Clavien, C., Chapuisat, M., 2012. Altruism across disciplines: one word, multiple meanings. Biol. Philos. 28, 125–140. https://doi.org/10.1007/s10539-012-9317-3

Clithero, J.A., Rangel, A., 2013. Informatic parcellation of the network involved in the computation of subjective value. Soc. Cogn. Affect. Neurosci. 9, 1289–1302. https://doi.org/10.1093/scan/nst106

Cornelissen, G., Dewitte, S., Warlop, L., 2011. Are social value orientations expressed automatically? Decision making in the dictator game. Personal. Soc. Psychol. Bull. 37, 1080–1090. https://doi.org/10.1177/0146167211405996

Cowell, J.M., Calma-Birling, D., Decety, J., 2018. Domain-general neural computations underlying prosociality during infancy and early childhood. Curr. Opin. Psychol. 20, 66–71. https://doi.org/10.1016/j.copsyc.2017.08.016

Cremers, H.R., Wager, T.D., Yarkoni, T., 2017. The relation between statistical power and inference in fMRI. PLoS One 12, e0184923. https://doi.org/10.1371/journal.pone.0184923

Critchley, H.D., Harrison, N.A., 2013. Visceral Influences on Brain and Behavior. Neuron 77, 624–638. https://doi.org/10.1016/j.neuron.2013.02.008

Critchley, H.D., Mathias, C.J., Dolan, R.J., 2001. Neural activity in the human brain relating to uncertainty and arousal during anticipation. Neuron 29. https://doi.org/10.1016/S0896-6273(01)00225-2

Crocker, J., Canevello, A., 2008. Creating and Undermining Social Support in Communal Relationships: The Role of Compassionate and Self-Image Goals. J. Pers. Soc. Psychol. 95, 555–575. https://doi.org/10.1037/0022-3514.95.3.555

Crocker, J., Canevello, A., Brown, A.A., 2017. Social Motivation: Costs and Benefits of

Selfishness and Otherishness. Annu. Rev. Psychol. 68, annurev-psych-010416-044145. https://doi.org/10.1146/annurev-psych-010416-044145

Croxson, P.L., Walton, M.E., O'Reilly, J.X., Behrens, T.E.J., Rushworth, M.F.S., 2009. Effort-based cost-benefit valuation and the human brain. J. Neurosci. 29, 4531–4541. https://doi.org/10.1523/JNEUROSCI.4515-08.2009

Curry, O.S., Rowland, L.A., Van Lissa, C.J., Zlotowitz, S., McAlaney, J., Whitehouse, H., 2018. Happy to help? A systematic review and meta-analysis of the effects of performing acts of kindness on the well-being of the actor. J. Exp. Soc. Psychol. 76, 320–329. https://doi.org/10.1016/j.jesp.2018.02.014

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

Damoiseaux, J.S., Rombouts, S.A.R.B., Barkhof, F., Scheltens, P., Stam, C.J., Smith, S.M., Beckmann, C.F., 2006. Consistent resting-state networks across healthy subjects. Proc. Natl. Acad. Sci. U. S. A. 103, 13848–13853. https://doi.org/10.1073/pnas.0601417103

Dana, J., Cain, D.M., Dawes, R.M., 2006. What you don't know won't hurt me: Costly (but quiet) exit in dictator games. Organ. Behav. Hum. Decis. Process. 100, 193–201. https://doi.org/10.1016/j.obhdp.2005.10.001

Das, E., Kerkhof, P., Kuiper, J., 2008. Improving the Effectiveness of Fundraising Messages: The Impact of Charity Goal Attainment, Message Framing, and Evidence on Persuasion. J. Appl. Commun. Res. 36, 161–175. https://doi.org/10.1080/00909880801922854

Davis, M.H., 1983. Measuring individual differences in empathy: Evidence for a multidimensional approach. J. Pers. Soc. Psychol. 44, 113–126. https://doi.org/10.1037/0022-3514.44.1.113

Dawes, C.T., Loewen, P.J., Schreiber, D., Simmons, A.N., Flagan, T., McElreath, R., Bokemper, S.E., Fowler, J.H., Paulus, M.P., 2012. Neural basis of egalitarian behavior. Proc. Natl. Acad. Sci. U. S. A. 109, 6479–6483. https://doi.org/10.1073/pnas.1118653109

De Martino, B., Bobadilla-Suarez, S., Nouguchi, T., Sharot, T., Love, B.C., 2017. Social information is integrated into value and confidence judgments according to its reliability. J. Neurosci. 37, 6066–6074. https://doi.org/10.1523/JNEUROSCI.3880-16.2017

De Martino, B., Camerer, C.F., Adolphs, R., 2010. Amygdala damage eliminates monetary loss aversion. Proc. Natl. Acad. Sci. 107, 3788–3792. https://doi.org/10.1073/pnas.0910230107

de Waal, F.B.M., 2008. Putting the altruism back into altruism: the evolution of empathy. Annu. Rev. Psychol. 59, 279–300. https://doi.org/10.1146/annurev.psych.59.103006.093625

de Waal, F.B.M.F., Leimgruber, K., Greenberg, A.R., 2008. Giving is self-rewarding for monkeys. Proc. Natl. Acad. Sci. U. S. A. 105, 13685–13689. https://doi.org/10.1073/pnas.0807060105

Decety, J., 2015. The neural pathways, development and functions of empathy. Curr. Opin. Behav. Sci. 3, 1–6. https://doi.org/10.1016/j.cobeha.2014.12.001

Decety, J., 2011. Dissecting the Neural Mechanisms Mediating Empathy. Emot. Rev. 3, 92–

108. https://doi.org/10.1177/1754073910374662

Decety, J., Bartal, I.B., Uzefovsky, F., Knafo-noam, A., 2015. Empathy as a driver of prosocial behaviour: Highly conserved neurobehavioural mechanisms across species. Philos. Trans. R. Soc. B, Biol. Sci. 371, 20150077. https://doi.org/http://dx.doi.org/10.1098/rstb.2015.0077

Decety, J., Cowell, J.M., 2015. Empathy, justice, and moral behavior. AJOB Neurosci. 6, 3–14. https://doi.org/10.1080/21507740.2015.1047055

Decety, J., Lamm, C., 2007. The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition. Neurosci. 13, 580–93. https://doi.org/10.1177/1073858407304654

Decety, J., Yoder, K.J., 2017. The Emerging Social Neuroscience of Justice Motivation. Trends Cogn. Sci. 21, 6–14. https://doi.org/10.1016/j.tics.2016.10.008

Dersimonian, R., Laird, N., 1986. Meta-Analysis in Clinical Trials. Stat. Med. 188, 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

Dickert, S., Kleber, J., Peters, E., Slovic, P., 2011a. Numeracy as a precursor to pro-social behavior: The impact of numeracy and presentation format on the cognitive mechanisms underlying donation decisions. Judgm. Decis. Mak. 6, 638–650.

Dickert, S., Sagara, N., Slovic, P., 2011b. Affective Motivations to Help Others: A Two-stage Model of Donation Decisions. J. Neurosci. 16, 6513–6525. https://doi.org/10.1002/bdm

Dickert, S., Västfjäll, D., Kleber, J., Slovic, P., 2015. Scope insensitivity: The limits of intuitive valuation of human lives in public policy. J. Appl. Res. Mem. Cogn. 4, 248–255. https://doi.org/10.1016/j.jarmac.2014.09.002

Dickert, S., Västfjäll, D., Kleber, J., Slovic, P., 2012. Valuations of human lives: normative expectations and psychological mechanisms of (ir)rationality. Synthese 1–11. https://doi.org/10.1007/s11229-012-0137-4

Domsch, S., Linke, J., Heiler, P.M., Kroll, A., Flor, H., Wessa, M., Schad, L.R., 2013. Increased BOLD sensitivity in the orbitofrontal cortex using slice-dependent echo times at 3 T. Magn. Reson. Imaging 31, 201–211. https://doi.org/10.1016/j.mri.2012.06.020

Drevets, W.C., Savitz, J., Trimble, M., 2008. The subgenual anterior cingulate cortex in mood disorders. CNS Spectr. 13, 663–681. https://doi.org/10.1017/S1092852900013754

Driscoll, D.M., Dal Monte, O., Solomon, J., Krueger, F., Grafman, J., 2012. Empathic Deficits in Combat Veterans with Traumatic Brain Injury: A Voxel-Based Lesion-Symptom Mapping Study. Cogn. Behav. Neurol. 25.

Drouvelis, M., Grosskopf, B., 2014. The Effects of Anger and Happiness on Pro-Social Behaviour 1–37.

Dunn, E.W., Aknin, L.B., Norton, M.I., 2014. Prosocial Spending and Happiness: Using Money to Benefit Others Pays Off. Curr. Dir. Psychol. Sci. 23, 41–47. https://doi.org/10.1177/0963721413512503

Dunn, E.W., Aknin, L.B., Norton, M.I., 2008. Spending money on others promotes happiness. Science (80-.). 319, 1687–1688. https://doi.org/10.1126/science.1150952

Einolf, C.J., 2008. Empathic concern and prosocial behaviors: A test of experimental results using survey data. Soc. Sci. Res. 37, 1267–1279. https://doi.org/10.1016/j.ssresearch.2007.06.003

Eisenberg, N., Fabes, R. a, Murphy, B., Karbon, M., Maszk, P., Smith, M., O'Boyle, C., Suh, K., 1994. The relations of emotionality and regulation to dispositional and situational empathy-related responding. J. Pers. Soc. Psychol. 66, 776–797. https://doi.org/10.1037/0022-3514.66.4.776

Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci. U. S. A. 113, 7900–7905. https://doi.org/10.1073/pnas.1602413113

Emonds, G., Declerck, C.H., Boone, C., Seurinck, R., Achten, R., 2014. Establishing cooperation in a mixed-motive social dilemma. An fMRI study investigating the role of social value orientation and dispositional trust. Soc. Neurosci. 9, 10–22. https://doi.org/10.1080/17470919.2013.858080

Emonds, G., Declerck, C.H., Boone, C., Vandervliet, E.J.M., Parizel, P.M., 2011. Comparing the neural basis of decision making in social dilemmas of people with different social value orientations, an fMRI study. J. Neurosci. Psychol. Econ. 4, 11–24. https://doi.org/10.1037/a0020151

Engel, C., 2011. Dictator games: A meta study. Exp. Econ. 14, 583–610. https://doi.org/10.1007/s10683-011-9283-7

Engel, C., Goerg, S.J., 2018. If the worst comes to the worst: Dictator giving when recipient's endowments are risky. Eur. Econ. Rev. 105, 51–70. https://doi.org/10.1016/j.euroecorev.2018.03.011

Engen, H.G., Singer, T., 2013. Empathy circuits. Curr. Opin. Neurobiol. 23, 275–282. https://doi.org/10.1016/j.conb.2012.11.003

Erlandsson, A., Björklund, F., Bäckström, M., 2017. Choice-justifications after allocating resources in helping dilemmas. Judgm. Decis. Mak. 12, 60–80. https://doi.org/10.1023/a:1007796310463

Erlandsson, A., Björklund, F., Bäckström, M., 2015. Emotional reactions, perceived impact and perceived responsibility mediate the identifiable victim effect, proportion dominance effect and in-group effect respectively. Organ. Behav. Hum. Decis. Process. 127, 1–14. https://doi.org/10.1016/j.obhdp.2014.11.003

Erlandsson, A., Björklund, F., Bäckström, M., 2014. Perceived Utility (not Sympathy) Mediates the Proportion Dominance Effect in Helping Decisions. J. Behav. Decis. Mak. 27, 37–47. https://doi.org/10.1002/bdm.1789

Erlandsson, A., Jungstrand, A., Västfjäll, D., 2016. Anticipated guilt for not helping and anticipated warm glow for helping are differently impacted by personal responsibility to help. Front. Psychol. 7, 1–19. https://doi.org/10.3389/fpsyg.2016.01475

Erlandsson, A., Nilsson, A., Västfjäll, D., 2018. Attitudes and Donation Behavior When Reading Positive and Negative Charity Appeals. J. Nonprofit Public Sect. Mark. 30, 444–474. https://doi.org/10.1080/10495142.2018.1452828

Everett, J. a. C., Faber, N.S., Crockett, M.J., 2015. Preferences and beliefs in ingroup favoritism. Front. Behav. Neurosci. 9, 1–21. https://doi.org/10.3389/fnbeh.2015.00015

Everitt, B.J., Robbins, T.W., 2016. Drug addiction: Updating actions to habits to compulsions ten years on. Annu. Rev. Psychol. 67, 23–50. https://doi.org/10.1146/annurev-psych-122414-033457

Falk, A., Fischbacher, U., 2006. A theory of reciprocity. Games Econ. Behav. 54, 293–315. https://doi.org/10.1016/j.geb.2005.03.001

Fehr, E., Fischbacher, U., 2004. Social norms and human cooperation. Trends Cogn. Sci. 8, 185–190. https://doi.org/10.1016/j.tics.2004.02.007

Fehr, E., Fischbacher, U., Gächter, S., 2002. Strong reciprocity, human cooperation, and the enforcement of social norms. Hum. Nat. 13, 1–25. https://doi.org/10.1007/s12110-002-1012-7

Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. Nature 415, 137–140. https://doi.org/10.1038/415137a

Fehr, T., Code, C., Herrmann, M., 2007. Common brain regions underlying different arithmetic operations as revealed by conjunct fMRI – BOLD activation. Brain Res. 2, 1–10. https://doi.org/10.1016/j.brainres.2007.07.043

FeldmanHall, O., Dalgleish, T., Evans, D., Mobbs, D., 2015. Empathic concern drives costly altruism. Neuroimage 105, 347–356. https://doi.org/10.1016/j.neuroimage.2014.10.043

FeldmanHall, O., Dalgleish, T., Mobbs, D., 2013. Alexithymia decreases altruism in real social decisions. Cortex 49, 899–904. https://doi.org/10.1016/j.cortex.2012.10.015

FeldmanHall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S., Mobbs, D., 2012a. Differential neural circuitry and self-interest in real vs hypothetical moral decisions. Soc. Cogn. Affect. Neurosci. 7, 743–751. https://doi.org/10.1093/scan/nss069

Feldmanhall, O., Glimcher, P.W., Baker, A.L., Phelps, E.A., 2016. Emotion and Decision-Making Under Uncertainty: Physiological arousal predicts increased gambling during ambiguity but not risk. J. Exp. Psychol. Gen. 124, 1255–1262. https://doi.org/10.1037/xge0000205

FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., Dalgleish, T., 2012b. What we say and what we do: The relationship between real and hypothetical moral choices. Cognition 123, 434–441. https://doi.org/10.1016/j.cognition.2012.02.001

Feng, C., Luo, Y.J., Krueger, F., 2015. Neural signatures of fairness-related normative decision making in the ultimatum game: A coordinate-based meta-analysis. Hum. Brain Mapp. 36, 591–602. https://doi.org/10.1002/hbm.22649

Fermin, A.S.R., Sakagami, M., Kiyonari, T., Li, Y., Matsumoto, Y., Yamagishi, T., 2016. Representation of economic preferences in the structure and function of the amygdala and prefrontal cortex. Sci. Rep. 6, 20982. https://doi.org/10.1038/srep20982

Fernandes, M.F., Randall, D.M., 1992. The Nature of Social Desirability Response Effects in Ethics Research. Bus. Ethics Q. 2, 183–205.

Fernandez, B., Leuchs, L., Sämann, P.G., Czisch, M., Spoormaker, V.I., 2017. Multi-echo EPI of human fear conditioning reveals improved BOLD detection in ventromedial prefrontal cortex. Neuroimage 156, 65–77. https://doi.org/10.1016/j.neuroimage.2017.05.005

Fetherstonhaugh, D., Slovic, P., Johnson, S.M., Friedrich, J., 1997. Insensitivity to the Value of Human Life: A Study of Psychophysical Numbing. J. Risk Uncertain. 14, 283–300. https://doi.org/10.1023/A:1007744326393

Fett, A.K.J., Gromann, P.M., Giampietro, V., Shergill, S.S., Krabbendam, L., 2014. Default

distrust? An fMRI investigation of the neural development of trust and cooperation. Soc. Cogn. Affect. Neurosci. 9, 395–402. https://doi.org/10.1093/scan/nss144

Filkowski, M., Cochran, R.N., Haas, B., 2016. Altruistic behavior: mapping responses in the brain. Neurosci. Neuroeconomics 5, 65–75. https://doi.org/10.2147/NAN.S87718

Fleming, S.M., Lau, H.C., 2014. How to measure metacognition. Front. Hum. Neurosci. https://doi.org/10.3389/fnhum.2014.00443

Forgiarini, M., Gallucci, M., Maravita, A., 2011. Racism and the empathy for pain on our skin. Front. Psychol. 2, 1–7. https://doi.org/10.3389/fpsyg.2011.00108

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc. Natl. Acad. Sci. U. S. A. 102, 9673–9678. https://doi.org/10.1073/pnas.0504136102

Frederick, S., Fischhoff, B., 1998. Scope (in)sensitivity in elicited valuations. Risk Decis. Policy 3, 109–123. https://doi.org/10.1080/135753098348239

Frederick, S., Loewenstein, G., O'donoghue, T., 2002. Time Discounting and Time Preference: A Critical Review. J. Econ. Lit. 40, 351–401. https://doi.org/10.1257/002205102320161311

Frey, B.S., Oberholzer-Gee, F., 1997. The cost of price incentives: An empirical analysis of motivation crowding-out. Am. Econ. Rev. 87, 746–755. https://doi.org/10.1207/2951373

Friedrich, J., Barnes, P., Chapin, K., Dawson, I., Garst, V., Kerr, D., 1999. Psychophysical Numbing: When Lives Are Valued Less as the Lives at Risk Increase. J. Consum. Psychol. 3, 277–299. https://doi.org/10.1207/s15327663jcp0803_05

Friedrich, J., McGuire, A., 2010. Individual differences in reasoning style as a moderator of the identifiable victim effect. Soc. Influ. 5, 182–201. https://doi.org/10.1080/15534511003707352

Frith, C.D., Frith, U., 2006. The neural basis of mentalizing. Neuron 50. https://doi.org/10.1016/j.neuron.2006.05.001

Frith, U., Frith, C.D., 2003. Development and neurophysiology of mentalizing. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 358, 459–73. https://doi.org/10.1098/rstb.2002.1218

Gabay, A.S., Radua, J., Kempton, M.J., Mehta, M.A., 2014. The Ultimatum Game and the brain: A meta-analysis of neuroimaging studies. Neurosci. Biobehav. Rev. 47, 549–558. https://doi.org/10.1016/j.neubiorev.2014.10.014

Gaesser, B., Horn, M., Young, L., 2015. When can imagining the self increase willingness to help others? Investigating whether the self-referential nature of episodic simulation fosters prosociality. Soc. Cogn. 33, 562–584. https://doi.org/10.1521/soco.2015.33.6.562

Gallardo, G., Fick, R., Wells, W., Deriche, R., Wassermann, D., 2017. Groupwise structural parcellation of the cortex: A sound approach based on logistic models, in: Mathematics and Visualization. pp. 99–112. https://doi.org/10.1007/978-3-319-54130-3_8

Gan, T., Wang, L., Zhang, L., Fang, W., Gu, R., 2016. The neural correlates of evaluating the outcome of prosocial-helping decisions. Neuroreport 1293–1298. https://doi.org/10.1097/WNR.0000000000000696

Garbarini, F., Boero, R., D'Agata, F., Bravo, G., Mosso, C., Cauda, F., Duca, S., Geminiani, G., Sacco, K., 2014. Neural correlates of gender differences in reputation building. PLoS One 9. https://doi.org/10.1371/journal.pone.0106285

Genevsky, A., Västfjäll, D., Slovic, P., Knutson, B., 2013. Neural underpinnings of the identifiable victim effect: Affect shifts preferences for giving. J. Neurosci. 33, 17188–17196. https://doi.org/10.1523/JNEUROSCI.2348-13.2013

Gershman, S.J., Markman, A.B., Otto, A.R., 2014. Retrospective revaluation in sequential decision making: A tale of two systems. J. Exp. Psychol. Gen. 143, 182–194. https://doi.org/10.1037/a0030844

Gęsiarz, F., Crockett, M.J., 2015. Goal-directed, habitual and Pavlovian prosocial behavior. Front. Behav. Neurosci. 9, 1–18. https://doi.org/10.3389/fnbeh.2015.00135

Gianotti, L.R.R., Dahinden, F.M., Baumgartner, T., Knoch, D., 2019. Understanding Individual Differences in Domain-General Prosociality: A Resting EEG Study. Brain Topogr. 32, 118–126. https://doi.org/10.1007/s10548-018-0679-y

Gneezy, U., Keenan, E.A., Gneezy, A., 2014. Avoiding overhead aversion in charity. Science (80-.). 346, 632–635. https://doi.org/10.1126/science.1253932

Gorgolewski, K.J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S.S., Maumet, C., Sochat, V. V., Nichols, T.E., Poldrack, R.A., Poline, J.-B., Yarkoni, T., Margulies, D.S., 2015. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. Front. Neuroinform. 9, 1–9. https://doi.org/10.3389/fninf.2015.00008

Gospic, K., Sundberg, M., Maeder, J., Fransson, P., Petrovic, P., Isacsson, G., Karlstrom, A., Ingvar, M., Karlström, A., Ingvar, M., 2013. Altruism costs-the cheap signal from amygdala. Soc. Cogn. Affect. Neurosci. 9, 1325–1332. https://doi.org/10.1093/scan/nst118

Grant, A.M., 2008. Does intrinsic motivation fuel the prosocial fire? Motivational synergy in predicting persistence, performance, and productivity. J. Appl. Psychol. 93, 48–58. https://doi.org/10.1037/0021-9010.93.1.48

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C.F., Auerbach, E.J., Douaud, G., Sexton, C.E., Zsoldos, E., Ebmeier, K.P., Filippini, N., Mackay, C.E., Moeller, S., Xu, J., Yacoub, E., Baselli, G., Ugurbil, K., Miller, K.L., Smith, S.M., 2014. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. Neuroimage 95, 232–247. https://doi.org/10.1016/j.neuroimage.2014.03.034

Gross, J., Emmerling, F., Vostroknutov, A., Sack, A.T., 2018. Manipulation of pro-sociality and rule-following with non-invasive brain stimulation. Sci. Rep. 8, 1827. https://doi.org/10.1038/s41598-018-19997-5

Grossman, Z., 2015. Self-signaling and social-signaling in giving. J. Econ. Behav. Organ. 117, 26–39. https://doi.org/10.1016/j.jebo.2015.05.008

Grynberg, D., Pollatos, O., 2015. Perceiving one's body shapes empathy. Physiol. Behav. 140, 54–60. https://doi.org/10.1016/j.physbeh.2014.12.026

Guinote, A., Cotzia, I., Sandhu, S., Siwa, P., 2015. Social status modulates prosocial behavior and egalitarianism in preschool children and adults. Proc. Natl. Acad. Sci. U. S. A. 112, 731–736. https://doi.org/10.1073/pnas.1414550112

Güth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. J. Econ. Behav. Organ. 3, 367–388. https://doi.org/10.1016/0167-

2681(82)90011-7

Habashi, M.M., Graziano, W.G., Hoover, A.E., 2016. Searching for the Prosocial Personality: A Big Five Approach to Linking Personality and Prosocial Behavior. Personal. Soc. Psychol. Bull. 42, 1177–1192. https://doi.org/10.1177/0146167216652859

Hackel, L.M., Zaki, J., Van Bavel, J.J., 2017. Social identity shapes social valuation: Evidence from prosocial behavior and vicarious reward. Soc. Cogn. Affect. Neurosci. 12, 1219–1228. https://doi.org/10.1093/scan/nsx045

Håkansson, J., Montgomery, H., 2003. Empathy as an Interpersonal Phenomenon. J. Soc. Pers. Relat. 20, 267–284. https://doi.org/10.1177/0265407503020003001

Haller, J., 2018. The role of central and medial amygdala in normal and abnormal aggression: A review of classical approaches. Neurosci. Biobehav. Rev. https://doi.org/10.1016/j.neubiorev.2017.09.017

Hallsson, B.G., Siebner, H.R., Hulme, O.J., 2018. Fairness, fast and slow: A review of dual process models of fairness. Neurosci. Biobehav. Rev. 89, 49–60. https://doi.org/10.1016/j.neubiorev.2018.02.016

Hansen, P., Kergozou, N., Knowles, S., Thorsnes, P., 2014. Developing Countries in Need: Which Characteristics Appeal most to People when Donating Money? J. Dev. Stud. 0388, 1–16. https://doi.org/10.1080/00220388.2014.925542

Harbaugh, W.T., Mayr, U., Burghart, D.R., 2007. Neural responses to taxation and voluntary giving reveal motives for charitable donations. Science (80-.). 316, 1622–1625. https://doi.org/10.1126/science.1140738

Hare, T.A., Camerer, C.F., Knoepfle, D.T., O'Doherty, J.P., Rangel, A., 2010. Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. J. Neurosci. 30, 583–590. https://doi.org/10.1523/JNEUROSCI.4089-09.2010

Haruno, M., Kawato, M., 2006. Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. J. Neurophysiol. 95, 948–959. https://doi.org/10.1152/jn.00382.2005

Haushofer, J., Shapiro, J., 2016. The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya. Q. J. Econ. 1973–2042. https://doi.org/10.1093/qje/qjw025.Advance

Hein, G., Lamm, C., Brodbeck, C., Singer, T., 2011. Skin conductance response to the pain of others predicts later costly helping. PLoS One 6, 1–6. https://doi.org/10.1371/journal.pone.0022759

Hein, G., Morishima, Y., Leiberg, S., Sul, S., Fehr, E., 2016. The brains functional network architecture reveals human motives. Science (80-.). 351, 1074–1078. https://doi.org/10.1126/science.aac7992

Hein, G., Silani, G., Preuschoff, K., Batson, C.D., Singer, T., 2010. Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. Neuron 68, 149–160. https://doi.org/10.1016/j.neuron.2010.09.003

Hill, P.F., Yi, R., Spreng, R.N., Diana, R.A., 2017. Neural congruence between intertemporal and interpersonal self-control: Evidence from delay and social discounting. Neuroimage 162, 186–198. https://doi.org/10.1016/j.neuroimage.2017.08.071

Holland, P.C., Gallagher, M., 2004. Amygdala-frontal interactions and reward expectancy. Curr. Opin. Neurobiol. https://doi.org/10.1016/j.conb.2004.03.007

Hortensius, R., de Gelder, B., 2018. From Empathy to Apathy: The Bystander Effect Revisited. Curr. Dir. Psychol. Sci. 27, 249–256. https://doi.org/10.1177/0963721417749653

Hu, J., Li, Y., Yin, Y., Blue, P.R., Yu, H., Zhou, X., 2017. How do self-interest and other-need interact in the brain to determine altruistic behavior? Neuroimage 157, 598–611. https://doi.org/10.1016/j.neuroimage.2017.06.040

Hu, X., Xu, Z., Mai, X., 2017. Social value orientation modulates the processing of outcome evaluation involving others. Soc. Cogn. Affect. Neurosci. 12, 1730–1739. https://doi.org/10.1093/scan/nsx102

Hu, Y., Hu, Yinying, Li, X., Pan, Y., Cheng, X., 2017. Brain-to-brain synchronization across two persons predicts mutual prosociality. Soc. Cogn. Affect. Neurosci. 12, 1835–1844. https://doi.org/10.1093/scan/nsx118

Hubbard, J., Harbaugh, W.T., Srivastava, S., Degras, D., Mayr, U., 2016. A general benevolence dimension that links neural, psychological, economic, and life-span data on altruistic tendencies. J. Exp. Psychol. Gen. 145, 1351–1358. https://doi.org/10.1037/xge0000209

Hutcherson, C.A., Bushong, B., Rangel, A., 2015. A neurocomputational model of altruistic choice and its implications. Neuron 87, 451–462. https://doi.org/10.1016/j.neuron.2015.06.031

Izuma, K., Saito, D.N., Sadato, N., 2010. Processing of the incentive for social approval in the ventral striatum during charitable donation. J. Cogn. Neurosci. 22, 621–631. https://doi.org/10.1162/jocn.2009.21228

Izuma, K., Saito, D.N., Sadato, N., 2008. Processing of social and monetary rewards in the human striatum. Neuron 58, 284–294. https://doi.org/10.1016/j.neuron.2008.03.020

Jackson, P.L., Brunet, E., Meltzoff, A.N., Decety, J., 2006. Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. Neuropsychologia 44, 752–761. https://doi.org/10.1016/j.neuropsychologia.2005.07.015

Janak, P.H., Tye, K.M., 2015. From circuits to behaviour in the amygdala. Nature. https://doi.org/10.1038/nature14188

Jenkins, A.C., Mitchell, J.P., 2010. Mentalizing under uncertainty: Dissociated neural responses to ambiguous and unambiguous mental state inferences. Cereb. Cortex 20, 404–410. https://doi.org/10.1093/cercor/bhp109

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841. https://doi.org/10.1016/S1053-8119(02)91132-8

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage 62, 782–90. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5, 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

Jeong, W., Chung, C.K., Kim, J.S., 2015. Episodic memory in aspects of large-scale brain networks. Front. Hum. Neurosci. 9, 1–15. https://doi.org/10.3389/fnhum.2015.00454

Jordan, M.R., Amir, D., Bloom, P., 2016. Are empathy and concern psychologically distinct? Emotion 16, 1107–1116. https://doi.org/10.1037/emo0000228

Kahneman, D., Knetsch, J.L., Thaler, R.H., 1986. Fairness and the assumptions of economics. J. Bus. 59, S285. https://doi.org/10.1086/296367

Kahneman, D., Tversky, A., 2013. Choices, Values, and Frames. pp. 269–278. https://doi.org/10.1142/9789814417358_0016

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47, 263–292.

Kanske, P., 2018. The social mind: disentangling affective and cognitive routes to understanding others. Interdiscip. Sci. Rev. 43, 115–124. https://doi.org/10.1080/03080188.2018.1453243

Kappes, A., Faber, N.S., Kahane, G., Savulescu, J., Crockett, M.J., 2018. Concern for Others Leads to Vicarious Optimism. Psychol. Sci. 29, 379–389. https://doi.org/10.1177/0956797617737129

Karns, C.M., Moore, W.E., Mayr, U., 2017. The Cultivation of Pure Altruism via Gratitude: A Functional MRI Study of Change with Gratitude Practice. Front. Hum. Neurosci. 11, 1–14. https://doi.org/10.3389/fnhum.2017.00599

Kawamichi, H., Tanabe, H.C., Takahashi, H.K., Sadato, N., 2013. Activation of the reward system during sympathetic concern is mediated by two types of empathy in a familiarity-dependent manner. Soc. Neurosci. 8, 90–100. https://doi.org/10.1080/17470919.2012.744349

Kawamichi, H., Yoshihara, K., Sugawara, S.K., Matsunaga, M., Makita, K., Hamano, Y.H., Tanabe, H.C., Sadato, N., 2015. Helping behavior induced by empathic concern attenuates anterior cingulate activation in response to others' distress. Soc. Neurosci. 0919, 1–14. https://doi.org/10.1080/17470919.2015.1049709

Kawamura, Y., Kusumi, T., 2018. Relationships between two types of reputational concern and altruistic behavior in daily life. Pers. Individ. Dif. 121, 19–24. https://doi.org/10.1016/j.paid.2017.09.003

Keltner, D., Kogan, A., Piff, P.K., Saturn, S.R., 2014. The sociocultural appraisals, values, and emotions (SAVE) framework of prosociality: Core processes from gene to meme. Annu. Rev. Psychol. 65, 425–460. https://doi.org/10.1146/annurev-psych-010213-115054

Kennerley, S.W., Dahmubed, A.F., Lara, A.H., Wallis, J.D., 2009. Neurons in the frontal lobe encode the value of multiple decision variables. J. Cogn. Neurosci. 21, 1162–1178. https://doi.org/10.1162/jocn.2009.21100

Kim, J.W., Kim, S.E., Kim, J.J., Jeong, B., Park, C.H., Son, A.R., Song, J.E., Ki, S.W., 2009. Compassionate attitude towards others' suffering activates the mesolimbic neural system. Neuropsychologia 47, 2073–2081. https://doi.org/10.1016/j.neuropsychologia.2009.03.017

King-Casas, B., Tomlin, D., Anen, C., Camerer, C.F., Quartz, S.R., Montague, P.R., 2005. Getting to know you: Reputation and trust in a two-person economic exchange. Science (80-.). 308, 78–83. https://doi.org/10.1126/science.1108062

Klimecki, O., Singer, T., 2012. Empathic Distress Fatigue Rather Than Compassion Fatigue? Integrating Findings from Empathy Research in Psychology and Social Neuroscience, in: Pathological Altruism. pp. 368–383. https://doi.org/10.1093/acprof:oso/9780199738571.003.0253

Klimecki, O.M., Mayer, S. V., Jusyte, A., Scheeff, J., Schönenberg, M., 2016. Empathy promotes altruistic behavior in economic interactions. Sci. Rep. 6, 785. https://doi.org/10.1038/srep31961

Knafo-Noam, A., Vertsberger, D., Israel, S., 2018. Genetic and environmental contributions to children's prosocial behavior: brief review and new evidence from a reanalysis of experimental twin data. Curr. Opin. Psychol. 20, 60–65. https://doi.org/10.1016/j.copsyc.2017.08.013

Knoch, D., Gianotti, L.R.R., Pascual-Leone, A., Treyer, V., Regard, M., Hohmann, M., Brugger, P., 2006. Disruption of right prefrontal cortex by low-frequency repetitive transcranial magnetic stimulation induces risk-taking behavior. J. Neurosci. 26, 6469–6472. https://doi.org/10.1523/jneurosci.0804-06.2006

Knoch, D., Schneider, F., Schunk, D., Hohmann, M., Fehr, E., 2009. Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. Proc. Natl. Acad. Sci. U. S. A. 106, 20895–9. https://doi.org/10.1073/pnas.0911619106

Knutsson, M., Martinsson, P., Wollbrant, C., 2013. Do people avoid opportunities to donate?. A natural field experiment on recycling and, charitable giving. J. Econ. Behav. Organ. 93, 71–77. https://doi.org/10.1016/j.jebo.2013.07.015

Koban, L., Pichon, S., Vuilleumier, P., 2014. Responses of medial and ventrolateral prefrontal cortex to interpersonal conflict for resources. Soc. Cogn. Affect. Neurosci. 9, 561–569. https://doi.org/10.1093/scan/nst020

Kogan, A., Oveis, C., Carr, E.W., Gruber, J., Mauss, I.B., Impett, E.A., Lowe, I. Van Der, Hui, B., Cheng, C., Shallcross, A., 2014. Vagal activity is quadratically related to prosocial traits, prosocial emotions, and observer perceptions of prosociality. J. Pers. Soc. Psychol. 107. https://doi.org/http://dx.doi.org/10.1037/a0037509

Konrath, S., Fuhrel-Forbis, A., Lou, A., Brown, S., 2012. Motives for volunteering are associated with mortality risk in older adults. Heal. Psychol. 31, 87–96. https://doi.org/10.1037/a0025226

Krach, S., Blümel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., van Os, J., Kircher, T., 2009. Are women better mindreaders? Sex differences in neural correlates of mentalizing detected with functional MRI. BMC Neurosci. 10. https://doi.org/10.1186/1471-2202-10-9

Krain, A.L., Wilson, A.M., Arbuckle, R., Castellanos, F.X., Milhama, M.P., 2006. Distinct neural mechanisms of risk and ambiguity: A meta-analysis of decision-making. Neuroimage 32, 477–484. https://doi.org/10.1016/j.neuroimage.2006.02.047

Krajbich, I., Bartling, B., Hare, T., Fehr, E., 2015. Rethinking fast and slow based on a critique of reaction-time reverse inference. Nat. Commun. 6, 7455. https://doi.org/10.1038/ncomms8455

Kuss, K., Falk, A., Trautner, P., Elger, C.E., Weber, B., Fliessbach, K., 2013. A reward prediction error for charitable donations reveals outcome orientation of donators. Soc. Cogn. Affect. Neurosci. 8, 216–223. https://doi.org/10.1093/scan/nsr088

Kuss, K., Falk, A., Trautner, P., Montag, C., Weber, B., Fliessbach, K., 2015. Neuronal correlates of social decision making are influenced by social value orientation - an

fMRI study. Front. Behav. Neurosci. 9, 1–8.
https://doi.org/10.3389/fnbeh.2015.00040

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, 2492–502. https://doi.org/10.1016/j.neuroimage.2010.10.014

Lamm, C., Rütgen, M., Wagner, I.C., 2019. Imaging empathy and prosocial emotions. Neurosci. Lett. 693, 49–53. https://doi.org/10.1016/j.neulet.2017.06.054

Lammers, J., Stapel, D. a., 2011. Power increases dehumanization. Gr. Process. Intergr. Relations 14, 113–126. https://doi.org/10.1177/1368430210370042

Lang, H., DeAngelo, G., Bongard, M., 2018. Theory of Mind and General Intelligence in Dictator and Ultimatum Games. Games 9, 16. https://doi.org/10.3390/g9020016

Lee, S., Feeley, T.H., 2016. The identifiable victim effect: a meta-analytic review. Soc. Influ. 4510, 0. https://doi.org/10.1080/15534510.2016.1216891

Levine, M., Prosser, A., Evans, D., Reicher, S., 2005. Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. Personal. Soc. Psychol. Bull. 31, 443–453. https://doi.org/10.1177/0146167204271651

Levine, M., Thompson, K., 2004. Identity, place, and bystander intervention: social categories and helping after natural disasters. J. Soc. Psychol. 144, 229–245. https://doi.org/10.3200/SOCP.144.3.229-245

Levy, D.J., Glimcher, P.W., 2012. The root of all value: A neural common currency for choice. Curr. Opin. Neurobiol. 22, 1027–1038. https://doi.org/10.1016/j.conb.2012.06.001

Li, M., Vietri, J., Galvani, A.P., Chapman, G.B., 2010. How do people value life? Psychol. Sci. 21, 163–7. https://doi.org/10.1177/0956797609357707

Li, Y., Zhao, M., 2019. Effects of the presence of others on prosocial behavior: Perceived face as mediator. Asian J. Soc. Psychol. 22, 193–202. https://doi.org/10.1111/ajsp.12358

Li, Z., Xia, S., Wu, X., Chen, Z., 2018. Analytical thinking style leads to more utilitarian moral judgments: An exploration with a process-dissociation approach. Pers. Individ. Dif. 131, 180–184. https://doi.org/10.1016/j.paid.2018.04.046

Liao, C., Wu, S., Luo, Y. jia, Guan, Q., Cui, F., 2018. Transcranial direct current stimulation of the medial prefrontal cortex modulates the propensity to help in costly helping behavior. Neurosci. Lett. 674, 54–59. https://doi.org/10.1016/j.neulet.2018.03.027

Lin, A., Adolphs, R., Rangel, A., 2012. Social and monetary reward learning engage overlapping neural substrates. Soc. Cogn. Affect. Neurosci. 7, 274–281. https://doi.org/10.1093/scan/nsr006

Lindsey, L.L.M., Kimo Ah Yun, Hill, J.B., 2007. Anticipated guilt as motivation to help unknown others: An examination of empathy as a moderator. Communic. Res. 34, 468–480. https://doi.org/10.1177/0093650207302789

Lischke, A., Lemke, D., Neubert, J., Hamm, A.O., Lotze, M., 2017. Inter-individual differences in heart rate variability are associated with inter-individual differences in mind-reading. Sci. Rep. 7, 1–6. https://doi.org/10.1038/s41598-017-11290-1

Liu, X., Hu, X., Shi, K., Mai, X., 2018. Empathy modulates the evaluation processing of

altruistic outcomes. Front. Psychol. 9, 1–10.
https://doi.org/10.3389/fpsyg.2018.00407

Lo, S., Andrews, S., 2015. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. Front. Psychol. 6, 1171.
https://doi.org/10.3389/fpsyg.2015.01171

Lockwood, P.L., 2016. The anatomy of empathy: Vicarious experience and disorders of social cognition. Behav. Brain Res. 311, 255–266.
https://doi.org/10.1016/j.bbr.2016.05.048

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. 113, 201603198. https://doi.org/10.1073/PNAS.1603198113

Lockwood, P.L., Apps, M.A.J.J., Roiser, J.P., Viding, E., 2015. Encoding of vicarious reward prediction in anterior cingulate cortex and relationship with trait empathy. J. Neurosci. 35, 13720–7. https://doi.org/10.1523/JNEUROSCI.1703-15.2015

Loewenstein, G., Small, D. a., 2007. The Scarecrow and the Tin Man: The vicissitudes of human sympathy and caring. Rev. Gen. Psychol. 11, 112–126.
https://doi.org/10.1037/1089-2680.11.2.112

Lole, L., Gonsalvez, C.J., Blaszczynski, A., Clarke, A.R., 2012. Electrodermal activity reliably captures physiological differences between wins and losses during gambling on electronic machines. Psychophysiology 49, 154–163.
https://doi.org/10.1111/j.1469-8986.2011.01290.x

Lotz, S., Schlösser, T., Cain, D.M., Fetchenhauer, D., 2013. The (in)stability of social preferences: Using justice sensitivity to predict when altruism collapses. J. Econ. Behav. Organ. 93, 141–148. https://doi.org/10.1016/j.jebo.2013.07.012

Luce, R.D., Raiffa, H., 1957. Games and decisions: Introduction and critical survey., Games and decisions: Introduction and critical survey. Wiley, Oxford, England.

Luo, J., 2018. The neural basis of and a common neural circuitry in different types of pro-social behavior. Front. Psychol. 9, 859. https://doi.org/10.3389/fpsyg.2018.00859

Ma, Q., Shen, Q., Xu, Q., Li, D., Shu, L., Weber, B., 2011. Empathic responses to others' gains and losses: An electrophysiological investigation. Neuroimage 54, 2472–2480.
https://doi.org/10.1016/j.neuroimage.2010.10.045

Ma, Y., Wang, C., Han, S., 2011. Neural responses to perceived pain in others predict real-life monetary donations in different socioeconomic contexts. Neuroimage 57, 1273–1280. https://doi.org/10.1016/j.neuroimage.2011.05.003

Mackey, S., Petrides, M., 2010. Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains. Eur. J. Neurosci. 32, 1940–1950.
https://doi.org/10.1111/j.1460-9568.2010.07465.x

Majdandžić, J., Amashaufer, S., Hummer, A., Windischberger, C., Lamm, C., 2016. The selfless mind: How prefrontal involvement in mentalizing with similar and dissimilar others shapes empathy and prosocial behavior. Cognition 157, 24–38.
https://doi.org/10.1016/j.cognition.2016.08.003

Majdandžić, J., Bauer, H., Windischberger, C., Moser, E., Engl, E., Lamm, C., 2012. The Human Factor: Behavioral and Neural Correlates of Humanized Perception in Moral Decision Making. PLoS One 7. https://doi.org/10.1371/journal.pone.0047698

Majerus, S., Cowan, N., Péters, F., Calster, L. Van, Phillips, C., Schrouff, J., 2016. Cross-Modal Decoding of Neural Patterns Associated with Working Memory: Evidence for Attention-Based Accounts of Working Memory. Cereb. Cortex 166–179. https://doi.org/10.1093/cercor/bhu189

Mak, L.E., Minuzzi, L., MacQueen, G., Hall, G., Kennedy, S.H., Milev, R., 2017. The default mode eetwork in healthy individuals: A systematic review and meta-analysis. Brain Connect. 7, 25–33. https://doi.org/10.1089/brain.2016.0438

Mars, R.B., Neubert, F.-X., Noonan, M.P., Sallet, J., Toni, I., Rushworth, M.F.S., 2012. On the relationship between the "default mode network" and the "social brain." Front. Hum. Neurosci. 6, 1–9. https://doi.org/10.3389/fnhum.2012.00189

Marsh, A.A., 2018. The neuroscience of empathy. Curr. Opin. Behav. Sci. 19, 110–115. https://doi.org/10.1016/j.cobeha.2017.12.016

Marsh, A.A., Stoycos, S.A., Brethel-Haurwitz, K.M., Robinson, P., VanMeter, J.W., Cardinale, E.M., 2014. Neural and cognitive characteristics of extraordinary altruists. Proc. Natl. Acad. Sci. U. S. A. 111, 15036–41. https://doi.org/10.1073/pnas.1408440111

Martin, Alia; Olson, K., 2015. Beyond Good and Evil: What Motivations Underlie Children's Prosocial Behaviour. Perspect. Psychol. Sci. 10, 159–175. https://doi.org/10.1177/1745691615568998

Martinsson, P., Myrseth, K.O.R., Wollbrant, C., 2012. Reconciling Pro-Social vs. Selfish Behavior: On the role of self-control. Judgm. Decis. Mak. 7, 304–315.

Masten, C.L., Morelli, S.A., Eisenberger, N.I., 2011. An fMRI investigation of empathy for "social pain" and subsequent prosocial behavior. Neuroimage 55, 381–388. https://doi.org/10.1016/j.neuroimage.2010.11.060

Mata, A., 2016. Proportion dominance in valuing lives : The role of deliberative thinking 11, 1–9.

Matthey, A., Regner, T., 2014. More than outcomes: The role of self-image in otherregarding behavior.

Mayr, M., 2012. The impact of numeracy on gain and non-loss donation decisions.

McAuliffe, W.H.B., Forster, D.E., Philippe, J., McCullough, M.E., 2018. Digital altruists: Resolving key questions about the empathy-altruism hypothesis in an internet sample. Emotion 18, 493–506. https://doi.org/10.1037/emo0000375

Meier, S., 2011. A Survey of Economic Theories and Field Evidence on Pro-Social Behavior. SSRN Electron. J. https://doi.org/10.2139/ssrn.917187

Mendez, M.F., 2009. The neurobiology of moral behavior: review and neuropsychiatric implications. CNS Spectr. 14, 608–20. https://doi.org/10.1017/S1092852900023853

Merritt, A.C., Effron, D. a., Monin, B., 2010. Moral Self-Licensing: When Being Good Frees Us to Be Bad. Soc. Personal. Psychol. Compass 4, 344–357. https://doi.org/10.1111/j.1751-9004.2010.00263.x

Milinski, M., Semmann, D., Bakker, T.C.M., Krambeck, H.J., 2001. Cooperation through indirect reciprocity: Image scoring or standing strategy? Proc. R. Soc. B Biol. Sci. 268, 2495–2501. https://doi.org/10.1098/rspb.2001.1809

Miller, E.M., Shankar, M.U., Knutson, B., McClure, S.M., 2014. Dissociating motivation from reward in human striatal activity. J. Cogn. Neurosci. 26, 1075–1084. https://doi.org/10.1162/jocn_a_00535

Miller, J.G., 2018. Physiological mechanisms of prosociality. Curr. Opin. Psychol. 20, 50–54. https://doi.org/10.1016/j.copsyc.2017.08.018

Mobbs, D., Yu, R., Meyer, M., Passamonti, L., Seymour, B., Calder, A.J., Schweizer, S., Frith, C.D., Dalgleish, T., 2009. A key role for similarity in vicarious reward. Science (80-.). 324, 900. https://doi.org/10.1126/science.1170539

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339, b2535. https://doi.org/10.1136/bmj.b2535

Molenberghs, P., 2013. The neuroscience of in-group bias. Neurosci. Biobehav. Rev. 37, 1530–1536. https://doi.org/10.1016/j.neubiorev.2013.06.002

Moll, J., De Oliveira-Souza, R., Basilio, R., Bramati, I.E., Gordon, B., Rodríguez-Nieto, G., Zahn, R., Krueger, F., Grafman, J., 2018. Altruistic decisions following penetrating traumatic brain injury. Brain 141, 1558–1569. https://doi.org/10.1093/brain/awy064

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., Grafman, J., 2006. Human fronto-mesolimbic networks guide decisions about charitable donation. Proc. Natl. Acad. Sci. U. S. A. 103, 15623–15628. https://doi.org/10.1073/pnas.0604475103

Morelli, S.A., Knutson, B., Zaki, J., 2018. Neural sensitivity to personal and vicarious reward differentially relate to prosociality and well-being. Soc. Cogn. Affect. Neurosci. 13, 831–839. https://doi.org/10.1093/scan/nsy056

Morelli, S.A., Rameson, L.T., Lieberman, M.D., 2014. The neural components of empathy: Predicting daily prosocial behavior. Soc. Cogn. Affect. Neurosci. 9, 39–47. https://doi.org/10.1093/scan/nss088

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. https://doi.org/10.1016/j.neuroimage.2014.12.056

Morey, R.A., McCarthy, G., Selgrade, E.S., Seth, S., Nasser, J.D., LaBar, K.S., 2012. Neural systems for guilt from actions affecting self versus others. Neuroimage 60, 683–692. https://doi.org/10.1016/j.neuroimage.2011.12.069

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

Müller, V.I., Cieslik, E.C., Laird, A.R., Fox, P.T., Radua, J., Mataix-Cols, D., Tench, C.R., Yarkoni, T., Nichols, T.E., Turkeltaub, P.E., Wager, T.D., Eickhoff, S.B., 2018. Ten simple rules for neuroimaging meta-analysis. Neurosci. Biobehav. Rev. https://doi.org/10.1016/j.neubiorev.2017.11.012

Murphy, R.O., Ackermann, K.A., Handgraaf, M.J.J., 2011. Measuring Social Value Orientation. Judgm. {&} Decis. Mak. 6, 771–781. https://doi.org/10.2139/ssrn.1804189

Nagel, J., Waldmann, M.R., 2013. Deconfounding distance effects in judgments of moral obligation. J. Exp. Psychol. Learn. Mem. Cogn. 39, 237–252. https://doi.org/10.1037/a0028641

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. U. S. A. 112, E2695–E2704. https://doi.org/10.1073/pnas.1410767112

Nichols, S., 2001. Mindreading and the cognitive architecture underlying altruistic motivation. Mind Lang. 16, 425–455. https://doi.org/10.1111/1468-0017.00178

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

Niemi, L., Wasserman, E., Young, L., 2018. The behavioral and neural signatures of distinct conceptions of fairness. Soc. Neurosci. 13, 399–415. https://doi.org/10.1080/17470919.2017.1333452

Niemi, L., Wasserman, E., Young, L., 2017. The behavioral and neural signatures of distinct conceptions of fairness. Soc. Neurosci. 00, 1–17. https://doi.org/10.1080/17470919.2017.1333452

Nook, E.C., Ong, D.C., Morelli, S.A., Mitchell, J.P., Zaki, J., 2016. Prosocial conformity: Prosocial norms generalize across behavior and empathy. Personal. Soc. Psychol. Bull. 42, 1045–1062. https://doi.org/10.1177/0146167298248001

Nowak, M. a, Sigmund, K., 2005. Evolution of indirect reciprocity. Nature 437, 1291–1298. https://doi.org/10.1038/nature04131

O'Brien, E., Kassirer, S., 2019. People Are Slow to Adapt to the Warm Glow of Giving. Psychol. Sci. 30, 193–204. https://doi.org/10.1177/0956797618814145

O'Connell, G., Christakou, A., Haffey, A.T., Chakrabarti, B., 2013. The role of empathy in choosing rewards from another's perspective. Front. Hum. Neurosci. 7, 174. https://doi.org/10.3389/fnhum.2013.00174

O'Connell, K., Brethel-Haurwitz, K.M., Rhoads, S.A., Cardinale, E.M., Vekaria, K.M., Robertson, E.L., Walitt, B., VanMeter, J.W., Marsh, A.A., 2019. Increased similarity of neural responses to experienced and empathic distress in costly altruism. Sci. Rep. 9. https://doi.org/10.1038/s41598-019-47196-3

Ojemann, J.G., Akbudak, E., Snyder, A.Z., McKinstry, R.C., Raichle, M.E., Conturo, T.E., 1997. Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. Neuroimage 6, 156–167. https://doi.org/10.1006/nimg.1997.0289

Olivola, C.Y., 2015. The Cognitive Psychology of Sensitivity to Human Fatalities: Implications for Life-Saving Policies. Policy Insights from Behav. Brain Sci. 2, 141–146. https://doi.org/10.1177/2372732215600887

Öngür, D., Ferry, A.T., Price, J.L., 2003. Architectonic subdivision of the human orbital and medial prefrontal cortex. J. Comp. Neurol. 460, 425–449. https://doi.org/10.1002/cne.10609

Ottoni-Wilhelm, M., Vesterlund, L., Xie, H., 2017. Why do people give? Testing pure and impure altruism. Am. Econ. Rev. 107, 3617–3633. https://doi.org/10.1257/aer.20141222

Palmer, C.J., Paton, B., Barclay, L., Hohwy, J., 2013a. Equality, Efficiency, and Sufficiency: Responding to Multiple Parameters of Distributive Justice During Charitable Distribution. Rev. Philos. Psychol. 4, 659–674. https://doi.org/10.1007/s13164-013-0157-0

Palmer, C.J., Paton, B., Ngo, T.T., Thomson, R.H., Hohwy, J., Miller, S.M., 2013b. Individual differences in moral behaviour: A role for response to risk and uncertainty? Neuroethics 6, 97–103. https://doi.org/10.1007/s12152-012-9158-4

Park, S.Q., Kahnt, T., Dogan, A., Strang, S., Fehr, E., Tobler, P.N., 2017. A neural link between generosity and happiness. Nat. Commun. 8, 1–10. https://doi.org/10.1038/ncomms15964

Penner, L.A., Dovidio, J.F., Piliavin, J.A., Schroeder, D.A., 2005. Prosocial behaviour: Multilevel Perspectives. Annu. Rev. Psychol. 56, 365–392. https://doi.org/10.1146/annurev.psych.56.091103.070141

Peters, J., Büchel, C., 2010. Neural representations of subjective reward value. Behav. Brain Res. 213, 135–141. https://doi.org/10.1016/j.bbr.2010.04.031

Petersen, M.B., Slothuus, R., Stubager, R., Togeby, L., 2011. Deservingness versus values in public opinion on welfare: The automaticity of the deservingness heuristic. Eur. J. Polit. Res. 50, 24–52. https://doi.org/10.1111/j.1475-6765.2010.01923.x

Peysakhovich, A., Nowak, M.A., Rand, D.G., 2014. Humans display a "cooperative phenotype" that is domain general and temporally stable. Nat. Commun. 5, 1–8. https://doi.org/10.1038/ncomms5939

Piff, P.K., Robinson, A.R., 2017. Social class and prosocial behavior: current evidence, caveats, and questions. Curr. Opin. Psychol. 18, 6–10. https://doi.org/10.1016/j.copsyc.2017.06.003

Pinazo, D., Barrós-loscertales, A., Peris, R., Ventura-campos, N., Ávila, C., 2016. Neuroticism Predisposes to Donation More Than Agreeableness : An fMRI Study 9, 100–108. https://doi.org/10.1037/npe0000054

Primes, G., Fieder, M., 2018. Real-life helping behaviours in North America: A genome-wide association approach. PLoS One 13, 1–16. https://doi.org/10.1371/journal.pone.0190950

Pronin, E., Olivola, C.Y., Kennedy, K. a., 2007. Doing Unto Future Selves As You Would Do Unto Others: Psychological Distance and Decision Making. Personal. Soc. Psychol. Bull. 34, 224–236. https://doi.org/10.1177/0146167207310023

Pulcu, E., Zahn, R., Moll, J., Trotter, P.D., Thomas, E.J., Juhasz, G., Deakin, J.F.W., Anderson, I.M., Sahakian, B.J., Elliott, R., 2014. Enhanced subgenual cingulate response to altruistic decisions in remitted major depressive disorder. NeuroImage Clin. 4, 701–710. https://doi.org/10.1016/j.nicl.2014.04.010

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Radua, J., Mataix-Cols, D., 2012. Meta-analytic methods for neuroimaging data explained. Biol. Mood Anxiety Disord. 2, 6. https://doi.org/10.1186/2045-5380-2-6

Radua, J., Mataix-Cols, D., Phillips, M.L., El-Hage, W., Kronhaus, D.M., Cardoner, N., Surguladze, S., 2012. A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. Eur. Psychiatry 27, 605–611. https://doi.org/10.1016/j.eurpsy.2011.04.001

Radua, J., Rubia, K., Canales-Rodríguez, E.J., Pomarol-Clotet, E., Fusar-Poli, P., Mataix-Cols, D., 2014. Anisotropic kernels for coordinate-based meta-analyses of neuroimaging studies. Front. Psychiatry 5, 1–8. https://doi.org/10.3389/fpsyt.2014.00013

Rameson, L.T., Morelli, S. a., Lieberman, M.D., 2012. The Neural Correlates of Empathy: Experience, Automaticity, and Prosocial Behavior. J. Cogn. Neurosci. 24, 235–245. https://doi.org/10.1162/jocn_a_00130

Rand, D.G., 2016. Cooperation , fast and slow : Meta-analytic evidence for a theory of social heuristics & self-interested deliberation. Psychol. Sci. https://doi.org/10.1177/0956797616654455

Rangel, A., Camerer, C.F., Montague, P.R., 2008. Neuroeconomics: The neurobiology of value-based decision-making. Nat. Rev. Neurosci. 9, 545–556. https://doi.org/10.1038/nrn2357.Neuroeconomics

Rapoport, A., 1987. Research Paradigms and Expected Utility Models for the Provision of Step-Level Public Goods. Psychol. Rev. 94, 74–83. https://doi.org/10.1037/0033-295X.94.1.74

Ribar, D.C., Wilhelm, M.O., 2002. Altruistic and Joy-of-Giving Motivations in Charitable Behavior. J. Polit. Econ. 110, 425–457. https://doi.org/10.1086/338750

Richman, S.B., DeWall, C.N., Wolff, M.N., 2015. Avoiding affection, avoiding altruism: Why is avoidant attachment related to less helping? Pers. Individ. Dif. 76, 193–197. https://doi.org/10.1016/j.paid.2014.12.018

Riekki, T., Svedholm-Häkkinen, A.M., Lindeman, M., 2018. Empathizers and systemizers process social information differently. Soc. Neurosci. 13, 616–627. https://doi.org/10.1080/17470919.2017.1368700

Rilling, J.K., Dagenais, J.E., Goldsmith, D.R., Glenn, A.L., Pagnoni, G., 2008. Social cognitive neural networks during in-group and out-group interactions. Neuroimage 41, 1447–1461. https://doi.org/10.1016/j.neuroimage.2008.03.044

Rilling, J.K., Gutman, D.A., Zeh, T.R., Pagnoni, G., Berns, G.S., Kilts, C.D., 2002. A neural basis for social cooperation. Neuron 35, 395–405. https://doi.org/10.1016/S0896-6273(02)00755-9

Ring, P., 2015. The framing effect and skin conductance responses. Front. Behav. Neurosci. 9, 1–7. https://doi.org/10.3389/fnbeh.2015.00188

Roy, A.K., Shehzad, Z., Margulies, D.S., Kelly, A.M.C., Uddin, L.Q., Gotimer, K., Biswal, B.B., Castellanos, F.X., Milham, M.P., 2009. Functional connectivity of the human amygdala using resting state fMRI. Neuroimage 45, 614–626. https://doi.org/10.1016/j.neuroimage.2008.11.030

RStudio Team, 2015. RStudio: Integrated Development for R. RStudio, Inc.

Ruci, L., van Allen, Z.M., Zelenski, J.M., 2018. Pro-social personality traits, helping behavior, and ego-depletion: Is helping really easier for the dispositionally pro-social? Pers. Individ. Dif. 120, 32–39. https://doi.org/10.1016/j.paid.2017.08.013

Ruff, C.C., Ugazio, G., Fehr, E., 2013. Changing social norm compliance with noninvasive brain stimulation. Science (80-.). 342, 482–484. https://doi.org/10.1126/science.1241399

Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M., 2014. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90, 449–468. https://doi.org/10.1016/j.neuroimage.2013.11.046

Sallet, J., Mars, R.B., Noonan, M.P., Neubert, F.-X., Jbabdi, S., O'Reilly, J.X., Filippini, N., Thomas, A.G., Rushworth, M.F., 2013. The organization of dorsal frontal cortex in humans and macaques. J. Neurosci. 33, 12255–12274. https://doi.org/10.1523/JNEUROSCI.5108-12.2013

Sassenrath, C., Pfattheicher, S., Keller, J., 2016. I might ease your pain, but only if you're sad: The impact of the empathized emotion in the empathy-helping association. Motiv. Emot. https://doi.org/10.1007/s11031-016-9586-2

Saxe, R., Haushofer, J., 2008. For Love or Money: A Common Neural Currency for??Social and Monetary Reward. Neuron 58, 164–165. https://doi.org/10.1016/j.neuron.2008.04.005

Schacter, D.L., Benoit, R.G., Szpunar, K.K., 2017. Episodic future thinking: mechanisms and functions. Curr. Opin. Behav. Sci. https://doi.org/10.1016/j.cobeha.2017.06.002

Schelling, T., 1968. The Life You Save May Be Your Own. Probl. public Expend.

Schneider-Hassloff, H., Straube, B., Nuscheler, B., Wemken, G., Kircher, T., 2015. Adult attachment style modulates neural responses in a mentalizing task. Neuroscience 303, 462–473. https://doi.org/10.1016/j.neuroscience.2015.06.062

Schreuders, E., Klapwijk, E.T., Will, G.J., Güroğlu, B., 2018. Friend versus foe: Neural correlates of prosocial decisions for liked and disliked peers. Cogn. Affect. Behav. Neurosci. 18, 127–142. https://doi.org/10.3758/s13415-017-0557-1

Schroeder, D.A., Dovidio, J.F., Sibicky, M.E., Matthews, L.L., Allen, J.L., 1988. Empathic concern and helping behavior: Egoism or altruism? J. Exp. Soc. Psychol. 24, 333–353. https://doi.org/10.1016/0022-1031(88)90024-8

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: A meta-analysis of functional brain imaging studies. Neurosci. Biobehav. Rev. 42, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Sepulcre, J., Liu, H., Talukdar, T., Martincorena, I., Thomas Yeo, B.T., Buckner, R.L., 2010. The organization of local and distant functional connectivity in the human brain. PLoS Comput. Biol. 6, 1–15. https://doi.org/10.1371/journal.pcbi.1000808

Sescousse, G., Caldú, X., Segura, B., Dreher, J.C., 2013. Processing of primary and secondary rewards: A quantitative meta-analysis and review of human functional neuroimaging studies. Neurosci. Biobehav. Rev. 37, 681–696. https://doi.org/10.1016/j.neubiorev.2013.02.002

Seu, I.B., 2016. 'The deserving': Moral reasoning and ideological dilemmas in public responses to humanitarian communications. Br. J. Soc. Psychol. 55, 739–755. https://doi.org/10.1111/bjso.12156

Shamay-Tsoory, S.G., 2011. The Neural Bases for Empathy. Neuroscientist 17, 18–24. https://doi.org/10.1177/1073858410379268

Shamay-Tsoory, S.G., Aharon-Peretz, J., Perry, D., 2009. Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. Brain 132, 617–27. https://doi.org/10.1093/brain/awn279

Sharot, T., 2011. The optimism bias. Curr. Biol. 21, R941–R945. https://doi.org/10.1016/j.cub.2011.10.030

Shdo, S.M., Ranasinghe, K.G., Gola, K.A., Mielke, C.J., Sukhanov, P. V., Miller, B.L., Rankin, K.P., 2018. Deconstructing empathy: Neuroanatomical dissociations between affect sharing and prosocial motivation using a patient lesion model. Neuropsychologia 116, 126–135. https://doi.org/10.1016/j.neuropsychologia.2017.02.010

Singer, P., 1972. Famine, Affluence, and Morality. Philos. Public Aff.

Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M.S., 2019. afex: Analysis of Factorial Experiments. R package version 0.25-1.

Slovic, P., 2010. The more who die, the less we care, in: Slovic, P., Erwann, M.-K. (Eds.), The Irrational Economist: Making Decisions in a Dangerous World.

Slovic, P., 2007. "If I look at the mass I will never act": Psychic numbing and genocide. Judgm. Decis. Mak. 2, 79–95. https://doi.org/10.1007/978-90-481-8647-1_3

Slovic, P., 2000. Thinking and Deciding Rationally About Catastrophic Losses of Human Lives, in: Michel-Kerjan, Slovic, P. (Eds.), The Irrational Economist: Future Directions in Behavioral Economics and Risk Management. Public Affairs Press.

Slovic, P., Västfjäll, D., Erlandsson, A., Gregory, R., 2017. Iconic photographs and the ebb and flow of empathic response to humanitarian disasters. Proc. Natl. Acad. Sci. 114. https://doi.org/10.1073/pnas.1613977114

Slovic, P., Zionts, D., Woods, A.K., Goodman, R., Jinks, D., Slovic, P., Zionts, D., Woods, A.K., Goodman, R., Jinks, D., 2011. Psychic Numbing and Mass Atrocity. Behav. Found. Policy 11–56. https://doi.org/10.2139/ssrn.1809951

Small, D. a., Loewenstein, G., Slovic, P., 2007. Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. Organ. Behav. Hum. Decis. Process. 102, 143–153. https://doi.org/10.1016/j.obhdp.2006.01.005

Small, D.A., 2010. Reference-dependent sympathy. Organ. Behav. Hum. Decis. Process. 112, 151–160. https://doi.org/10.1016/j.obhdp.2010.03.001

Small, D.A., Loewenstein, G., 2003. Helping a Victim or Helping the Victim: Altruism and Identifiabilit. J. Risk UncertaintyJournal Risk Uncertain. 26, 5–16. https://doi.org/10.1023/A

Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155. https://doi.org/10.1002/hbm.10062

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL, in: NeuroImage. https://doi.org/10.1016/j.neuroimage.2004.07.051

Sonne, J.W.H., Gash, D.M., 2018. Psychopathy to altruism: Neurobiology of the selfish-selfless spectrum. Front. Psychol. 9, 1–18. https://doi.org/10.3389/fpsyg.2018.00575

Spaans, J.P., Peters, S., Crone, E.A., 2018. Neural reward-related reactions to monetary gains for self and charity. Cogn. Affect. Behav. Neurosci. https://doi.org/10.3758/s13415-018-00672-1

Staib, M., Castegnetti, G., Bach, D.R., 2015. Optimising a model-based approach to inferring fear learning from skin conductance responses. J. Neurosci. Methods 255, 131–138. https://doi.org/10.1016/j.jneumeth.2015.08.009

Stanley, D.A., Sokol-Hessner, P., Fareri, D.S., Perino, M.T., Delgado, M.R., Banaji, M.R., Phelps, E.A., 2012. Race and reputation: Perceived racial group trustworthiness influences the neural correlates of trust decisions. Philos. Trans. R. Soc. B Biol. Sci. 367, 744–753. https://doi.org/10.1098/rstb.2011.0300

Steinbeis, N., Bernhardt, B.C., Singer, T., 2012. Impulse control and underlying functions of the left DLPFC mediate age-related and age-independent individual differences in strategic social behavior. Neuron 73, 1040–1051. https://doi.org/10.1016/j.neuron.2011.12.027

Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B., Sack, A.T., 2014. Be nice if you have to - the neurobiological roots of strategic fairness. Soc. Cogn. Affect. Neurosci. 10, 790–796. https://doi.org/10.1093/scan/nsu114

Strombach, T., Jin, J., Weber, B., Kenning, P., Shen, Q., Ma, Q., Kalenscher, T., 2014. Charity begins at home: Cultural differences in social discounting and generosity. J. Behav. Decis. Mak. 27, 235–245. https://doi.org/10.1002/bdm.1802

Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I.I., Tobler, P.N., Kalenscher, T., 2015. Social discounting involves modulation of neural value signals by temporoparietal junction. Proc. Natl. Acad. Sci. USA 112, 1619–1624. https://doi.org/10.1073/pnas.1414715112

Stürmer, S., Snyder, M., Omoto, A.M., 2005. Prosocial Emotions and Helping: The Moderating Role of Group Membership. J. Pers. Soc. Psychol. 88, 532–546. https://doi.org/10.1037/0022-3514.88.3.532

Sul, S., Tobler, P.N., Hein, G., Leiberg, S., Jung, D., Fehr, E., Kim, H., 2015. Spatial gradient in value representation along the medial prefrontal cortex reflects individual differences in prosociality. Proc. Natl. Acad. Sci. 112, 201423895. https://doi.org/10.1073/pnas.1423895112

Sze, J.A., Gyurak, A., Goodkind, M.S., Levenson, R.W., 2012. Greater Emotional Empathy and Prosocial Behavior in Late Life. Emotion 12, 1129–1140. https://doi.org/10.1037/a0025011

Tabibnia, G., Lieberman, M.D., 2007. Fairness and cooperation are rewarding: Evidence from social cognitive neuroscience. Ann. N. Y. Acad. Sci. 1118, 90–101. https://doi.org/10.1196/annals.1412.001

Tanaka, S.C., Yamada, K., Yoneda, H., Ohtake, F., 2014. Neural Mechanisms of Gain – Loss Asymmetry in Temporal Discounting. J. Neurosci. 34, 5595–5602. https://doi.org/10.1523/JNEUROSCI.5169-12.2014

Tappin, B.M., Capraro, V., 2018. Doing good vs. avoiding bad in prosocial choice: A refined test and extension of the morality preference hypothesis. J. Exp. Soc. Psychol. 79, 64–70. https://doi.org/10.1016/j.jesp.2018.06.005

Telle, N.-T., Pfister, H.-R., 2015. Positive Empathy and Prosocial Behavior: A Neglected Link. Emot. Rev. 1754073915586817-. https://doi.org/10.1177/1754073915586817

Telzer, E.H., Masten, C.L., Berkman, E.T., Lieberman, M.D., Andrew, J., 2010. Gaining while giving: An fMRI study of the rewards of family assistance among White and Latino youth. Soc. Neurosci. 5, 508–518. https://doi.org/10.1080/17470911003687913.Gaining

Telzer, E.H., Masten, C.L., Berkman, E.T., Lieberman, M.D., Fuligni, A.J., 2011. Neural regions associated with self control and mentalizing are recruited during prosocial behaviors towards the family. Neuroimage 58, 242–249. https://doi.org/10.1016/j.neuroimage.2011.06.013

Thomas, K.A., Freitas, J. De, Descioli, P., Pinker, S., 2016. Recursive mentalizing and common knowledge in the bystander effect recursive mentalizing and common

knowledge in the bystander effect. J Exp Psychol Gen. 145, 621–9.
https://doi.org/10.1037/xge0000153

Tom, S.M., Fox, C.R., Trepel, C., Poldrack, R.A., 2007. The neural basis of loss aversion in
decision-making under risk. Science (80-.). 315, 515–518.
https://doi.org/10.1126/science.1134239

Torstveit, L., Sütterlin, S., Lugo, R.G., 2016. Empathy, guilt proneness, and gender: Relative
contributions to prosocial behaviour. Eur. J. Psychol. 12, 260–270.
https://doi.org/10.5964/ejop.v12i2.1097

Touré-Tillery, M., Fishbach, A., 2017. Too far to help: The effect of perceived distance on
the expected impact and likelihood of charitable action. J. Pers. Soc. Psychol. 112,
860–876. https://doi.org/10.1037/pspi0000089

Tovote, P., Fadok, J.P., Lüthi, A., 2015. Neuronal circuits for fear and anxiety. Nat. Rev.
Neurosci. https://doi.org/10.1038/nrn3945

Tusche, A., Bockler, A., Kanske, P., Trautwein, F.-M., Singer, T., 2016. Decoding the
charitable brain: Empathy, perspective taking, and attention shifts differentially
predict altruistic giving. J. Neurosci. 36, 4719–4732.
https://doi.org/10.1523/JNEUROSCI.3392-15.2016

Ugazio, G., Majdandžić, J., Lamm, C., 2014. Are Empathy and Morality Linked? Insights from
Moral Psychology, Social and Decision Neuroscience, and Philosophy. Empathy Moral.
155–171.

UNICEF, 2018. Annual report.

Urbanska, K., McKeown, S., Taylor, L.K., 2019. From injustice to action: The role of empathy
and perceived fairness to address inequality via victim compensation. J. Exp. Soc.
Psychol. 82, 129–140. https://doi.org/10.1016/j.jesp.2019.01.010

Van Bavel, J.J., FeldmanHall, O., Mende-Siedlecki, P., 2015. The neuroscience of moral
cognition: From dual processes to dynamic systems. Curr. Opin. Psychol. 6, 167–172.
https://doi.org/10.1016/j.copsyc.2015.08.009

Van Iwaarden, J., Van Der Wiele, T., Williams, R., Moxham, C., 2009. Charities: How
important is performance to donors? Int. J. Qual. Reliab. Manag. 26, 5–22.
https://doi.org/10.1108/02656710910924143

Van Lange, P. a. M., Bekkers, R., Schuyt, T.N.M., Vugt, M. Van, 2007. From Games to Giving:
Social Value Orientation Predicts Donations to Noble Causes. Basic Appl. Soc. Psych.
29, 375–384. https://doi.org/10.1080/01973530701665223

Varnum, M.E.W., Shi, Z., Chen, A., Qiu, J., Han, S., 2014. When "Your" reward is the same as
"My" reward: Self-construal priming shifts neural responses to own vs. friends'
rewards. Neuroimage 87, 164–169.
https://doi.org/10.1016/j.neuroimage.2013.10.042

Vassena, E., Holroyd, C.B., Alexander, W.H., 2017. Computational models of anterior
cingulate cortex: At the crossroads between prediction and effort. Front. Neurosci.
11, 316. https://doi.org/10.3389/fnins.2017.00316

Västfjäll, D., Slovic, P., Mayorga, M., 2015. Pseudoinefficacy: negative feelings from children
who cannot be helped reduce warm glow for children who can be helped. Front.
Psychol. 6, 616. https://doi.org/10.3389/fpsyg.2015.00616

Västfjäll, D., Slovic, P., Mayorga, M., Peters, E., 2014. Compassion fade: Affect and charity

are greatest for a single child in need. PLoS One 9.
https://doi.org/10.1371/journal.pone.0100115

Vitaglione, G.D., Barnett, M. a, 2003. Assessing a new dimension of empathy: Empathic anger as a predictor of helping and punishing desires. Motiv. Emot. 27, 301–325. https://doi.org/10.1023/A:1026231622102

Vives, M.L., Feldmanhall, O., 2018. Tolerance to ambiguous uncertainty predicts prosocial behavior. Nat. Commun. 9, 2156. https://doi.org/10.1038/s41467-018-04631-9

Vroome, T. De, Verkuyten, M., Martinovic, B., 2014. Host National Identification of Immigrants in the Netherlands. Int. Migr. Rev. 48, 1–27. https://doi.org/10.1111/imre.12063

Wager, T.D., Lindquist, M., Kaplan, L., 2007. Meta-analysis of functional neuroimaging data: Current and future directions. Soc. Cogn. Affect. Neurosci. 2, 150–158. https://doi.org/10.1093/scan/nsm015

Wager, T.D., Lindquist, M.A., Nichols, T.E., Kober, H., Van Snellenberg, J.X., 2009. Evaluating the consistency and specificity of neuroimaging data using meta-analysis. Neuroimage. https://doi.org/10.1016/j.neuroimage.2008.10.061

Wang, J.X., Hughes, E., Fernando, C., Czarnecki, W.M., Duenez-Guzman, E.A., Leibo, J.Z., 2018. Evolving intrinsic motivations for altruistic behavior. Preprint 1–12.

Waytz, A., Mitchell, J.P., 2011. Two mechanisms for simulating other minds: Dissociations between mirroring and self-projection. Curr. Dir. Psychol. Sci. 20, 197–200. https://doi.org/10.1177/0963721411409007

Weiland, S., Hewig, J., Hecht, H., Mussel, P., Miltner, W.H.R., 2012. Neural correlates of fair behavior in interpersonal bargaining. Soc. Neurosci. 7, 537–551. https://doi.org/10.1080/17470919.2012.674056

Weiskopf, N., Hutton, C., Josephs, O., Turner, R., Deichmann, R., 2007. Optimized EPI for fMRI studies of the orbitofrontal cortex: Compensation of susceptibility-induced gradients in the readout direction. Magn. Reson. Mater. Physics, Biol. Med. 20, 39–49. https://doi.org/10.1007/s10334-006-0067-6

Weng, H.Y., Fox, A.S., Shackman, A.J., Stodola, D.E., Caldwell, J.Z.K.K., Olson, M.C., Rogers, G.M., Davidson, R.J., 2013. Compassion Training Alters Altruism and Neural Responses to Suffering. Psychol. Sci. 24, 1171–1180. https://doi.org/10.1177/0956797612469537

Weyant, J.M., 1978. Effects of mood states, costs, and benefits on helping. J. Pers. Soc. Psychol. 36, 1169–1176. https://doi.org/10.1037/0022-3514.36.10.1169

Wiech, K., Kahane, G., Shackel, N., Farias, M., Savulescu, J., Tracey, I., 2013. Cold or calculating? Reduced activity in the subgenual cingulate cortex reflects decreased emotional aversion to harming in counterintuitive utilitarian judgment. Cognition 126, 364–372. https://doi.org/10.1016/j.cognition.2012.11.002

Wilkes, B.L., Gonsalvez, C.J., Blaszczynski, A., 2017. Capturing SCL and HR changes to win and loss events during gambling on electronic machines. Int. J. Psychophysiol. 78, 265–272. https://doi.org/10.1016/j.ijpsycho.2010.08.008

Woolrich, M.W., Behrens, T.E.J., Beckmann, C.F., Jenkinson, M., Smith, S.M., 2004. Multilevel linear modelling for FMRI group analysis using Bayesian inference. Neuroimage 21, 1732–1747. https://doi.org/10.1016/j.neuroimage.2003.12.023

Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage 14, 1370–1386. https://doi.org/10.1006/nimg.2001.0931

Yamagishi, T., Matsumoto, Y., Kiyonari, T., Takagishi, H., Li, Y., Kanai, R., Sakagami, M., 2017. Response time in economic games reflects different types of decision conflict for prosocial and proself individuals. Proc. Natl. Acad. Sci. 114, 6394–6399. https://doi.org/10.1073/pnas.1608877114

Yamagishi, T., Mifune, N., Li, Y., Shinada, M., Hashimoto, H., Horita, Y., Miura, A., Inukai, K., Tanida, S., Kiyonari, T., Takagishi, H., Simunovic, D., 2013. Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. Organ. Behav. Hum. Decis. Process. 120, 260–271. https://doi.org/10.1016/j.obhdp.2012.06.002

Yamagishi, T., Takagishi, H., Fermin, A. de S.R., Kanai, R., Li, Y., Matsumoto, Y., 2016. Cortical thickness of the dorsolateral prefrontal cortex predicts strategic choices in economic games. Proc. Natl. Acad. Sci. 201523940. https://doi.org/10.1073/pnas.1523940113

Yankouskaya, A., Humphreys, G., Stolte, M., Stokes, M., Moradi, Z., Sui, J., 2017. An anterior–posterior axis within the ventromedial prefrontal cortex separates self and reward. Soc. Cogn. Affect. Neurosci. 12, 1859–1868. https://doi.org/10.1093/scan/nsx112

Yarkoni, T., Ashar, Y.K., Wager, T.D., 2015. Interactions between donor Agreeableness and recipient characteristics in predicting charitable donation and positive social evaluation. PeerJ 1–11. https://doi.org/10.7717/peerj.1089

Yi, R., Pickover, A., Stuppy-Sullivan, A.M., Baker, S., Landes, R.D., 2016. Impact of episodic thinking on altruism. J. Exp. Soc. Psychol. 65, 74–81. https://doi.org/10.1016/j.jesp.2016.03.005

Yost-Dubrow, R., Dunham, Y., 2018. Evidence for a relationship between trait gratitude and prosocial behaviour. Cogn. Emot. 32, 397–403. https://doi.org/10.1080/02699931.2017.1289153

Yu, H., Siegel, J.Z., Crockett, M.J., 2019. Modeling Morality in 3-D: Decision-Making, Judgment, and Inference. Top. Cogn. Sci. 11, 409–432. https://doi.org/10.1111/tops.12382

Zagefka, H., 2018. It could have been me: Proximity motivates disaster giving. Int. J. Nonprofit Volunt. Sect. Mark. 23, 1–6. https://doi.org/10.1002/nvsm.1587

Zagefka, H., James, T., 2015. The psychology of charitable donations to disaster victims and beyond. Soc. Issues Policy Rev. 9, 155–192. https://doi.org/10.1111/sipr.12013

Zahn, R., de Oliveira-Souza, R., Bramati, I., Garrido, G., Moll, J., 2009. Subgenual cingulate activity reflects individual differences in empathic concern. Neurosci. Lett. 457, 107–110. https://doi.org/10.1016/j.neulet.2009.03.090

Zahn, R., de Oliveira-Souza, R., Moll, J., 2019. Moral Motivation and the Basal Forebrain. Neurosci. Biobehav. Rev. 108, 127065. https://doi.org/10.1016/j.neubiorev.2019.10.022

Zaki, J., Mitchell, J.P., 2016. Prosociality as a Form of Reward Seeking, in: Positive Neuroscience. pp. 57–72. https://doi.org/10.1093/acprof:oso/9780199977925.003.0005

Zaki, J., Mitchell, J.P., 2011. Equitable decision making is associated with neural markers of intrinsic value. Proc. Natl. Acad. Sci. 108, 19761–19766.

https://doi.org/10.1073/pnas.1112324108

Zanon, M., Novembre, G., Zangrando, N., Chittaro, L., Silani, G., 2014. Brain activity and prosocial behavior in a simulated life-threatening situation. Neuroimage 98, 134–146. https://doi.org/10.1016/j.neuroimage.2014.04.053

Zaval, L., Markowitz, E.M., Weber, E.U., 2015. How Will I Be Remembered? Conserving the Environment for the Sake of One's Legacy. Psychol. Sci. 26, 231–236. https://doi.org/10.1177/0956797614561266

Zheng, H., Zhu, L., 2013. Neural mechanism of proposer's decision-making in the ultimatum and dictator games. Neural Regen. Res. 8, 357–62. https://doi.org/10.3969/j.issn.1673-5374.2013.04.008

Zink, C.F., Pagnoni, G., Martin-Skurski, M.E., Chappelow, J.C., Berns, G.S., 2004. Human striatal responses to monetary reward depend on saliency. Neuron 42, 509–517. https://doi.org/10.1016/S0896-6273(04)00183-7

Supplementary materials

1. Method

1.1. Additional analysis details

In addition to the criteria for inclusion, studies with coordinates were required to use the same statistical threshold for all brain regions, often meaning coordinates identified in whole brain analysis rather than more liberally thresholded ROIs, although studies could have employed different thresholds to each other (Radua et al., 2012).

If t-values were not available, we calculated them from the z-values or p-values in the manuscripts. If a threshold p-value was given, rather than an exact one, this provides a minimum t-value (e.g. if $p < 0.01$ then the t-value will be at least the value which corresponds to $p = 0.01$).

Two studies in the altruistic group (Izuma et al., 2010; Weiland et al., 2012) were identified as also having data eligible for strategic decisions. However, data from the same participants in different groups violates independence assumptions, which would make p-values inaccurate in the meta-analysis. We allocated these studies to the altruistic group, as the group with fewer papers identified, to make the groups as equal as possible.

1.2. Covariate coding

A covariate for *comparator complexity* was used in the analysis with codes ranging from 1 to 4. Rest contrasts were coded 1 as minimum complexity. Contrasts with selfish decisions were coded 4 due to the multiple processes involved and consequences for the self and another person. Within the group of visuomotor controls, there were differences in whether the control condition had no consequence or a consequence coded 2 and 3 respectively. For example, some studies used dictator game trials without an opportunity to be altruistic as controls, which affect the participants' winnings, whereas other controls had no consequence (e.g. simply press a button).

We chose consequence as the determining factor for coding visuomotor controls as it links to some of the key processes of interest, including intrinsic / extrinsic motivation, and anticipatory processes linked to reward certainty and reward timing. However, there were also differences in the degree to which the control task was cognitively taxing, for example playing a lottery requires non-social risk calculation, which is more difficult, compared to controls that simply required participants to click a target on the screen. Codes based on this level of difficulty did not always align with consequence i.e. some controls were difficult but had no consequences, others were easy but had consequences etc. Results for the two sets of analyses showed very similar results and we only report those based on consequence here to avoid repetition.

1.3. Subgroup of studies with a selfish contrast

In addition to contrasts that included all data, we also ran specific contrasts on a subgroup of studies with a control that was the selfish alternative to the prosocial choice,

the most common control provided. This was to enable the use of additional functions of AES:SDM that are not possible when including covariates. These included the examination of heterogeneity of findings was assessed using between-study variance analysis, with significance showing larger between-study variance than would result from sampling error alone. Whether areas of significant activation showed significant heterogeneity was tested through a simple overlap analysis. Robustness of findings was estimated using jackknife sensitivity analysis which repeats the analysis as many times as there are studies, excluding one each time to test replicability. This analysis was quantified as a proportion of replications voxel by voxel using SPM12 (Statistical Parametric Mapping, http://www.fil.ion.ucl.ac.uk/spm) to give significant voxels in each jackknife repeat a value of 1, summing maps and dividing by the total number of repetitions.

Some studies in the strategic group provided maps for two control conditions e.g. a rest contrast and contrast with selfish decisions. As selfish decisions were the most common control, rest and visuomotor alternatives were favoured for the analysis using the covariate to better match the altruistic group which had more of these non-selfish contrasts. However, data from the selfish contrast was included in this subgroup analysis.

1.4. ROI analysis

To test the anatomical prediction that altruistic decisions would be associated with greater posterior and strategic decisions more anterior vmPFC activity, activation estimates were extracted across an anatomical axis. The location of the axis was defined by spheres used previously (Campbell-Meiklejohn *et al.*, 2016). These 10mm spheres spanned the superior medial gyrus, dorsal to the gyrus rectus. The axis began in area 10 m/14 m (Mackey and Petrides, 2010; Öngür et al., 2003), at MNI coordinates [0 mm, 32 mm, −16 mm] and ended at the frontal pole at MNI coordinates [0mm, 64mm, 8mm], the approximate centre of ventromedial area 10 (Supplementary Figure 1).

As effect sizes should not be averaged from related voxels, we created an axis of single voxels following the line of the spheres down the midline (x = 0, see Supplementary Figure 1). The posterior end of the axis was therefore within a region associated with choice value (Bartra et al., 2013; Levy and Glimcher, 2012) whereas the anterior end was a region associated with abstract reasoning about mental states (Amodio and Frith, 2006). Values were also extracted at x = 4 and x = -4 to better represent the area covered by the spheres and as the longitudinal fissure at x = 0 means there is no data for the most anterior 3 spheres.

For plots, effect sizes (Hedges' G) were extracted from the meta-analytic maps for each voxel for every study. To formally test these relationships, the weighted effect sizes were examined by mixed-effect linear regressions using the lme4 package (Bates et al., 2015) in the R environment (R Development Core Team, 2006). This was done for the

original analysis (in which spheres for some studies had values of 0 due to dropout) and also only including voxels with data present.
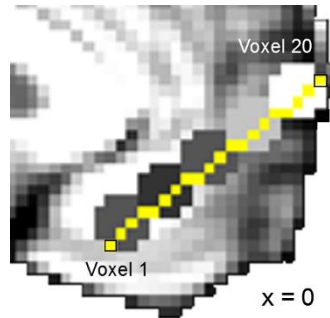
Supplementary Figure 1. vmPFC ROI analysis: location of voxels and spheres

1.5. Adjusted analysis accounting for dropout

Maps recreated from coordinates were not adjusted as 0 values represent too greater distance from a peak to receive an effect size estimation, rather than missing data. However, it is possible that the original datasets missed peak activations in vmPFC. Missing data were assumed to be at random as no factor showed significant differences in, an effect on, or correlation with the number of voxels covered (date of the study, number of participants, control condition, task used, MNI / Talairach space; all ps > 0.1).

As the adjusted analysis resulted in voxels having differing numbers of studies included in the calculation of their effect size, the permutation test for significance was not appropriate. Maps of vmPFC were instead thresholded at an uncorrected SDM-Z value of 2.3. This common value for thresholding is close to the average of the critical z values generated in the permutation tests for the original analyses and AES:SDM analyses run with the 50% of maps with the best coverage. This uncorrected threshold may initially seem liberal. However, as noted by Radua *et al.* (2012), the fact that coordinate studies have large numbers of voxels with null effect sizes from being far from any peak makes z-values under the empirical distributions given by the permutation tests associated with lower p-values than they would be under the normal distribution. Reported values are described specifically as SDM-Z as they do not follow the standard normal distribution.

2. Results

Supplementary Table 1.

Peak activations from modelled contrasts using complexity coordinate for altruistic vs. rest

Peak label	BA*	coordinates			SDM-Z	Vox
		x	y	z		
Altruistic > rest						
L inferior occipital cortex		-32	-82	-12	7.26	3037
R inferior occipital cortex		38	-86	-6	7.83	2296
L superior lateral occipital cortex	7	-26	-70	46	5.48	1127
Paracingulate gyrus	32	0	14	48	5.29	947
L middle frontal gyrus	48	-42	28	24	5.60	813
R superior lateral occipital cortex	7	36	-64	38	4.55	379
R insula	48	32	20	8	4.67	151
L thalamus		-6	-22	14	4.21	63
L thalamus	27	-20	-32	0	4.41	52
R hippocampus		28	-30	-4	4.64	48
R thalamus		16	-18	18	4.72	43
R frontal pole	45	46	36	24	3.93	16
Cerebellum vermic lobule IV / V		2	-44	0	4.34	10
Rest > altruistic						
L superior lateral occipital cortex	39	-48	-76	22	6.52	2696
Frontal medial cortex	11	4	54	-12	5.32	2655
R middle temporal gyrus	21	66	-48	6	4.84	2603
Posterior cingulate gyrus		-8	-36	42	4.05	1948
L middle temporal gyrus		-52	-26	-12	4.9	956
R middle temporal gyrus	20	56	-12	-16	4.79	924
L precuneus cortex		-30	-56	8	4.11	419
L superior frontal gyrus		-24	22	38	4.50	269
R inferior frontal gyrus	45	52	34	0	4.02	147
L orbitofrontal cortex	47	-30	34	-14	4.03	113
L inferior frontal gyrus	45	-50	32	4	3.37	83
L occipital pole	18	-6	-94	24	3.19	74
R frontal pole		22	34	-10	3.51	22
R cerebellum, hemispheric lobule IV / V	30	16	-34	-18	2.63	17
R postcentral gyrus	3	30	-38	70	2.36	17
L parahippocampal gyrus	30	-14	-36	-14	2.83	14

Note. L = left, R = right, BA = Brodmann area, coordinates (x, y, z) in MNI space, SDM-Z= SDM z-value of activation, Vox = number of voxels in the cluster. *If cluster falls within *more* than one BA, none is reported either by AES:SDM software, or here.

Supplementary Table 2.

Peak activations from modelled contrasts using complexity coordinate for strategic vs. rest

| Peak label	BA*	coordinates			SDM-Z	Vox
		x	y	z		
Strategic > rest						
L occipital cortex		-34	-64	-14	8.06	4825
R occipital cortex		38	-84	-6	8.49	3646
Paracingulate / anterior cingulate cortex		12	28	24	6.08	1341
L inferior frontal gyrus	45	-46	34	16	6.21	1004
R insula	48	32	20	6	5.46	340
R caudate		16	-12	22	6.15	224
R frontal pole		28	38	20	5.09	219
Posterior cingulate gyrus		4	-32	24	5.33	182
R frontal pole	45	48	40	22	4.81	115
L frontal pole	10	-28	50	10	4.97	80
R hippocampus		32	-24	-8	5.72	51
Anterior cingulate gyrus		6	-6	32	4.95	39
L putamen		-24	0	2	4.80	35
Cerebellum, vermic lobule VI	18	6	-68	-14	4.77	30
L insula	48	-32	16	4	4.70	29
R precentral gyrus	44	48	8	30	4.50	24
R frontal operculum cortex	38	50	18	-4	4.55	16
R frontal pole	10	30	52	10	4.49	16
Brainstem		8	-40	-28	5.02	13
L thalamus		-20	-30	0	4.63	10
Rest > strategic						
L superior lateral occipital cortex	39	-56	-62	26	6.49	3625
R middle temporal gyrus	21	60	-8	-18	5.71	3027
Frontal medial cortex	11	4	54	-12	4.44	2577
R precentral gyrus	6	8	-14	78	4.02	2061
L inferior frontal gyrus	45	-50	32	2	4.44	481
L middle frontal gyrus		-26	28	44	3.99	367
L precuneus cortex		-26	-52	8	2.80	265
R cerebellum, crus II		24	-84	-36	3.44	116
L parahippocampal gyrus		-28	-44	6	3.94	104
R parahippocampal gyrus		22	-20	-24	4.22	91
L parahippocampal gyrus	30	-20	-20	-28	4.52	82
L precentral gyrus	6	-8	-14	78	3.72	58
L cerebellum, hemispheric lobule IV/V	30	-14	-36	-16	3.57	40
R lingual gyrus		32	-52	4	2.37	30
R central opercular cortex	48	40	-18	22	1.95	24
R precuneus cortex		16	-38	40	2.35	12

Note. L = left, R = right, BA = Brodmann area, coordinates (x, y, z) in MNI space, SDM-Z= SDM z-value of activation, Vox = number of voxels in the cluster. *If cluster falls within *more* than one BA, none is reported either by AES:SDM software, or here.

2.1. Results from the subgroup of studies with a selfish contrast only

2.1.1. *Means, heterogeneity & robustness*

Limiting altruistic studies to those with a selfish control showed activity in right nucleus accumbens (NuAcc), vmPFC, ACC, left AI, orbitofrontal cortex (OFC), bilateral precuneus, left inferior dlPFC, posterior cingulate cortex (PCC). Selfish decisions activated bilateral posterior insula, left dlPFC (area 46D), bilateral putamen, right caudate nucleus, bilateral pSTS and left amygdala (Supplementary Figure 2a & Supplementary Table 3).

Jackknife sensitivity analysis showed these clusters were robust with the majority of voxels replicating in most repetitions of the analysis with one study removed. Heterogeneity analysis showed no significant between-study variation in any of the voxels significantly active in either direction (Supplementary Figure 3a).
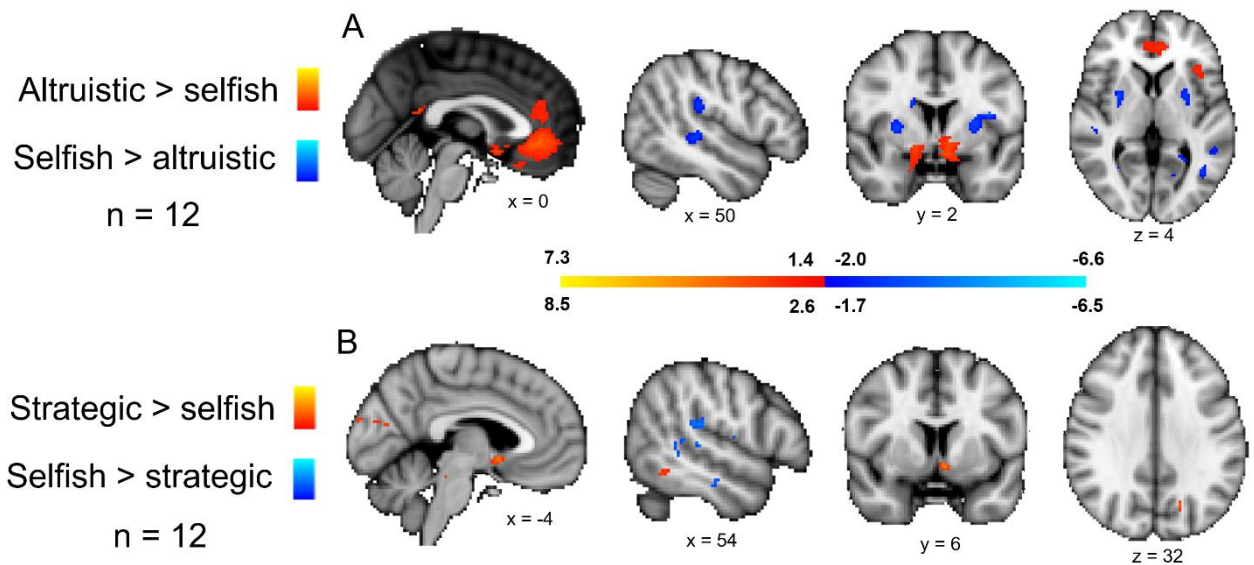
Supplementary Figure 2. Mean activations from **A:** altruistic and **B:** strategic meta-analytic maps from specific contrasts; only those with a selfish control (thresholded with permutation analysis run in AES:SDM). Coronal images in radiological orientation (right = left)
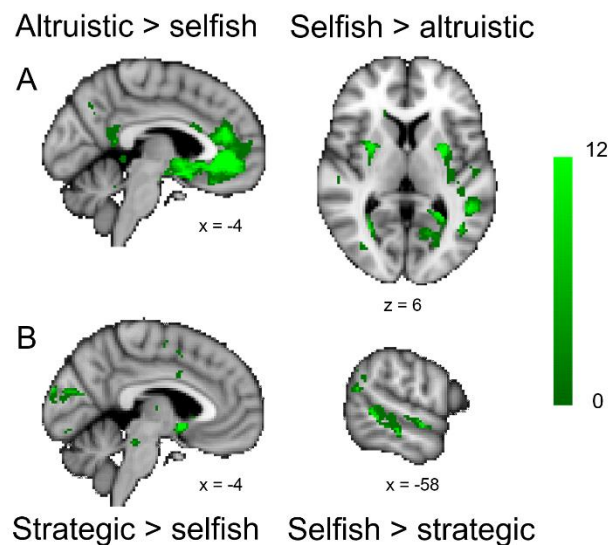
Supplementary Table 3.
Peak activations from specific contrasts (studies with a selfish control only) for altruistic vs. selfish

Peak label	BA*	coordinates			SDM-Z	Vox
		x	y	z		
Altruistic > selfish						
Anterior cingulate cortex	11	-2	38	-6	3.94	1738
R striatum	25	14	0	-12	3.05	276
L frontal operculum cortex	47	-38	24	4	2.31	42
Posterior cingulate cortex		0	-48	18	2.66	25
L precuneus cortex		-10	-66	36	2.17	23
R supramarginal gyrus		60	-36	50	2.42	16
L middle temporal gyrus		-60	-54	-8	2.37	13
L frontal pole	46	-48	50	-4	2.20	11
R precuneus cortex		8	-68	36	2.05	11
Selfish > altruistic						
L central opercular cortex		-42	-14	22	3.84	433
L postcentral / precentral gyri	6	-50	-10	50	2.97	232
L lingual gyrus		-26	-58	-4	2.88	142
L middle frontal gyrus	46	-28	36	30	2.83	105
L inferior lateral occipital cortex	37	-40	-64	0	3.05	90
R parietal operculum cortex		46	-26	24	2.80	88
R middle temporal gyrus	21	48	-32	-2	2.75	85
L precentral gyrus	6	-32	-12	54	2.58	85
L middle temporal gyrus	21	-48	-48	6	2.69	74
R caudate nucleus		18	-16	26	3.05	67
R putamen	48	30	2	6	2.73	71
R superior lateral occipital cortex		34	-74	14	2.68	61
L supramarginal gyrus	41	-46	-40	24	2.68	51
R temporal occipital fusiform cortex		34	-46	-8	2.60	25
L postcentral gyrus	3	-34	-32	58	2.36	25
R inferior temporal gyrus		46	-54	-4	2.42	22
R hippocampus	20	36	-14	-18	2.69	21
L inferior temporal gyrus		-48	-14	-26	2.64	20
L precentral gyrus	4	-6	-26	60	2.33	20
R intracalcarine cortex		30	-62	6	2.80	17
L postcentral gyrus	3	-44	-24	58	2.27	16
L postcentral gyrus	48	-62	-10	16	2.45	15
R precentral gyrus	6	22	-14	66	2.31	15
R superior parietal lobule	1	30	-44	68	2.46	14
R superior parietal lobule	5	20	-52	64	2.23	13
L superior temporal gyrus		-44	-8	-18	2.44	12
L lingual gyrus	18	-10	-78	-10	2.29	11
L amygdala		-30	-8	-18	2.45	10

Note. L = left, R = right, BA = Brodmann area, coordinates (x, y, z) in MNI space, SDM-Z= SDM z-value of activation, Vox = number of voxels in the cluster. *If cluster falls within *more* than one BA, none is reported either by AES:SDM software, or here.

Isolating the studies with a selfish contrast for strategic decisions showed significant activation in left NuAcc, right dlPFC (area 46), left precuneus, left amygdala, right cerebellum, right AI and right frontal pole. Selfish decisions in strategic tasks, compared to prosocial ones, showed activation in left TPJ, bilateral posterior temporal sulci, left temporal pole, left hippocampus, right supplementary motor area, right cerebellum and right posterior insula (Supplementary Figure 2b & Supplementary Table 4).

No heterogeneity was found in these significant voxels and activations for strategic > selfish were robust. Some peaks for where selfish > strategic showed a small, robust region surrounded by voxels only replicated in a small number of jackknife repeats (Supplementary Figure 3b).

Supplementary Figure 3. Jack-knife analysis for **A:** altruistic decisions and **B:** strategic decisions. Number of jack-knife runs the voxel was significant in, n = 12 so maximum replicability.

Supplementary Table 4.
Peak activations where strategic > selfish from specific contrasts (studies with a selfish control only)

Peak label	BA*	coordinates			SDM-Z	Vox
		x	y	z		
Strategic > selfish						
L occipital pole		-14	-92	18	4.47	77
R frontal pole		38	44	2	4.60	71
R middle temporal gyrus	37	58	-54	-10	4.17	48
L striatum		-6	8	-8	5.14	34
L occipital pole	18	-26	-96	-12	4.77	34
R frontal operculum cortex		32	26	10	4.58	16
Brain stem		-2	-30	-22	4.39	17
Occipital pole	18	-2	-98	14	3.91	17
L striatum		-10	20	-8	4.45	15
Supracalcarine cortex	18	-2	-74	18	3.78	14
L precuneus cortex		-18	-68	32	4.20	13
R cerebellum hemispheric lobule IV / V	19	10	-48	-16	4.40	12
R inferior lateral occipital cortex	18	30	-88	-4	4.02	12
L amygdala		-12	-12	-14	3.74	11
L inferior lateral occipital cortex	19	-46	-80	-2	4.34	10
R inferior lateral occipital cortex	19	50	-72	-12	4.20	10
Selfish > strategic						
R precentral gyrus		8	-16	52	4.59	250
R parietal operculum cortex		40	-24	24	5.21	92
R parietal operculum cortex	48	52	-28	18	3.19	81
L superior temporal gyrus	22	-54	-12	-8	3.81	53
R middle temporal gyrus	21	60	-40	-2	3.56	53
R postcentral gyrus	4	22	-34	66	3.50	36
R precentral gyrus	4	16	-26	70	2.91	32
L middle temporal gyrus	21	-58	-44	0	3.09	31
R insula	48	36	-12	6	2.95	21
L middle temporal gyrus	20	-58	-26	-14	2.97	19
R middle temporal gyrus	20	58	-14	-18	2.85	18
L angular gyrus	22	-60	-60	22	3.04	17
R cerebellum, crus II		30	-82	-36	3.32	15
L superior temporal gyrus		-50	-20	-4	3.10	15
R inferior temporal gyrus	20	44	-22	-22	3.42	15
R insula		34	-8	18	3.51	13
R postcentral gyrus	4	50	-12	38	2.90	13
L angular gyrus	39	-40	-56	22	2.95	13
R central opercular cortex		46	-8	20	3.10	13
L temporal pole	21	-50	8	-26	3.40	12
L hippocampus		-18	-40	6	2.82	11

Note. L = left, R = right, BA = Brodmann area, coordinates (x, y, z) in MNI space, SDM-Z= SDM z-value of activation, Vox = number of voxels in the cluster. *If cluster falls within *more* than one BA, none is reported either by AES:SDM software, or here.

2.1.2. *Comparisons and overlap*

We found overlap between altruistic and strategic decisions in the specific contrasts in left NuAcc for prosocial > selfish and a small area of parietal operculum cortex was common for selfish > prosocial (Supplementary Figure 4a).

Like the modelled contrast, altruistic decisions showed greater activation in sgACC and left TPJ and additionally in this analysis, right pSTS and left lateral occipital cortex. Strategic decisions showed greater activation in the right caudate nucleus, right frontal pole, right dlPFC (area 46) and left posterior insula than altruistic decisions. Several sensorimotor regions, including the fusiform face area, also showed more activity during strategic decisions (Supplementary Figure 4b & Supplementary Table 5).
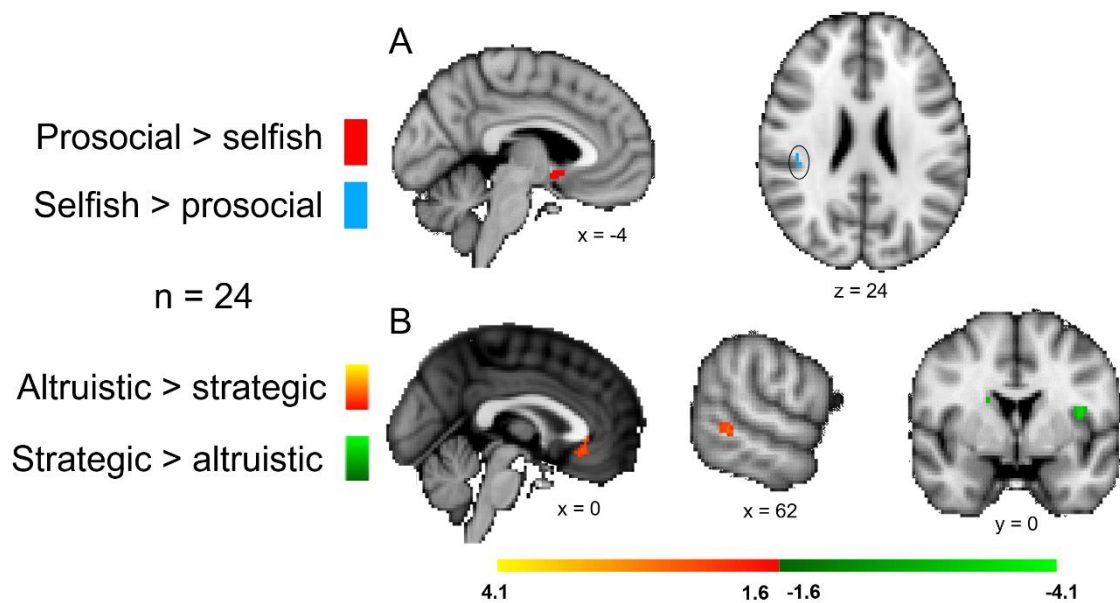
Supplementary Figure 4. **A:** overlap and **B:** comparisons between altruistic and strategic decisions using specific contrasts; only those with a selfish control (thresholded with permutation analysis run in AES:SDM). Coronal images in radiological orientation (right = left)

Supplementary Table 5.

Regions showing significantly greater activation in altruistic or strategic studies in the specific contrasts (studies with a selfish control only)

Peak label	BA*	coordinates			SDM-Z	Vox
		x	y	z		
Altruistic > strategic						
R middle temporal gyrus	21	66	-44	0	2.88	58
Anterior cingulate gyrus		-6	34	0	2.97	43
L superior lateral occipital cortex	39	-46	-74	42	2.97	29
L angular gyrus	22	-60	-58	24	2.76	22
Strategic > altruistic						
Intracalcarine cortex		-6	-74	16	3.79	102
L precentral gyrus	6	-42	-8	62	4.38	78
L postcentral gyrus	3	-32	-38	58	3.81	64
L central opercular cortex		-50	-18	20	3.59	55
L central opercular / insular cortex	48	-40	-4	12	3.91	49
L superior lateral occipital cortex	18	-20	-82	32	3.52	48
R frontal pole	10	28	52	10	3.50	35
L occipital pole	18	-16	-92	20	3.51	24
L inferior lateral occipital cortex		-42	-64	6	3.63	21
L lingual gyrus		-18	-68	-2	3.59	17
L precentral gyrus	6	-28	-22	62	3.17	15
L superior parietal lobule	5	-22	-52	72	3.38	13
R caudate		18	6	20	3.33	13
L occipital fusiform gyrus		-30	-78	-8	3.31	13
R inferior lateral occipital cortex		34	-76	10	3.39	12

Note. L = left, R = right, BA = Brodmann area, coordinates (x, y, z) in MNI space, SDM-Z= SDM z-value of activation, Vox = number of voxels in the cluster. *If cluster falls within *more* than one BA, none is reported either by AES:SDM software, or here.

2.2. Additional ROI results

As with the results at x = 4 reported in the main text, linear models on vmPFC voxel effect sizes at x = 0 and x = -4 showed interactions between vmPFC location and decision group. Activation for altruistic decisions was greater in posterior voxels whereas strategic activation was greater in anterior regions (Supplementary Figure 5). Models were run using (i) all the data, including zero values and (ii) only the studies with data in that voxel, on a voxel-by-voxel basis. In all cases, a mixed-effects model outperformed a fixed-effects model.

At x = 0, both models showed significant interactions, (i) $t(22)=2.27$, $p=.032$; (ii) $t(20.8)=2.83$, $p=.010$. For x = -4, the model only including studies with data present (ii) showed a significant interaction, $t(22.2)=2.42$, $p=.024$ but when all data was included (i), this only reached trend significance, $t(22)=1.99$, $p=.058$.
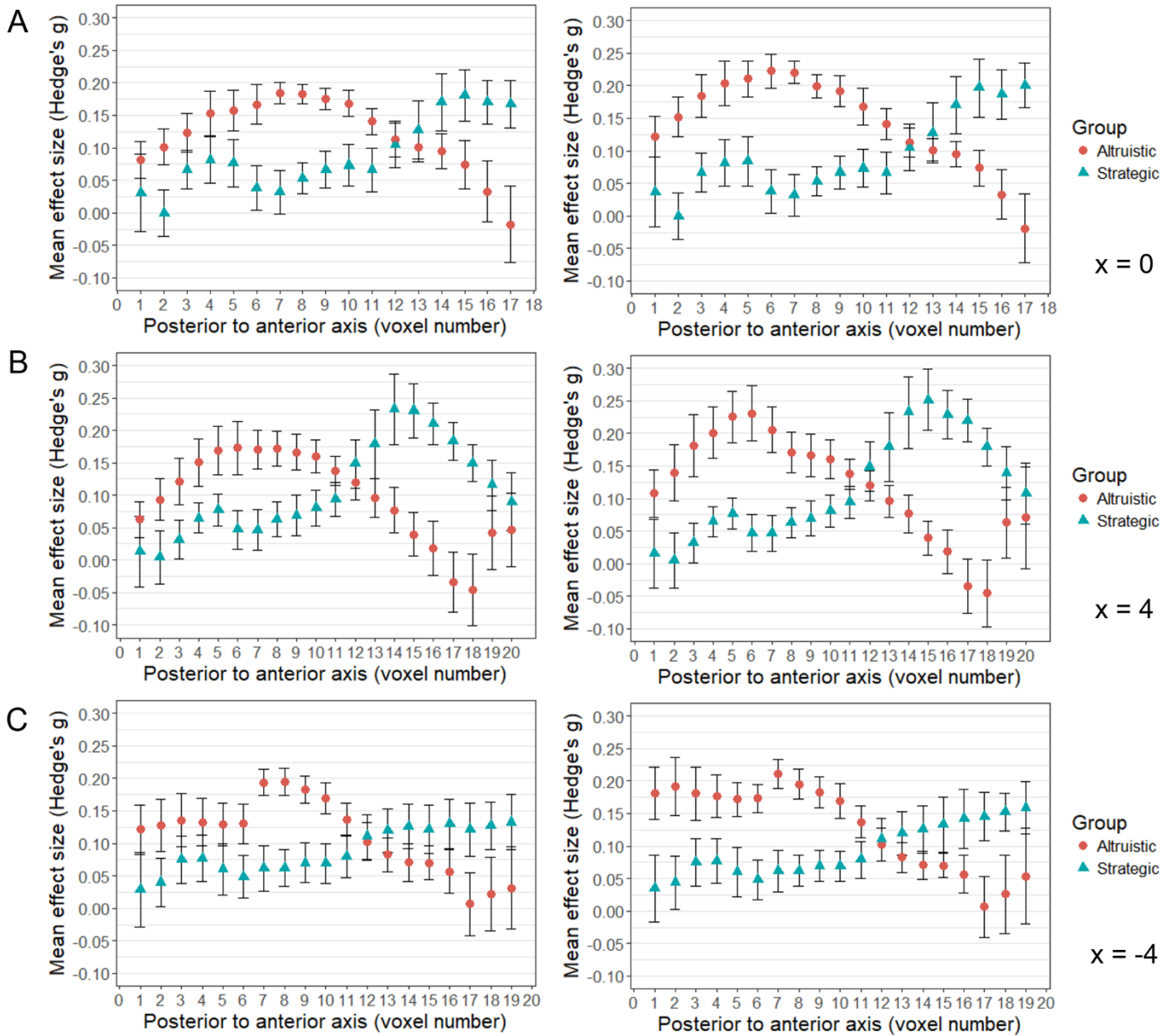
Supplementary Figure 5. vmPFC ROI results for **A:** x = 0, **B:** x = 4 and **C:** x = -4. Left column: using all the data, including zero values, right column: using only the studies with data in that voxel, on a voxel-by-voxel basis. Error bars are standard error.

3. Papers eligible for the meta-analysis

Bault, N., Pelloux, B., Fahrenfort, J.J., Ridderinkhof, K.R., van Winden, F., 2015. Neural dynamics of social tie formation in economic decision-making. Soc. Cogn. Affect. Neurosci. 10, 877–884. doi:10.1093/scan/nsu138

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., Fehr, E., 2008. Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron 58, 639–650. doi:10.1016/j.neuron.2008.04.009

Brosch, T., Coppin, G., Scherer, K.R., Schwartz, S., Sander, D., 2011. Generating value(s): Psychological value hierarchies reflect context-dependent sensitivity of the reward system. Soc. Neurosci. 6, 198–208. doi:10.1080/17470919.2010.506754

Brunnlieb, C., Nave, G., Camerer, C.F., Schosser, S., Vogt, B., Münte, T.F., Heldmann, M., 2016. Vasopressin increases human risky cooperative behavior. Proc. Natl. Acad. Sci. 113, 2051–2056. doi:10.1073/pnas.1518825113

Chen, X., Hackett, P.D., DeMarco, A.C., Feng, C., Stair, S., Haroon, E., Ditzen, B., Pagnoni, G., Rilling, J.K., 2016. Effects of oxytocin and vasopressin on the neural response to unreciprocated cooperation within brain regions involved in stress and anxiety in men and women. Brain Imaging Behav. 581–593. doi:10.1007/s11682-015-9411-7

Chiu, P.H., Kayali, M.A., Kishida, K.T., Tomlin, D., Klinger, L.G., Klinger, M.R., Montague, P.R., 2008. Self responses along cingulate cortex reveal quantitative neural phenotype for high-functioning autism. Neuron 57, 463–473. doi:10.1016/j.neuron.2007.12.020

Delgado, M.R., Frank, R.H., Phelps, E.A., 2005. Perceptions of moral character modulate the neural systems of reward during the trust game. Nat. Neurosci. 8, 1611–1618. doi:10.1038/nn1575

Dulebohn, J., Davison, R., Lee, S., Conlon, D., McNamara, G., Sarinopoulos, I., 2015. Gender differences in justice evaluations: Evidence from fMRI. J Appl Psychol 101, 151–170. doi:http://dx.doi.org/10.1037/apl0000048

Emonds, G., Declerck, C.H., Boone, C., Seurinck, R., Achten, R., 2014. Establishing cooperation in a mixed-motive social dilemma. An fMRI study investigating the role of social value orientation and dispositional trust. Soc. Neurosci. 9, 10–22. doi:10.1080/17470919.2013.858080

Emonds, G., Declerck, C.H., Boone, C., Vandervliet, E.J.M., Parizel, P.M., 2012. The cognitive demands on cooperation in social dilemmas: An fMRI study. Soc. Neurosci. 7, 494–509. doi:10.1080/17470919.2012.655426

Emonds, G., Declerck, C.H., Boone, C., Vandervliet, E.J.M., Parizel, P.M., 2011. Comparing the neural basis of decision making in social dilemmas of people with different social value orientations, an fMRI study. J. Neurosci. Psychol. Econ. 4, 11–24. doi:10.1037/a0020151

Fahrenfort, J.J., van Winden, F., Pelloux, B., Stallen, M., Ridderinkhof, K.R., 2012. Neural correlates of dynamically evolving interpersonal ties predict prosocial behavior. Front. Neurosci. 6, 1–14. doi:10.3389/fnins.2012.00028

Fareri, D.S., Chang, L.J., Delgado, M.R., 2012. Effects of direct social experience on trust decisions and neural reward circuitry. Front. Neurosci. 6, 1–17. doi:10.3389/fnins.2012.00148

FeldmanHall, O., Dalgleish, T., Evans, D., Mobbs, D., 2015. Empathic concern drives costly

altruism. Neuroimage 105, 347–356. doi:10.1016/j.neuroimage.2014.10.043

FeldmanHall, O., Dalgleish, T., Mobbs, D., 2013. Alexithymia decreases altruism in real social decisions. Cortex 49, 899–904. doi:10.1016/j.cortex.2012.10.015

FeldmanHall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S., Mobbs, D., 2012. Differential neural circuitry and self-interest in real vs hypothetical moral decisions. Soc. Cogn. Affect. Neurosci. 7, 743–751. doi:10.1093/scan/nss069

Fermin, A.S.R., Sakagami, M., Kiyonari, T., Li, Y., Matsumoto, Y., Yamagishi, T., 2016. Representation of economic preferences in the structure and function of the amygdala and prefrontal cortex. Sci. Rep. 6, 20982. doi:10.1038/srep20982

Fett, A.K.J., Gromann, P.M., Giampietro, V., Shergill, S.S., Krabbendam, L., 2014. Default distrust? An fMRI investigation of the neural development of trust and cooperation. Soc. Cogn. Affect. Neurosci. 9, 395–402. doi:10.1093/scan/nss144

Fouragnan, E., Chierchia, G., Greiner, S., Neveu, R., Avesani, P., Coricelli, G., 2013. Reputational priors magnify striatal responses to violations of trust. J. Neurosci. 33, 3602–3611. doi:10.1523/JNEUROSCI.3086-12.2013

Garbarini, F., Boero, R., D'Agata, F., Bravo, G., Mosso, C., Cauda, F., Duca, S., Geminiani, G., Sacco, K., 2014. Neural correlates of gender differences in reputation building. PLoS One 9. doi:10.1371/journal.pone.0106285

Genevsky, A., Västfjäll, D., Slovic, P., Knutson, B., 2013. Neural underpinnings of the identifiable victim effect: Affect shifts preferences for giving. J. Neurosci. 33, 17188–17196. doi:10.1523/JNEUROSCI.2348-13.2013

Gospic, K., Sundberg, M., Maeder, J., Fransson, P., Petrovic, P., Isacsson, G., Karlström, A., Ingvar, M., 2013. Altruism costs-the cheap signal from amygdala. Soc. Cogn. Affect. Neurosci. 9, 1325–1332. doi:10.1093/scan/nst118

Gradin, V.B., Pérez, A., Macfarlane, J.A., Cavin, I., Waiter, G., Tone, E.B., Dritschel, B., Maiche, A., Steele, J.D., 2016. Neural correlates of social exchanges during the Prisoner's Dilemma game in depression. Psychol. Med. 46, 1289–1300. doi:10.1017/S0033291715002834

Greening, S., Norton, L., Virani, K., Ty, A., Mitchell, D., Finger, E., 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. doi:10.3758/s13415-013-0213-3

Gromann, P.M., Shergill, S.S., de Haan, L., Meewis, D.G.J., Fett, a-K.J., Korver-Nieberg, N., Krabbendam, L., 2014. Reduced brain reward response during cooperation in first-degree relatives of patients with psychosis: An fMRI study. Psychol. Med. 1–10. doi:10.1017/S0033291714000737

Gunther Moor, B., Güroğlu, B., Op de Macks, Z.A., Rombouts, S.A.R.B., van der Molen, M.W., Crone, E.A., 2012. Social exclusion and punishment of excluders: Neural correlates and developmental trajectories. Neuroimage 59, 708–717. doi:10.1016/j.neuroimage.2011.07.028

Güroğlu, B., Will, G.J., Crone, E.A., 2014. Neural correlates of advantageous and disadvantageous inequity in sharing decisions. PLoS One 9, e107996. doi:10.1371/journal.pone.0107996

Harbaugh, W.T., Mayr, U., Burghart, D.R., 2007. Neural responses to taxation and voluntary giving reveal motives for charitable donations. Science (80-.). 316, 1622–1625.

doi:10.1126/science.1140738

Hare, T.A., Camerer, C.F., Knoepfle, D.T., O'Doherty, J.P., Rangel, A., 2010. Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. J. Neurosci. 30, 583–590. doi:10.1523/JNEUROSCI.4089-09.2010

Hein, G., Morishima, Y., Leiberg, S., Sul, S., Fehr, E., 2016. The brains functional network architecture reveals human motives. Science (80-.). 351, 1074–1078. doi:10.1126/science.aac7992

Hein, G., Silani, G., Preuschoff, K., Batson, C.D., Singer, T., 2010. Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. Neuron 68, 149–160. doi:10.1016/j.neuron.2010.09.003

Hutcherson, C.A., Bushong, B., Rangel, A., 2015. A neurocomputational model of altruistic choice and its implications. Neuron 87, 451–462. doi:10.1016/j.neuron.2015.06.031

Izuma, K., Saito, D.N., Sadato, N., 2010. Processing of the incentive for social approval in the ventral striatum during charitable donation. J. Cogn. Neurosci. 22, 621–631. doi:10.1162/jocn.2009.21228

King-Casas, B., Tomlin, D., Anen, C., Camerer, C.F., Quartz, S.R., Montague, P.R., 2005. Getting to know you: Reputation and trust in a two-person economic exchange. Science (80-.). 308, 78–83. doi:10.1126/science.1108062

Kircher, T., Blümel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., van Os, J., Krach, S., 2009. Online mentalising investigated with functional MRI. Neurosci. Lett. 454, 176–181. doi:10.1016/j.neulet.2009.03.026

Klapwijk, E.T., Lelieveld, G.J., Aghajani, M., Boon, A.E., van der Wee, N.J.A., Popma, A., Vermeiren, R.R.J.M., Colins, O.F., 2016. Fairness decisions in response to emotions: A functional MRI study among criminal justice-involved boys with conduct disorder. Soc. Cogn. Affect. Neurosci. 11, 674–682. doi:10.1093/scan/nsv150

Koban, L., Pichon, S., Vuilleumier, P., 2014. Responses of medial and ventrolateral prefrontal cortex to interpersonal conflict for resources. Soc. Cogn. Affect. Neurosci. 9, 561–569. doi:10.1093/scan/nst020

Krach, S., Blümel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., van Os, J., Kircher, T., 2009. Are women better mindreaders? Sex differences in neural correlates of mentalizing detected with functional MRI. BMC Neurosci. 10. doi:10.1186/1471-2202-10-9

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., Kircher, T., 2008. Can machines think? Interaction and perspective taking with robots investigated via fMRI. PLoS One 3, e25. doi:10.1371/journal.pone.0002597

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., Heinecke, A., Grafman, J., 2007. Neural correlates of trust. Proc. Natl. Acad. Sci. U. S. A. 104, 20084–20089. doi:10.1073/pnas.0710103104

Kuss, K., Falk, A., Trautner, P., Elger, C.E., Weber, B., Fliessbach, K., 2013. A reward prediction error for charitable donations reveals outcome orientation of donators. Soc. Cogn. Affect. Neurosci. 8, 216–223. doi:10.1093/scan/nsr088

Kuss, K., Falk, A., Trautner, P., Montag, C., Weber, B., Fliessbach, K., 2015. Neuronal correlates of social decision making are influenced by social value orientation - an fMRI study. Front. Behav. Neurosci. 9, 1–8. doi:10.3389/fnbeh.2015.00040

Lelieveld, G.-J., Van Dijk, E., Güroğlu, B., Van Beest, I., Van Kleef, G. a., Rombouts, S.A.R.B., Crone, E.A., 2013. Behavioral and neural reactions to emotions of others in the distribution of resources. Soc. Neurosci. 8, 52–62. doi:10.1080/17470919.2012.735621

Li, J., Xiao, E., Houser, D., Montague, P.R., 2009. Neural responses to sanction threats in two-party economic exchange. Proc. Natl. Acad. Sci. U. S. A. 106, 16835–40. doi:10.1073/pnas.0908855106

Macoveanu, J., Ramsoy, T.Z., Skov, M., Siebner, H.R., Fosgaard, T.R., 2016. The neural bases of framing effects in social dilemmas. J. Neurosci. Psychol. Econ. 9, 14–28. doi:10.1037/npe0000050

Majdandžić, J., Amashaufer, S., Hummer, A., Windischberger, C., Lamm, C., 2016. The selfless mind: How prefrontal involvement in mentalizing with similar and dissimilar others shapes empathy and prosocial behavior. Cognition 157, 24–38. doi:10.1016/j.cognition.2016.08.003

McCabe, K., Houser, D., Ryan, L., Smith, V., Trouard, T., 2001. A functional imaging study of cooperation in two-person reciprocal exchange. Proc. Natl. Acad. Sci. U. S. A. 98, 11832–11835. doi:10.1073/pnas.211415698

McClure-Tone, E.B., Nawa, N.E., Nelson, E.E., Detloff, A.M., Fromm, S.J., Pine, D.S., Ernst, M., 2011. Preliminary findings: Neural responses to feedback regarding betrayal and cooperation in adolescent anxiety disorders. Dev. Neuropsychol. 36, 453–472. doi:936795781 [pii]\r10.1080/87565641.2010.549876

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., Grafman, J., 2006. Human fronto-mesolimbic networks guide decisions about charitable donation. Proc. Natl. Acad. Sci. U. S. A. 103, 15623–15628. doi:10.1073/pnas.0604475103

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

Phan, K.L., Sripada, C.S., Angstadt, M., McCabe, K., 2010. Reputation for reciprocity engages the brain reward center. Proc. Natl. Acad. Sci. U. S. A. 107, 13099–13104. doi:10.1073/pnas.1008137107

Pulcu, E., Zahn, R., Moll, J., Trotter, P.D., Thomas, E.J., Juhasz, G., Deakin, J.F.W., Anderson, I.M., Sahakian, B.J., Elliott, R., 2014. Enhanced subgenual cingulate response to altruistic decisions in remitted major depressive disorder. NeuroImage Clin. 4, 701–710. doi:10.1016/j.nicl.2014.04.010

Ramsøy, T.Z., Skov, M., Macoveanu, J., Siebner, H.R., Fosgaard, T.R., 2015. Empathy as a neuropsychological heuristic in social decision-making. Soc. Neurosci. 10, 179–191. doi:10.1080/17470919.2014.965341

Riedl, R., Mohr, P.N.C., Kenning, P.H., Davis, F.D., Heekeren, H.R., 2014. Trusting humans and avatars: A brain imaging study based on evolution theory. J. Manag. Inf. Syst. 30, 83–114. doi:10.2753/MIS0742-1222300404

Rilling, J.K., Glenn, A.L., Jairam, M.R., Pagnoni, G., Goldsmith, D.R., Elfenbein, H.A., Lilienfeld, S.O., 2007. Neural correlates of social cooperation and non-cooperation as a function of psychopathy. Biol. Psychiatry 61, 1260–1271. doi:10.1016/j.biopsych.2006.07.021

Rilling, J.K., Goldsmith, D.R., Glenn, A.L., Jairam, M.R., Elfenbein, H.A., Dagenais, J.E., Murdock, C.D., Pagnoni, G., 2008. The neural correlates of the affective response to unreciprocated cooperation. Neuropsychologia 46, 1256–1266.

doi:10.1016/j.neuropsychologia.2007.11.033

Rilling, J.K., Gutman, D.A., Zeh, T.R., Pagnoni, G., Berns, G.S., Kilts, C.D., 2002. A neural basis for social cooperation. Neuron 35, 395–405. doi:10.1016/S0896-6273(02)00755-9

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

Sakaiya, S., Shiraito, Y., Kato, J., Ide, H., Okada, K., Takano, K., Kansaku, K., 2013. Neural correlate of human reciprocity in social interactions. Front. Neurosci. 7, 1–12. doi:10.3389/fnins.2013.00239

Schneider-Hassloff, H., Straube, B., Nuscheler, B., Wemken, G., Kircher, T., 2015. Adult attachment style modulates neural responses in a mentalizing task. Neuroscience 303, 462–473. doi:10.1016/j.neuroscience.2015.06.062

Sharp, C., Burton, P.C., Ha, C., 2011. "Better the devil you know": A preliminary study of the differential modulating effects of reputation on reward processing for boys with and without externalizing behavior problems. Eur. Child Adolesc. Psychiatry 20, 581–592. doi:10.1007/s00787-011-0225-x

Smith-Collins, A.P.R., Fiorentini, C., Kessler, E., Boyd, H., Roberts, F., Skuse, D.H., 2013. Specific neural correlates of successful learning and adaptation during social exchanges. Soc. Cogn. Affect. Neurosci. 8, 887–897. doi:10.1093/scan/nss079

Sripada, C., Angstadt, M., Banks, S., 2009. Functional neuroimaging of mentalizing during the trust game in social anxiety disorder. Neuroreport 20, 984–989. doi:10.1097/WNR.0b013e32832d0a67

Sripada, C., Angstadt, M., Liberzon, I., McCabe, K., Phan, K.L., 2013. Aberrant reward center response to partner reputation during a social exchange game in generalized social phobia. Depress. Anxiety 30, 353–361. doi:10.1002/da.22091

Stanley, D.A., Sokol-Hessner, P., Fareri, D.S., Perino, M.T., Delgado, M.R., Banaji, M.R., Phelps, E.A., 2012. Race and reputation: Perceived racial group trustworthiness influences the neural correlates of trust decisions. Philos. Trans. R. Soc. B Biol. Sci. 367, 744–753. doi:10.1098/rstb.2011.0300

Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I.I., Tobler, P.N., Kalenscher, T., 2015. Social discounting involves modulation of neural value signals by temporoparietal junction. Proc. Natl. Acad. Sci. USA 112, 1619–1624. doi:10.1073/pnas.1414715112

Suzuki, S., Niki, K., Fujisaki, S., Akiyama, E., 2011. Neural basis of conditional cooperation. Soc. Cogn. Affect. Neurosci. 6, 338–347. doi:10.1093/scan/nsq042

Telzer, E.H., Fuligni, A.J., Lieberman, M.D., Galván, A., 2013. Ventral striatum activation to prosocial rewards predicts longitudinal declines in adolescent risk taking. Dev. Cogn. Neurosci. 3, 45–52. doi:10.1016/j.dcn.2012.08.004

Telzer, E.H., Ichien, N., Qu, Y., 2015. The ties that bind: Group membership shapes the neural correlates of in-group favoritism. Neuroimage 115, 42–51. doi:10.1016/j.neuroimage.2015.04.035

Telzer, E.H., Masten, C.L., Berkman, E.T., Lieberman, M.D., Andrew, J., 2010. Gaining while giving: An fMRI study of the rewards of family assistance among White and Latino youth. Soc. Neurosci. 5, 508–518. doi:10.1080/17470911003687913.Gaining

Telzer, E.H., Masten, C.L., Berkman, E.T., Lieberman, M.D., Fuligni, A.J., 2011. Neural regions associated with self control and mentalizing are recruited during prosocial behaviors towards the family. Neuroimage 58, 242–249. doi:10.1016/j.neuroimage.2011.06.013

Tomlin, D., Kayali, M.A., King-Casas, B., Anen, C., Camerer, C.F., Quartz, S.R., Montague, P.R., 2006. Agent-specific responses in the cingulate cortex during economic exchanges. Science (80-.). 312, 1047–1050. doi:10.1126/science.1125596

Tusche, A., Bockler, A., Kanske, P., Trautwein, F.-M., Singer, T., 2016. Decoding the charitable brain: Empathy, perspective taking, and attention shifts differentially predict altruistic giving. J. Neurosci. 36, 4719–4732. doi:10.1523/JNEUROSCI.3392-15.2016

Van Hoorn, J., Van Dijk, E., Güroğlu, B., Crone, E.A., 2016. Neural correlates of prosocial peer influence on public goods game donations during adolescence. Soc. Cogn. Affect. Neurosci. nsw013. doi:10.1093/scan/nsw013

Wardle, M.C., Fitzgerald, D.A., Angstadt, M., Sripada, C.S., McCabe, K., Luan Phan, K., 2013. The caudate signals bad reputation during trust decisions. PLoS One 8, 1–9. doi:10.1371/journal.pone.0068884

Watanabe, T., Takezawa, M., Nakawake, Y., Kunimatsu, A., Yamasue, H., Nakamura, M., Miyashita, Y., Masuda, N., 2014. Two distinct neural mechanisms underlying indirect reciprocity. Proc. Natl. Acad. Sci. U. S. A. 111, 3990–3995. doi:10.1073/pnas.1318570111

Weiland, S., Hewig, J., Hecht, H., Mussel, P., Miltner, W.H.R., 2012. Neural correlates of fair behavior in interpersonal bargaining. Soc. Neurosci. 7, 537–551. doi:10.1080/17470919.2012.674056

Will, G.-J., Crone, E.A., van Lier, P.A.C., Güroğlu, B., 2016a. Neural correlates of retaliatory and prosocial reactions to social exclusion: Associations with chronic peer rejection. Dev. Cogn. Neurosci. 19, 288–297. doi:10.1016/j.dcn.2016.05.004

Will, G.-J., Crone, E.A., van Lier, P.A.C., Güroğlu, B., 2016b. Longitudinal links between childhood peer acceptance and the neural correlates of sharing. Dev. Sci. 1–13. doi:10.1111/desc.12489

Zaki, J., Mitchell, J.P., 2011. Equitable decision making is associated with neural markers of intrinsic value. Proc. Natl. Acad. Sci. 108, 19761–19766. doi:10.1073/pnas.1112324108

Zheng, H., Zhu, L., 2013. Neural mechanism of proposer's decision-making in the ultimatum and dictator games. Neural Regen. Res. 8, 357–62. doi:10.3969/j.issn.1673-5374.2013.04.008

4. Supplementary references

Amodio, D.M., Frith, C.D., 2006. Meeting of minds: The medial frontal cortex and social cognition. Nat. Rev. 7, 268–277. doi:10.1038/nrn1884

Bartra, O., McGuire, J.T., Kable, J.W., 2013. The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage 76, 412–427. doi:10.1016/j.neuroimage.2013.02.063

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67. doi:10.18637/jss.v067.i01

Campbell-Meiklejohn, D., Simonsen, A., Frith, C.D., Daw, N.D., 2016. Independent neural computation of value from the confidence of others. J. Neurosci. 37, 4490–15. doi:10.1523/JNEUROSCI.4490-15.2016

Izuma, K., Saito, D.N., Sadato, N., 2010. Processing of the incentive for social approval in the ventral striatum during charitable donation. J. Cogn. Neurosci. 22, 621–631. doi:10.1162/jocn.2009.21228

Levy, D.J., Glimcher, P.W., 2012. The root of all value: A neural common currency for choice. Curr. Opin. Neurobiol. 22, 1027–1038. doi:10.1016/j.conb.2012.06.001

Mackey, S., Petrides, M., 2010. Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains. Eur. J. Neurosci. 32, 1940–1950. doi:10.1111/j.1460-9568.2010.07465.x

Öngür, D., Ferry, A.T., Price, J.L., 2003. Architectonic subdivision of the human orbital and medial prefrontal cortex. J. Comp. Neurol. 460, 425–449. doi:10.1002/cne.10609

Radua, J., Mataix-Cols, D., Phillips, M.L., El-Hage, W., Kronhaus, D.M., Cardoner, N., Surguladze, S., 2012. A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. Eur. Psychiatry 27, 605–611. doi:10.1016/j.eurpsy.2011.04.001

Weiland, S., Hewig, J., Hecht, H., Mussel, P., Miltner, W.H.R., 2012. Neural correlates of fair behavior in interpersonal bargaining. Soc. Neurosci. 7, 537–551. doi:10.1080/17470919.2012.674056