Prosocial apathy when helping others is too much effort

Patricia L. Lockwood[1*], Mathilde Hamonet[1], Samuel H. Zhang, Anya Ratnavel[1], Florentine U. Salmony[1], Masud Husain[1,2]^, Matthew A. J. Apps[1]^

[1] Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, United Kingdom

[2] Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford OX3 9DU, United Kingdom

*Corresponding author. Department of Experimental Psychology, 9 South Parks Road University of Oxford, Oxford, OX1 3UD, United Kingdom, patricia.lockwood@psy.ox.ac.uk

^Indicates equal senior contribution

Summary

Prosocial acts – those that are costly to ourselves but benefit others – are a central component of human co-existence[1–3]. While the financial and moral costs of prosocial behaviours are well understood[4–6], everyday prosocial acts do not typically come at such costs. Instead, they require effort. Here, using computational modelling of an effort-based task we show that people are prosocially apathetic. They are less willing to choose to initiate highly effortful acts that benefit others compared to benefitting themselves. Moreover, even when choosing to initiate effortful prosocial acts, people show superficiality, exerting less force into actions that benefit others than themselves. These findings replicated, were present when the other was anonymous or not, and when choices were made to earn rewards or avoid losses. Importantly, the least prosocially motivated people had higher subclinical levels of psychopathy and social apathy. Thus, although people sometimes 'help out', they are less motivated to benefit others and sometimes 'superficially prosocial', which may characterise everyday prosociality and its disruption in social disorders.

The willingness to be prosocial – to perform acts that benefit others – is a central component of human social and moral behaviour. Prosocial behaviours are argued to be a key driver of bonding within groups and facilitate social cohesion[1–5,7]. Acts of prosociality are often reduced in those with criminal levels of psychopathy, in healthy individuals lower in empathy, and after brain damage[4,7–13]. Prosociality has commonly been investigated in terms of people's willingness to benefit others when these acts come at a personal cost[2,4,6,14]. Typically these costs have been financial, such as when donating to other individuals or charities,[6] or moral, such as deciding whether to profit from others' harm[4,5]. Yet in day-to-day life financial and moral costs are often negligible. Instead, prosocial acts require motivation to exert effort. Here, we use computational modeling to precisely characterise how effort costs influence motivation to benefit ourselves and other people.

Theoretical accounts of motivation highlight that there are at least two key components that define the willingness to exert effort. First, prior to an action being performed, its costs and benefits are evaluated[15–20]. When rewards are high, or effort costs low, the value of a behavior will be high and people will be motivated to choose to perform that action. However, when rewards are low or the effort required is high, the subjective value of reward is reduced. Recent accounts suggest that such 'devaluation' or 'discounting' of rewards by effort can be quantified using computational modelling approaches[17,21–27]. Using such models, it is possible to precisely characterise the form that the devaluation takes, and the extent to which each individual devalues rewards by effort costs. Second, once chosen, actions must be energised to such a degree that they are sufficiently forceful[28–30], in order to obtain the desired outcome. While this framework has been used extensively to provide quantifiable measures of motivation to reward oneself, surprisingly such models have not been applied to understand how we motivate actions that benefit others.

Are people prosocial or are they more selfish when it comes to putting in effort for others? As far as we are aware there are no studies that have quantified either willingness or force exerted when people have to put in physical effort for others' benefit. However, results of studies using economic games are often consistent with a largely selfish view of human behaviour[6,31]. While individuals are willing to cooperate and share financial resources with others, they still value their own financial gains above those of other people[6,31,32]. In contrast, studies of moral decision-making have painted a more positive picture of human behaviour. When individuals trade-off a financial cost to themselves against a number of painful shocks delivered to them or to an anonymous stranger, they appear to be 'hyperaltruistic'[4,5]. They are willing to pay more money to avoid painful shocks being delivered to another person, than they are to themselves. Thus, if prosocial effort costs are like moral choices, people might be *more* willing to exert effort to benefit others over themselves and therefore should be *hyper-motivated* to benefit others. On the other hand, if effort costs are treated similarly to financial costs, people may be prosocially apathetic and far less willing to work to benefit others.

Levels of self-motivation and prosocial behaviours vary considerably between people. Self-motivation, in terms of the willingness to exert effort for oneself, has been shown to relate to state and trait levels of apathy[17,33]. Apathy is as a reduction in the willingness to exert goal directed behaviours[34]. Multiple lines of evidence are emerging that as well as being common in several neurological and psychiatric disorders, apathy is also present to varying degrees in the general population[17,35–39], and can be reliably measured as a dimensional construct[35]. This dimensional approach to understanding psychiatric and clinical disorders is consistent with the recent RDoC approach to uncover the mechanisms of psychopathology in health and disease[40].

Individuals who have high levels of behavioural activation apathy are less likely to exert effort to benefit themselves[17,33]. However, apathy is not a unitary construct, and recent evidence demonstrates that another key component of lack of motivation can relate to social apathy[35]. Those who are more socially apathetic report less engagement in social behaviours[35]. Here we use the term "prosocial apathy" to refer to reduced motivation to put in effort for others benefit, relative to one's own benefit. Whether individuals who are more socially apathetic also show reduced willingness to exert prosocial acts and to put in effort to benefit others is currently unknown. Outside the domain of motivation, moral and financial prosocial choices often vary considerably between individuals[4,7,41,42]; and levels of psychopathic traits are negatively correlated with trait measures of prosociality[43] and self-reported enjoyment of prosocial interactions[44]. However, the significance of such trait level associations for behavior is unclear. For these reasons, we also tested whether those high in psychopathic traits or social apathy are less willing to exert effort to benefit others.

To examine whether people are hyper-motivated or apathetic when choosing or energising acts that benefit other individuals, we designed an effort-based decision-making task based on rodent, primate and human research[15,45,46]. Across two studies, participants made choices about whether they were willing to exert an offered level of effort for rewards and subsequently executed the chosen effort (**Fig.1**). To precisely quantify effort, we used different levels of grip-force, with the levels of effort scaled as a percentage of each individual's maximal voluntary contraction (MVC). Participants made choices between a baseline no effort (0% of MVC), low reward option (1 credit) and a variable 'offer', that was higher in reward (2-10 credits) but also higher in effort (30-70%). We also examined choices and force exerted when avoiding losing rewards (2-10 credits) if they chose the 'offer'.

Once they made their choice, participants were required to squeeze the handle to the required degree of force. Crucially, on half of the trials they made choices in order to earn payment for themselves (Self condition) and on the other half of trials they had the option to put in effort to reward another person (Other condition). Using computational modelling approaches we quantified how willing people were to choose to exert effort to benefit themselves versus others and the mathematical function that characterized the discounting of reward. We also examined how energised people's actions were by measuring the force exerted at each level of effort. We could therefore test our key question of whether effort is costlier when benefitting others or ourselves and whether trait levels of apathy and psychopathy explain individual differences in prosocial apathy.

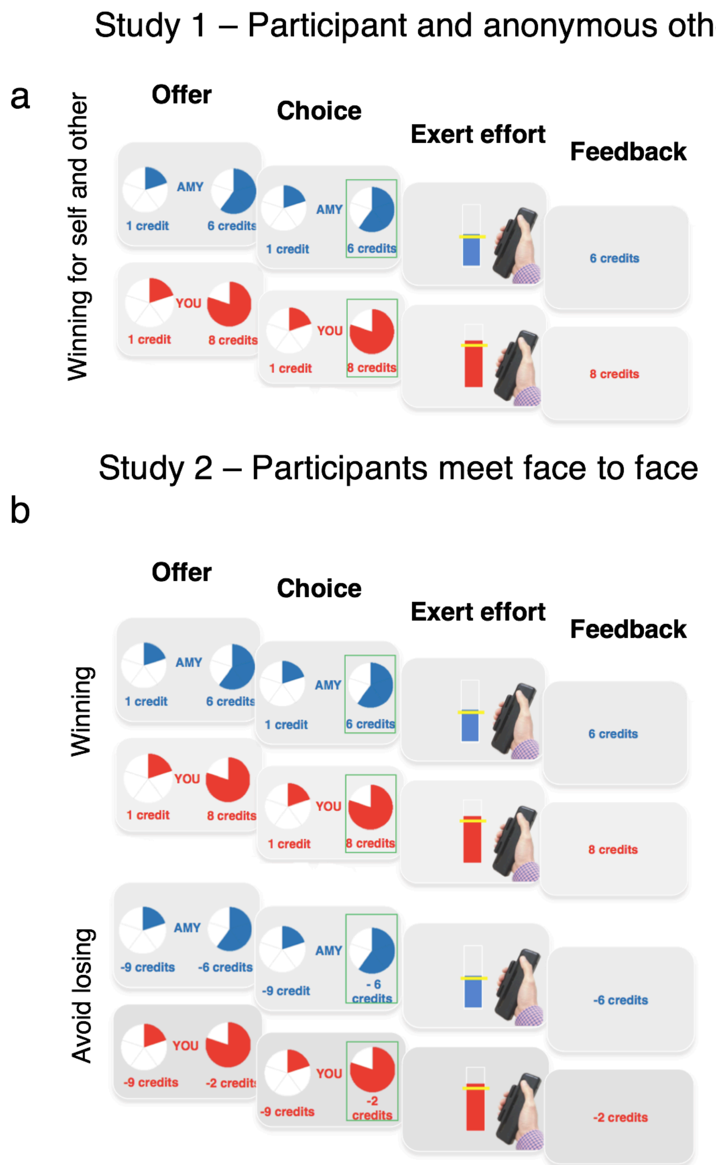

Figure 1. Prosocial Motivation Measure (PMM) for Self vs. Other

a. In Study 1 participants were assigned to two different roles in the experiment completely anonymously. They made choices between a fixed low effort, low reward option (shown on left) or a variable higher effort (30%, 40%, 50%, 60%, 70% maximum voluntary contraction, MVC) but higher reward option (2-10 credits). Once a choice was made the chosen force was exerted on a handheld dynamometer in the participant's dominant hand (Exert effort). Only if the required level was reached for 1s out of a 3s window would the offered reward be obtained, otherwise 0 credits would be delivered. On half the trials, the credits on offer were for themselves (Self: shown in red). On the other half of trials, credits were for the other participant (Other: blue). **b.** In Study 2 participants met face-to-face with a confederate (whom they believed to be another participant taking part in the same study). Participants took part in two sessions in counterbalanced order. In one session, they made choices *to win money* for the other person and themselves; in another session, they made choices *to avoid losing money* for the other person and themselves.

Using ANOVAs on proportion of acceptance of the ‘variable offer’, we first examined whether the willingness to choose to put in effort to obtain rewards differs by agency (Self or anonymous Other) of the beneficiary. In *Study 1*, there was a significant Agent x Effort level interaction ($F(4,188) = 11.21$, $p<.001$; **Fig 2A**), as well as main effects of Effort and Agent (Effort: $F(4,188) = 128.01$, $p<.001$; Agent: $F(1,47) = 54.93$, $p<.001$). Thus, as the offer increased in *effort,* people were less willing to choose it. However, this effect was further exaggerated when another person was the beneficiary, with a greater reduction in the willingness to exert higher levels of effort for the other person. Therefore, when the costs are low, people are willing to choose to benefit both themselves and others. However, when the costs are high, people are less willing to put in effort to help others than themselves. Are people less incentivized by *rewards* for others? We found a significant main effect of Reward ($F(4,188) = 52.05$, $p<.001$) and Agent ($F(1,47) = 54.93$, $p<.001$) and an interaction between Agent and Reward ($F(4,188) = 6.34$, $p<.001$). Thus, as rewards increased, people were more incentivized for self versus other (**Fig 2A**).

In Study 2 we aimed to replicate the findings of Study 1 and also examine whether prosocial motivation could be increased by changing the nature of the other person and the rewarding incentives. Previous studies suggest that prosocial behaviour can be affected by whether this is public or private[43,47,48]. For example, the introduction of an audience increases levels of prosociality[43,47,48], while anonymity reduces it[49]. Therefore, we examined whether face-to-face contact with the person who would receive the outcomes would increase the effort exerted for others. In addition, we also included an *avoid loss* condition as studies of moral harm have often used negative valence to the benefit on offer (e.g. losing money to avoid harm[4,5]). Moreover, previous research has suggested that self-motivation can be increased when the aim is to avoid the loss of rewards, rather than trying to earn them[50,51]. Thus, Study 2 allowed us to examine three questions. First, do the effects of effort replicate when minimal contact is made with another person whom the participant is led to believe is the ‘Other’ person? Second, is there a loss aversion effect that results in higher levels of prosocial motivation? And finally, is there still an increased willingness to exert higher levels of effort when avoiding losses?

First, examining the win session only (**Fig 2B**), we replicated the findings of Study 1 for effort (Agent x Effort interaction ($F(4,176)=18.02$, $p<.001$) but not for reward, (Agent x Reward interaction $F(4,176)=2.22$, $p=.069$). This suggests that even when participants met the person who would be receiving the outcomes, their prosocial motivation did not increase in terms of the effort they were willing to expend. Moreover, in this analysis, effort levels were collapsed across reward and the variable offer was always associated with an effort cost. Thus, the interaction between agent and effort across both studies suggests that it is a reduced willingness to exert *effort* for others that consistently drives reduced choices of 'offers' for others and not only a reduced sensitivity to rewards that others will receive.

Analysis of behaviour from the losing session (**Fig 2C**) showed that framing the task as avoiding losses for self and other did not change motivation. Again, participants were less willing to choose to put in effort to avoid losing money for the other person compared to themselves at increasing levels of effort (Agent x Effort interaction ($F(4,176)=13.59$, $p<.001$)). They also demonstrated significantly lower loss sensitivity for others compared to themselves (Agent x Reward interaction ($F(4,176)=3.10$ $p<.017$)). These findings were also replicated when including winning and losing in the same ANOVA model (See Supplementary Results). Comparing choice behaviour between study 1 and 2, we did not find a significant interaction between sample x agent x effort ($F(4,364)= 2.13$ $p=.08$).

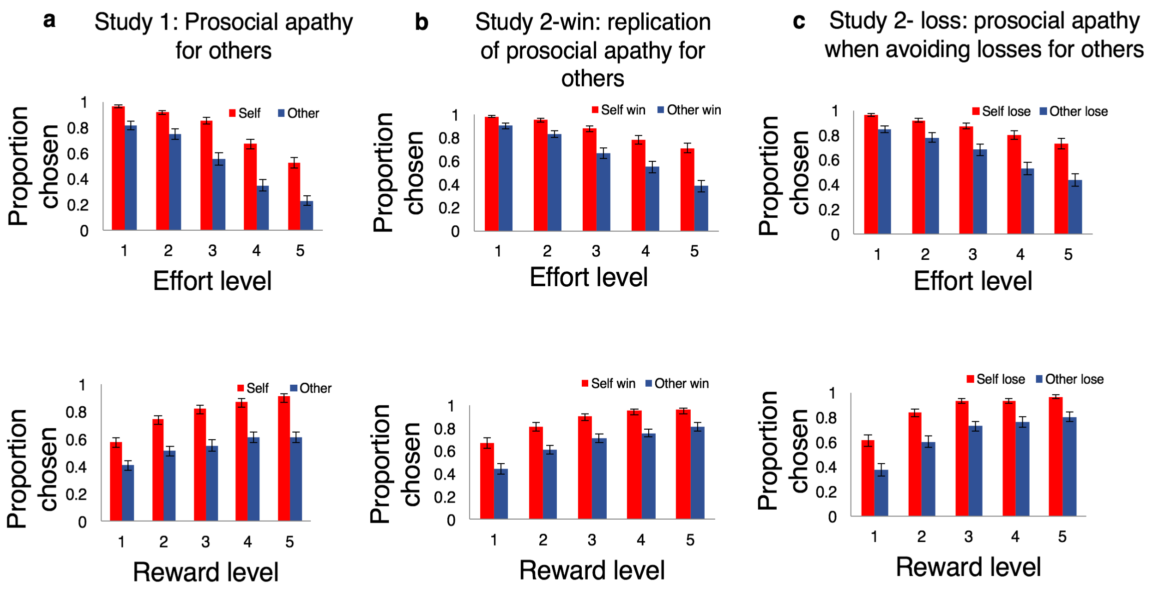

Figure 2. Prosocial apathy when deciding to exert effort to reward others.

A. Proportion of higher effort/higher reward option chosen over baseline option (lower reward, lower effort) plotted against effort (top panel) and reward levels (bottom panels) in *Study 1 (n=48)*. Effort levels 1-5 correspond to 30, 40, 50, 60 and 70% of a participant's maximum voluntary contraction (MVC). Reward levels 1-5 correspond to 2, 4, 6, 8 and 10 credits. Participants chose the high effort/higher reward option more frequently for self (red bars) than other (blue bars), with this difference increasing with effort level. They were also less reward sensitive for other people than themselves. Error bars show S.E.M. **B.** In *Study 2 (n=45)*, people again chose the higher effort/higher reward option more frequently for self than other, with this effect increasing with effort level (top panel). Participants were also less reward sensitive for others compared to self (bottom panel). Error bars show S.E.M. **C.** When making choices to avoid losing for self or other in *Study 2*, participants chose the higher effort/lower loss option more frequently for self than other (top) and were less loss sensitive for self than other. Error bars show S.E.M.

Next, to precisely quantify the *subjective* influence of effort on rewards for self and other we created a range of computational models that characterised how rewards were being devalued by the amount of effort. This approach allowed us to quantify people's motivation to put in effort to reward others using a single parameter, that meaningfully characterizes how motivation is influenced by the balance between effort and reward. The model also allowed us to use this parameter to test for associations between discounting rates and self-report measures, which can provide more interpretable results than using individual difference measures as covariates in ANOVAs that contain multiple factors, with multiple levels.

Models were fitted to behavioural data using the softmax function (See Materials and Methods and Supplementary Methods). Each model contained idiosyncratic parameters characterizing (‘*K*’) the degree to which a reward was devalued by effort, and ‘noise’ parameters characterizing the stochasticity of choices (β). There were two features that were varied to create the model space. First, we varied the mathematical function that characterises the form of the discounting (i.e., whether rewards are devalued linearly, hyperbolically, or parabolically by physical effort[22,52]). Second, we compared models which tested whether people devalue or ‘discount’ rewards by effort to the same degree on self or other trials, or instead used separate discounting rates. We therefore created two classes of models that had either the same parameters to characterise discounting (*K*) on self and other trials (models 1-6) or separate ones (models 7-12, **Fig 3.**). Within these models, we tested a further two classes of models that characterized whether separate parameters for levels of noise (β, softmax) (models 4-6, 10-12), or single parameters for noise (models 1-3, 7-9) best explained behaviour.

The winning model in Study 1 was a parabolic model in which separate parameters characterised the devaluation of rewards for self and other trials, but had a single noise parameter (model 7, **Fig 3A**). This model was best able to explain participants’ choices when correcting for the number of parameters in the models. The Bayes Information Criterion (BIC) scores of this model were very close to a parabolic model that had separate discounting parameters for self and other but also separate parameters for noise (BIC two discount, one noise parameter 5591.69; BIC separate *Kself* and *Kother*, one noise parameter = 5574.28; **Fig 3A,D**). However, choice behavior of the majority of participants (32 out of 48) was explained better by the model with a single noise parameter. Thus, devaluation of rewards by effort for self and other can be characterised in the following manner:

$$SV = R - KE^2$$

$$K = \begin{cases} Kself \text{ if self trial} \\ Kother \text{ if other trial} \end{cases}$$

In this model, SV is the subjective net-value of a variable offer of given Effort (E) and Reward (R). The extent to which rewards are subjectively discounted is dependent on the discount parameter (*K*) which is different on self and other trials. A high *K* indicates participants are discounting rewards by the effort to a higher degree.

The winning model (model 7) was replicated with the winning session data from Study 2 (BIC = 3671.72; **Fig 3B**) and in the majority of participants (33/45 people) compared to the separate noise model (model 10). Furthermore, the same model was again replicated as the best fitting to the avoiding loss session of Study 2 (BIC = 3597.85, **Fig 3C**), and in the majority of participants (40/45 participants). We also ran model comparisons with the combined winning and losing data from Study 2 to test whether separate win and loss parameters were also needed to explain participants' behaviour. This additional family of models also showed that separate parameters were needed to explain self and other discounting, but also that separate parameters were explained the discounting of rewards and losses (see Supplementary Fig. 1).

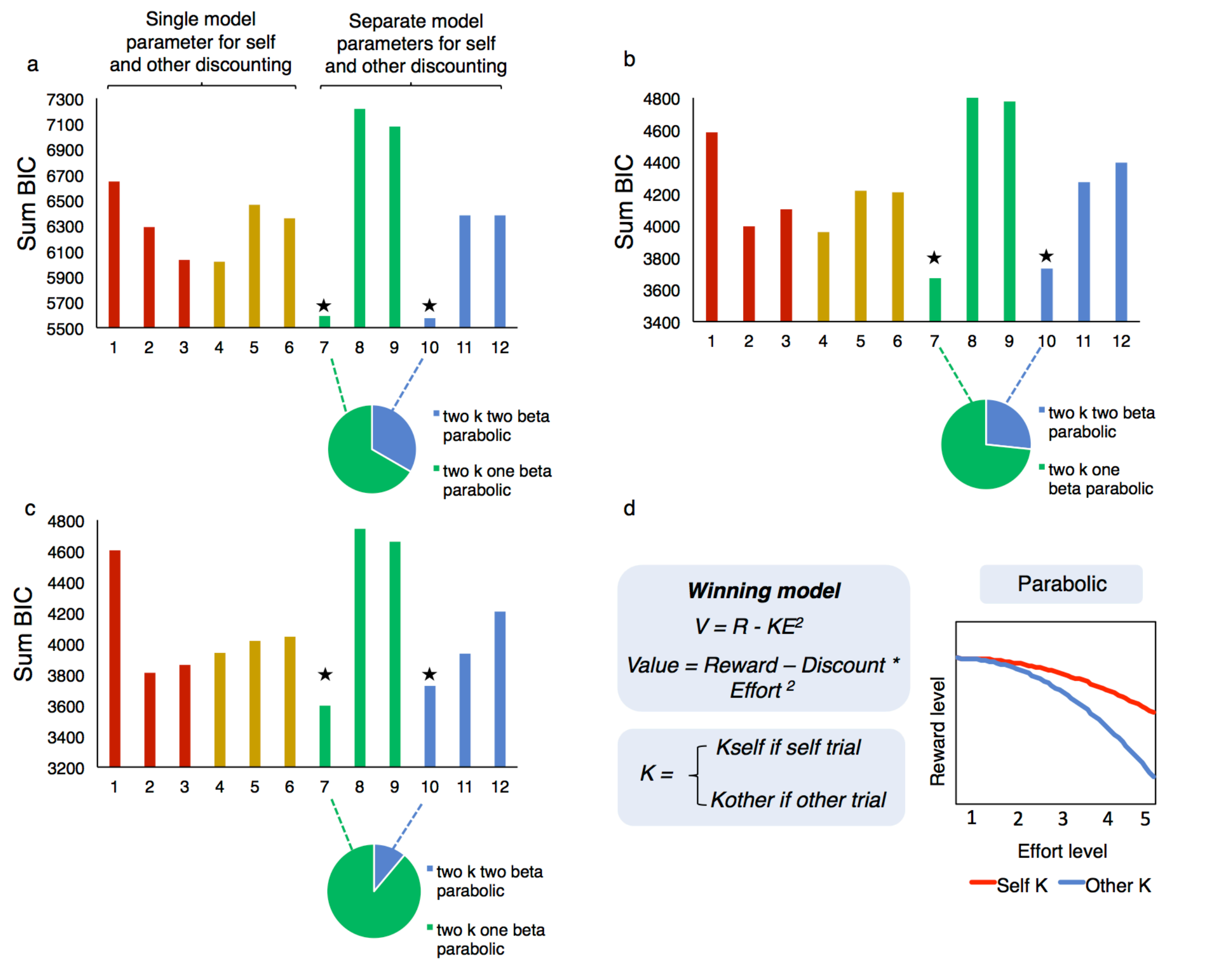

Figure 3. Model comparison robustly shows across two studies that a model with separate discount parameters for self and other best explains behaviour. **A.** Model comparison results from Study 1 with x-axis depicting model number and y-axis the sum of Bayesian information Criterion (BIC) Score. Models with parabolic, linear and hyperbolic discounting functions with either single (models 1-6) or separate discount (K) parameters (models 7-10) for self and other and/or single (red and green) or separate noise (β) parameters (yellow and blue) for self and other. A parabolic model with separate parameters for self and other discounting but a single noise parameter best explained participant's behavior and in the majority of subjects (model 7 in all studies), which was determined by having the lowest summed BIC score in combination with explaining behaviour in the majority of participants. Pie chart shows the proportion of participants the winning model explains behaviour in (green) compared to the same model with separate noise parameters (blue). Asterisks show the two models with lowest BIC scores. **B** Model comparison results from Study 2-win replicated the winning model from Study 1. Asterisks show the two models with lowest BIC scores. **C** Model comparison results from Study 2-lose also replicated the winning model from Study 1 and Study 2-win. Asterisks show the two models with lowest BIC scores. **D** Mathematical formula for the winning parabolic model. The graph shows discount parameters for the median K scores for self and other at reward level 4 for illustrative purposes.

To support the univariate analyses outlined at the beginning of the results, we also compared the discounting parameters from the model for Self and Other. The discounting parameters were higher for other compared to self in both studies and also for the winning data (one participant was excluded from Study-2 win due to very poor model fit resulting in a K value>3SDs above the mean; Wilcoxon Signed-Rank Test: Study 1 Kother > Kself, Z = 5.60, p<.001; Replicated in Study 2-win Z = 5.01, p<.001 and Study 2-loss Z = 4.41, p<.001). However, the parameters were also correlated, albeit only weakly in Study 1, suggesting that there is some baseline level of motivation that affects discounting for *both* self and other (Study 1 *r*=.26, *p*=.08 (marginal) Study 2-win *r*=.39, *p*=.008; Study 2-loss *r* =.35, *p*=.017). We also compared discount rates for self and other across the two samples. Intriguingly, we found that participants discounted rewards by effort for others to a lesser degree in study 2 compared to study 1 (z=-2.83, *p*=.005), but no greater discounting of rewards for self (z=-1.49, *p*=.14). This suggests that meeting another person face to face did in fact increase prosocial motivation, but still not to the same level as self motivation. Note that using, ANOVAs we did not find a significant interaction between agent, effort and sample, suggesting that the model may have provided a more sensitive measure of the balance between reward and effort on participant choices. In sum, these findings support the notion that people are less willing to choose to put in effort to obtain rewards for others, than they are for themselves. That is, they are more apathetic when being prosocial, than when benefiting themselves.

Existing studies suggest that there may be substantial variability in the extent to which people are prosocial[4,7,13,41,42]. Consequently, we investigated whether individual differences in self-reported motivation could explain some of the variance in behaviour. To ensure adequate power for these analyses we pooled together self-report and behavioural data from Study 1 and the winning session of Study 2 (*N*= 92; one participant excluded for extremely poor model fit providing discount (*K*) values > 3 SDs above mean). Our first set of questions related to whether self-report levels of motivation would correlate with willingness to put in effort on the task. We used the

Apathy-Motivation Index (AMI), a self-report measure of levels of motivation with strong psychometric properties[35] (see Supplementary Methods).

The AMI divides apathy into three different domains: Behavioural activation (Self) motivation, emotional sensitivity and Social (Other) motivation. Our main hypothesis was that individuals who were less willing to put in effort on the 'Other' trials would show higher levels of social apathy on the AMI. Consistent with this view, there was a significant positive correlation between levels of Social Apathy on the AMI and 'Other' K values from the model (r = .22, p =.03). Previous studies have shown that willingness to put in effort to obtain rewards for ourselves correlates with behavioural activation[17,33]. In line with these studies, we found that people who had higher discounting for 'Self' (higher self K) also had higher levels of behavioural activation apathy on the AMI (r = .34, p =.002). In contrast, there was no correlation between the 'Other' K parameter and the behavioural activation subscale. Thus, variation in the willingness to exert effort for others in the task appears to relate specifically to everyday self-reported *social apathy* whereas willingness to put in effort to benefit ourselves related to *behavioural apathy*.

Another key individual difference that may affect willingness to help others is our level of psychopathic traits. Psychopathy is a disorder characterised by a constellation of cognitive and behavioural atypicalities including callousness, shallow affect, lack of guilt, antisocial behaviour and impulsivity[53–56]. In the typical population, psychopathic traits can be reliably measured, with these traits existing on a continuum, and often mirroring associations with related constructs that are found in clinical samples[57–59]. Individuals high in psychopathic traits have consistently been found to be less willing to engage in prosocial behaviours[43,44]. In contrast, they may also engage in behaviours associated with impulsivity and antisociality[53,60,61]. Psychopathic traits are therefore prototypically linked to selfish behaviour and a reduced willingness to help others relative to oneself. On this basis, we predicted that psychopathic traits would be associated with a lower willingness to put in effort for rewards for others relative to oneself. As hypothesized, the difference in the discounting parameters between motivation for Other and Self (Kother – Kself) was correlated with total psychopathy scores (r= .24, p=.02; Kother, r=.05, p=.66; Kself, r=-.15, p=.15). Follow up-analyses to determine if this effect was associated with specific aspects of psychopathy showed that there was a correlation with affective-interpersonal (core) psychopathic traits (r= .23, p = .03; KOther, r=.10, p=.37, Kself r=-.10, p=.35) but not lifestyle antisocial traits (r=-.09, p=.40; KOther, r=.-.12, p=.28, Kself r=-.19, p=.08) (see Supplementary Methods). Thus, those higher in psychopathic traits devalue rewards by effort to a greater degree when another person is the beneficiary, relative to when they are putting in effort to help themselves. These findings suggest there might be utility in using our paradigm to understand prosocial deficits in clinical disorders.

Even when subjects do choose to exert effort for others, do they subsequently energise the actions to the same degree? To address this question, we ran linear regressions predicting the area under the curve (see Materials and Methods) of the force applied on each trial. Only trials where participants chose the variable offer were included in this analysis (since these were the only trials where they chose to put in effort). In Study 1 we found an Agent x Effort interaction indicating that as the amount of effort required increased there was an greater difference in the total amount of force exerted (average β = -0.04, Agent x Effort interaction, p<.001). Reduced levels of force exerted at the higher levels of effort for others, relative to self, drove this effect (**Fig 4A**). There were also main effects of Effort, Agent and Reward (Effort: β = .86, p<.001; Agent: β = -0.05, p<.001; Reward β = 0.04, p<.001). Thus, in Study 1 the force exerted increased as the effort required increased, but this happened to a lesser degree when exerting effort to benefit another person.

These findings were replicated in Study 2 for winning (Agent: β = -.08, p<.001; β = 0.51, Effort: β = .85, p<.001; Reward: β = .05, p<.001; Agent x Effort interaction, β = -.03, p<.001). However, for losses there was a main effect of agent on force exerted (β = -0.06, p<.001) but no significant interaction between agent and effort (β = -0.02, p=.14). Given the effects of effort across study 1 and 2, and the interactions between effort and agent when winning rewards in both studies, these results consistently show that people apply less grip force when exerting effort to reward others compared to themselves. Moreover, when benefitting others, at higher levels of effort prosocial acts are energized to a lesser degree when winning monetary rewards (See Supplementary **Fig. 2**).

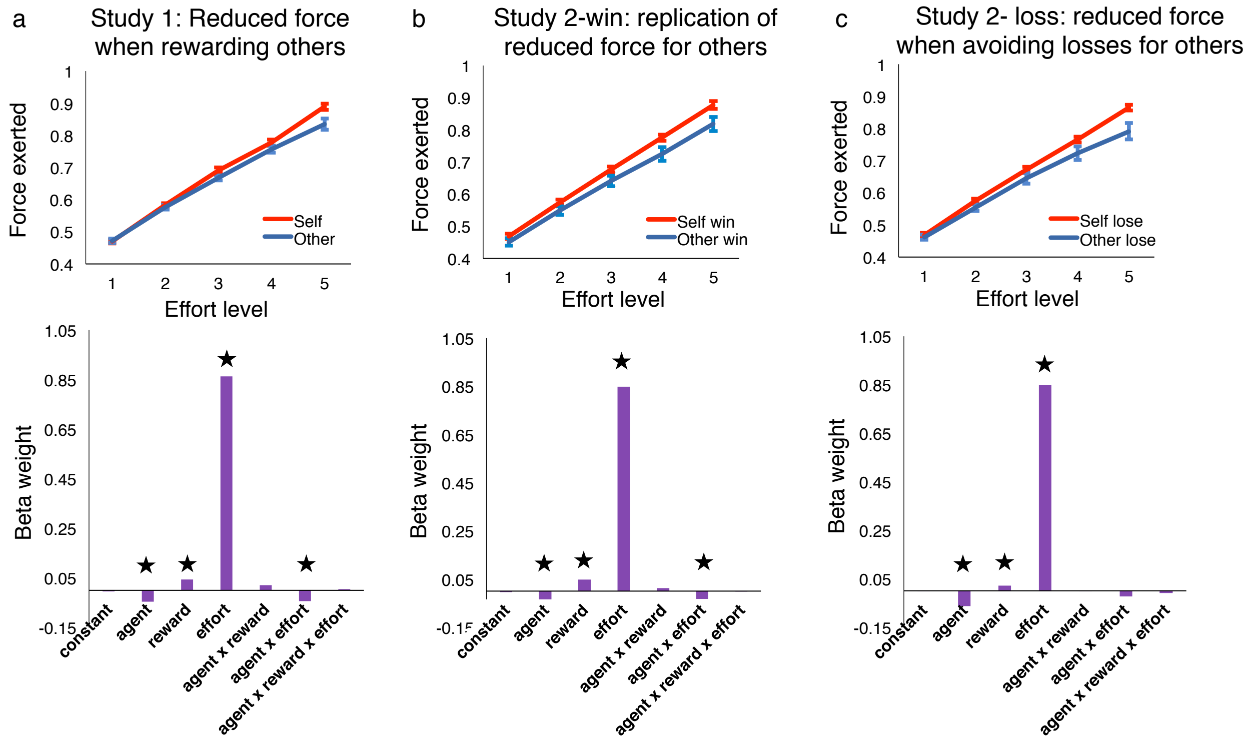

Figure 4. Reduced force when exerting effort to help others.

Even when choosing the experimental option for other, participants applied less grip force than when applying force to reward themselves. **A.** Force exerted as the total area under the curve during the ‘effort period’ on each trial (top panel). Participants applied less grip force when rewarding the other person compared to self at the higher levels of effort. In a linear regression, Agent, Reward, Effort and the Agent x Effort interaction were all significant predictors of force exerted (all $p<.001$). Asterisks show significant beta weight ($p<.001$). Error bars show S.E.M. **B.** Reduced force when rewarding others was replicated in Study 2 (top). As in Study 1, Agent, Reward, Effort and Agent x Effort interaction were significant predictors of force exerted (all $p<.001$). Asterisks show significant beta weight ($p<.001$). Error bars show S.E.M. **C.** Participants also applied less force when avoiding losses for others compared to self (top). Agent, Reward and Effort were again all significant predictors of force exerted (all $p<.001$). Asterisks show significant beta weight. Error bars show S.E.M. ($p<.001$). Together, these results show that people are superficially prosocial, even after choosing to exert effort participants apply less grip force to reward other people than themselves.

There is considerable debate in philosophy, economics and psychology as to whether humans are truly prosocial, willing to suffer costs to themselves to benefit others[2,3,6,14,62]. While empirical studies have quantified how willing people are to donate to charity or individuals and to cooperate with others[4–7,63,64], a crucial factor of everyday prosocial acts is they come at the cost of quantums of effort that must be exerted. Here we derived a computational model, through Bayesian model comparison approaches, that characterises how motivated people are to exert effort to benefit themselves versus others. On two different metrics of effort, which were set to people's own level of strength, we found that people are in fact prosocially apathetic. Although they will perform prosocial acts, they choose to do so less frequently than acts that benefit themselves (**Fig. 2**). Moreover, even after choosing to do so, they put in less force (**Fig. 4**). These effects were replicated across two studies, either when winning rewards or when avoiding losses, and when the other person was completely anonymous or not.

To characterize the *subjective* influence of effort on rewards for self and other we examined a range of computational models that describe how rewards are devalued by the amount of effort required to obtain them. The winning model was a parabolic discount with separate parameters for devaluation of rewards for Self and Other person trials and a single noise parameter (**Fig. 3**). This was robust across studies, regardless of whether people had to exert effort to gain reward or avoid losing it. Moreover, there was a significant correlation between levels of self-reported social apathy and the model discount (K) values for how people devalued reward for another person. The difference in the discounting parameters between motivation for Self and Other also correlated with affective-interpersonal psychopathic traits, consistent with greater discounting of rewards for others compared to self in people who have higher levels of psychopathy.

Overall, these results extend recent research examining the boundaries of prosocial behaviour. Economic theories posit that humans are inherently self-interested and predict that a rational agent would not exhibit any prosocial behaviours unless there was a benefit to themselves[6,32]. However, research into moral dilemmas has shown that people can be hyper-altruistic and desire to benefit others *more* than themselves[4,5]. Prosocial motivation therefore does not appear to completely accord with classical economic theories or research examining prosocial behaviour in terms of moral decisions. Instead, people are willing to put in effort to help others, suggesting that they can be altruistic, but they are less motivated to do so than for themselves, which we refer to as prosocial apathy.

A plethora of research has now shown that while people can process information about others, when the same information pertains to ourselves, it is processed with greater fidelity and speed[65]. People therefore show a ‘self-bias’ for many sources of information. In the prosocial domain, this self-bias is also observed when people learn to perform actions that help others, with individuals being faster to learn what actions will benefit themselves[7]. Our results suggest that there may be a similar self-bias for motivation to exert effort. People are indeed willing to exert effort for others and energise prosocial acts, but to a lesser degree than they will for themselves.

One important benefit of our design is that we had two metrics of prosocial motivation: both choices and the measurement of force even when an individual chose to exert effort. We were therefore able to identify a previously unobserved, but intuitive, phenomena; people show ‘superficial prosociality’. They indicate they are willing to put in effort to help others, but when it comes to performing a prosocial act, they do not put in effort to the same degree as if they were doing it for their own benefit. Thus, strikingly we find that even when people do show a seeming willingness to benefit others, they are not in fact as motivated to benefit another person, particularly when the effort required is high (**Fig. 4**). Moreover, although we found consistent results across both metrics, there was no *a priori* reason to expect this to be the case.

It could be argued that participants in our study were simply complying with a social norm that prosocial acts should not be directed towards strangers. However, such a heuristic or norm would logically operate at all levels of reward and effort. That is, people should always choose not to benefit another, regardless of whether the other person will get a small or large reward. We found consistent interactions between effort and agency in people's decisions, but less consistent reward by agency effects. At low levels of effort, people were relatively similar in their willingness to choose to exert effort to benefit self or other, compared to at the higher levels of effort where the disparity in motivation between self and other was greater. Similar interaction effects were also present in the people's energisation of actions that benefitted self or other. Thus, people's willingness to exert effort for others really did depend on the costs of acting, suggesting that participants were being influenced by the effort costs and not simply using a heuristic that benefiting strangers is undesirable.

There is emerging evidence that the systems in the brain that might be crucial for computing subjective value of exerting effort, and guide our choices of whether it is worth allocating resources, are partially distinct from those that are engaged in mobilising the resources that are required to actually execute the actions[28,29,66]. It is therefore striking that we found evidence of reduced motivation to benefit others relative to self in both choices and the energisation of actions. This result also throws into question whether the hyper-altruistic behaviours observed for morally driven prosocial acts[4,5] may also be due to those studies only examining choice behaviours. Perhaps, if metrics of people's actual energisation of acts that prevent harm to Self vs. Other were examined, their superficiality might lead to a reduction in our assessment of how altruistic people are.

Prosocial acts can be driven by the reputation of another person or the possibility that prosocial acts will be reciprocated[2,14,43,47–49]. In order to maintain experimental control, we tried to limit the effects of reputation and reciprocity by keeping the other participant anonymous in Study 1, and informing the participant that the other person would not know how much money had been earned or lost in Study 2. We also did not observe any interactions between participant gender and the willingness to exert effort for self or other, suggesting similar levels of prosocial motivation in males and females. These results show that, regardless of context, people are less motivated to incur costs to benefit others. However, we did find increased motivation for benefitting others in the second experiment compared to the first, and therefore meeting face to face with another may indeed increase prosocial motivation, even if it still does not lead to same levels of motivation as when benefitting ourselves. Future studies that manipulate social context, such as reputation of another person, the reciprocity of acts, or that examine the allocation of effort between individuals may be able to determine whether prosocial motivation can be raised to an equal or higher level than self motivation. By examining these effects in ecologically valid situations, the generalisability of our results across the spectrum of social situations could be determined.

An alternative account of prosocial behaviour has come from the notion that people experience a ‘warm glow’ or ‘vicarious reward’ from acts that benefit others. This hypothesis is motivated by the fact that areas of the brain that process rewards others receive overlap with those that process one’s own rewards[67–71]. This could potentially account for one element of the results presented here, specifically that people were motivated to perform acts that benefit others at all. However, it cannot account for why people show less motivation to benefit others. Intriguingly there is growing evidence that there are areas of the brain that play important roles in processing information specifically about costs and benefits that pertain to others and not ourselves[69,72,73]. The anterior cingulate gyrus (ACCg) and the subgenual anterior cingulate cortex (sgACC) have been shown to process this information *only* when processing the costs and benefits for others, and not ourselves, and when learning to be prosocial[7,16,69,72,74–76]. These regions also show variability in processing such information that correlates with trait levels of empathy, a crucial driver of prosocial behaviour[7,62,75]. Moreover, there is growing evidence that exerting effort to cooperate or compete with others is driven by separable neural circuits[77]. This would suggest that processing information about the benefits for others might be encoded in a separate system, that may motivate prosocial acts, and its variability drives individual differences in prosocial motivation.

Although overall participants were prosocially apathetic, we also provide evidence that this tendency varies between people[4,7,17,33,41,42]. While self-reported levels of behavioural activation apathy were associated with increased devaluation of rewards for self, replicating previous findings[17,33], social apathy related specifically to how motivated people were to put in effort for others. Moreover, apathy in the self or social domain only related to the willingness to exert effort in the self or other domain respectively. This suggests that a key component of social apathy might be the lack of willingness to engage in effortful social acts, which can be characterized by a specific computational measure. Previous studies have also suggested that self-reported levels of prosociality are lower amongst those with high psychopathic traits[43,44]. Here, we show that lower prosocial motivation on the behavioural level is also apparent in those with higher levels of affective-interpersonal psychopathic traits. Such findings hold promise for using this measure in future studies to precisely quantify levels of prosocial motivation in individuals with clinical levels of apathy and psychopathy

Using computational modelling on an effort-based decision-making task, we show that people show a self-bias in motivation. They are less willing to choose to put in effort and, subsequently energise prosocial acts, less than similarly effortful self-benefitting acts. We also show that people can be superficially prosocial, they are willing to put in effort to help others, but energise the same actions to a lesser degree than they would do if they were the beneficiary. These effects replicated across two studies, were present both when the other person was fully anonymous or if they had met face to face, and also occurred regardless of whether the aim was to win more money, or avoid losses. Such effects were correlated with individual differences in apathy and psychopathy, highlighting the potential for this framework to enlighten our understanding of prosocial behaviour and its links to disorders of social cognition.

Methods

Study 1

Participants

Fifty-three participants (mean age 24.4, SD 4.0, 31 female) aged 19-35 took part. Participants were recruited through university databases. The sample size was based on the study by Crockett et al[4], with additional participants to account for potential exclusions after testing (see below). All participants provided written informed consent and the study was approved by the Oxford University Medical Sciences Inter Divisional Research Ethics Committee. Exclusion criteria included previous or current neurological or psychiatric disorder (as reported by the participants) and non-normal or non-corrected to normal vision. Five participants were excluded from the study because they reported a disbelief in the deception. Participants were instructed the money they would receive at the end of the experiment, as well as the earnings of the confederate participant, would depend on their task performance and would vary between £10-15.

Design

Participants completed 5 blocks of 30 trials (150 total) comprised of 75 decisions for themselves and 75 decisions for the other person. Breaks were included to avoid participant fatigue. Each trial involved a choice between a baseline option that consisted of gaining 1 credit for no effort or an alternative experimental 'offer' that varied in the level of effort (30%, 40%, 50%, 60%, 70% maximum voluntary contraction (MVC)) and level of reward (2, 4, 6, 8 or 10 credits; **Fig.1**).

Apparatus

Stimulus presentation was programmed on a PC using MATLAB (The Math-Works Inc., USA) and Psychtoolbox. Force was recorded using a hand-held TSD121B-MRI (BIOPAC Systems Inc., USA). The PC screen provided subjects with real-time visual feedback on the force being exerted.

Procedure

Role assignment procedure

To ensure that participants believed that their choices were truly resulting in outcomes for another person they were told that there was a second participant taking part in the experiment, but did not see the other participant (who was in fact a confederate) following the procedure described by Crockett et al[4,5]. Participants were told that selecting a ball from a box would randomly assign them to the different roles. Participants were handed a black glove and told not to speak so that the identity of either participant could not be uncovered. A second experimenter arrived in the room, bringing the confederate participant with them who was handed a second glove but remained behind the other side of the door at all times, without ever being seen by the participant. Participants were asked to place their hands in front of the door and wave to one another to ensure it was clear that there was another person there. The experimenter then tossed a coin to decide who would pick from the box first. Each participant selected a ball and was told which role in the experiment they were assigned to.

Task procedure

Participants were asked to grip a handheld dynometer with as much force as possible to determine their MVC. This ensured that although individuals differ in their strength, the effort levels used in the experiment would be relative to that. This measurement was then used as a subject specific threshold for the levels of effort required to obtain rewards in the main task. In the experimental task, participants made decisions between a baseline low effort (0% of MVC) option that gained 1 credit and a variable offer in which more credits (2, 4, 6, 8 or 10 credits) were available but also required more force (30%, 40%, 50%, 60%, 70% of the MVC – represented by segments in a pie chart). The effort and reward levels were varied independently over trials, with each effort-reward combination sampled three times for each agent. There were 150 trials in total, with 75 Self trials where participants chose between the offer and the baseline for themselves and 75 Other trials when they made these decisions for the other person. To obtain the rewards on each trial participants had to apply a force that exceeded the required level for a total of 1s out of a 3s window. Failure to do resulted in 0 credits being delivered. 1 credit was used for the baseline condition to ensure that there was a clear incentive to choose the baseline if the value was not considered worth it, rather than choosing the offer and then not exerting any effort at all. If a choice was not selected 0 credits were delivered. All trials, regardless of the choice made (or if no response was made), lasted for the same duration. This ensured that choices were not influenced by discounting effects of temporal delay rather than effort (e.g.[78]). Indeed, failure rates were very low in the main experiment (<3% study 1; <3.5% Study 2-win; <3.2% Study 2-lose), indicating that subjects were almost always able to achieve a required amount of force. The fact that failure rates were so low also helps to rule our potential effects of risk aversion (e.g.[79]) that may interact with effort discounting, as participants had a very high probability of receiving the rewards from the options they chose.

Prior to the decision-making task, participants experienced each effort level three times across 18 trials. They also learnt to associate each level of effort with the elements in the pie chart. They were instructed that if only one element of the pie chart was shown then 0% force was required and that this was the baseline offer, equivalent to a "rest". However, they still had to grip the dynometer in their hand. During the training session, only 1 credit was on offer and participant's were instructed this credit would not count towards their payment, and they did not choose whether to opt out of exerting the effort.

Study 2

Participants

Fifty-eight participants (mean age 25.7, SD 6.1, 31 female) took part in two sessions. The sample size was based on study 1. Participants were recruited through university databases. All participants provided written informed consent and the study was approved by Oxford University Medical Sciences Inter Divisional Research Ethics Committee. Thirteen participants were excluded from the study because they showed a strong disbelief in the deception. Exclusion criteria included previous or current neurological or psychiatric disorder (as reported by the participants) and non-normal or non-corrected to normal vision. Participants were instructed the money they would receive at the end of the experiment, as well as the earnings of the confederate participant, would depend on their task performance and would vary between £15-25.

Design

There were two additions to the design of Study 1. Firstly, participants performed a session where the aim was to win points as in study 1, but they also included a second session, where the aim was to avoid losing points. Secondly, to avoid participant fatigue, they performed these sessions on separate days (at least 1 day apart), in a counterbalanced order across participants. In the loss session, participants made decisions between a baseline option of -9 credits (equivalent to winning 1 point) for 0% MVC or a variable offer where the offers would be losing 0, -2, -4, -6, -8 credits for themselves or the other participant. If participants chose to accept the experimental option but did not obtain the reward they lost 10 credits. The incentive structure was therefore identical, but the framing was distinct.

Procedure

Role assignment procedure

In the second study, we used a role assignment procedure where participants met another person face to face. This allowed us to examine whether increasing the familiarity and removing the anonymous nature of the other players made people less prosocially apathetic. At the beginning of the experiment participants were introduced to another participant who was in fact a confederate of the experiment. The two participants were instructed that they would be asked to perform different tasks and that only one of them would be in charge of determining the extra amount of money that both would earn (**Fig.1C**).

Study 1 and 2

Statistical analysis of behavioural data

Analyses of behavioural data were performed in SPSS 22 (Armonk, New York: IBM Corp). We examined differences between conditions in choice behaviour for effort and reward and self/other using separate repeated measures analyses of variance (ANOVAs). For the comparison between self and other discounting (k) rates from the model comparison we used Wilcoxon-Signed Rank Tests as the data were not normally distributed. For participants who never chose the variable offer, their discount (k) values were set to 0, as they did not discount rewards by effort (4 participants in Study 2). We also examined bivariate associations between the Kother, Kself and Kother-Kself using Spearman's non-parametric correlation coefficient, also due to non-normality of the data. For the correlations between psychometric scores and computational parameters we only performed planned comparisons that were a priori and hypothesized driven. On the basis of previous findings we hypothesized that discount rates only for others would be associated with social apathy[35], and discount rates only for self would be associated with behavioural apathy[17,33]. For correlations with psychopathic traits, we assessed the correlation with the total score and report the correlations with the two subscales as exploratory follow-up analyses. All comparisons were corrected for the false-discover rate (FDR) using the Benjamini and Hochberg FDR procedure[80]. To test whether there were differences between the force exerted for other and self we calculated the area under the curve (AUC) of the voluntary contraction trace recorded from a hand-held TSD121B-MRI (BIOPAC Systems Inc., USA) using the function 'trapz' in Matlab 2011b (Mathworks). This standard function computes the integral of Y with respect to X using the trapezoidal method. We standardized this value by the maximum level of force exerted to take into account baseline differences in participant's MVC. We then used linear regression analyses to predict the amount of force exerted by reward level, effort level, agent and interactions between these variables. For all analyses, the alpha level was set to $p<.05$ two-tailed.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

The computer code that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Fehr, E. & Camerer, C. F. Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* **11,** 419–427 (2007).
2. Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* **425,** 785–91 (2003).
3. Fehr, E. & Rockenbach, B. Human altruism: economic, neural, and evolutionary perspectives. *Curr. Opin. Neurobiol.* **14,** 784–90 (2004).
4. Crockett, M. J., Kurth-nelson, Z., Siegel, J. Z., Dayan, P. & Dolan, R. J. Harm to others outweighs harm to self in moral decision making. *Proc. Natl. Acad. Sci.* **112,** 201424572 (2015).
5. Crockett, M. J. *et al.* Dissociable Effects of Serotonin and Dopamine on the Valuation of Harm in Moral Decision Making. *Curr. Biol.* **25,** 1852–1859 (2015).
6. Engel, C. Dictator games: a meta study. *Exp. Econ.* **14,** 583–610 (2011).
7. Lockwood, P. L., Apps, M. A. J., Valton, V., Viding, E. & Roiser, J. P. Neurocomputational mechanisms of prosocial learning and links to empathy. *Proc. Natl. Acad. Sci.* **113,** 9763–9768 (2016).
8. Anderson, N. E. & Kiehl, K. A. The psychopath magnetized: insights from brain imaging. *Trends Cogn. Sci.* **16,** 52–60 (2012).
9. Batson, C. D. in *The handbook of social psychology, 4th edition* (New York: McGraw-Hill., 1998).
10. Blair, J., Mitchell, D. & Blair, K. *The Psychopath: emotion and the Brain.* (Blackwell Publishing, 2005).
11. Blair, R. J. R., Jones, L., Clark, F. & Smith, M. Is the Psychopath 'morally insane'? *Personal. Individ. Differ.* **19,** 741–752 (1995).

12. Krajbich, I., Adolphs, R., Tranel, D., Denburg, N. L. & Camerer, C. F. Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *J. Neurosci.* **29,** 2188–2192 (2009).
13. Lockwood, P. L., Seara-Cardoso, A. & Viding, E. Emotion regulation moderates the association between empathy and prosocial behavior. *PloS One* **9,** e96555 (2014).
14. Fehr, E. & Fischbacher, U. Social norms and human cooperation. *Trends Cogn. Sci.* **8,** 185–190 (2004).
15. Apps, M. A. J., Grima, L. L., Manohar, S. & Husain, M. The role of cognitive effort in subjective reward devaluation and risky decision-making. *Sci. Rep.* **5,** 16880 (2015).
16. Apps, M. A. & Ramnani, N. The Anterior Cingulate Gyrus Signals the Net Value of Others' Rewards. *J. Neurosci.* **34,** 6190–6200 (2014).
17. Bonnelle, V. *et al.* Characterization of reward and effort mechanisms in apathy. *J. Physiol.-Paris* (2014). doi:http://dx.doi.org/10.1016/j.jphysparis.2014.04.002
18. Kurniawan, I. T. *et al.* Choosing to Make an Effort: The Role of Striatum in Signaling Physical Effort of a Chosen Action. *J. Neurophysiol.* **104,** 313–321 (2010).
19. Vassena, E. *et al.* Overlapping neural systems represent cognitive effort and reward anticipation. *PloS One* **9,** (2014).
20. Inzlicht, M., Bartholow, B. D. & Hirsh, J. B. Emotional foundations of cognitive control. *Trends Cogn. Sci.* **19,** 126–132 (2015).
21. Bonnelle, V., Manohar, S., Behrens, T. & Husain, M. Individual Differences in Premotor Brain Systems Underlie Behavioral Apathy. *Cereb. Cortex* **26,** 807–819 (2016).
22. Hartmann, M. N., Hager, O. M., Tobler, P. N. & Kaiser, S. Parabolic discounting of monetary rewards by physical effort. *Behav. Processes* **100,** 192–196 (2013).

23. Klein-Flügge, M. C., Kennerley, S. W., Friston, K. & Bestmann, S. Neural Signatures of Value Comparison in Human Cingulate Cortex during Decisions Requiring an Effort-Reward Trade-off. *J. Neurosci.* **36,** 10002–10015 (2016).

24. Klein-Flügge, M. C., Kennerley, S. W., Saraiva, A. C., Penny, W. D. & Bestmann, S. Behavioral modeling of human choices reveals dissociable effects of physical effort and temporal delay on reward devaluation. *PLoS Comput. Biol.* **11,** e1004116 (2015).

25. Skvortsova, V., Palminteri, S. & Pessiglione, M. Learning To Minimize Efforts versus Maximizing Rewards: Computational Principles and Neural Correlates. *J. Neurosci.* **34,** 15621 (2014).

26. Westbrook, A. & Braver, T. S. Cognitive effort: A neuroeconomic approach. *Cogn. Affect. Behav. Neurosci.* **15,** 395–415 (2015).

27. Westbrook, A. & Braver, T. S. Dopamine Does Double Duty in Motivating Cognitive Effort. *Neuron* **89,** 695–710 (2016).

28. Ang, Y.-S., Manohar, S. & Apps, M. A. J. Commentary: Noradrenaline and Dopamine Neurons in the Reward/Effort Trade-off: A Direct Electrophysiological Comparison in Behaving Monkeys. *Front. Behav. Neurosci.* **9,** (2015).

29. Varazzani, C., San-Galli, A., Gilardeau, S. & Bouret, S. Noradrenaline and dopamine neurons in the reward/effort trade-off: a direct electrophysiological comparison in behaving monkeys. *J. Neurosci.* **35,** 7866–7877 (2015).

30. Manohar, S. G. *et al*. Reward Pays the Cost of Noise Reduction in Motor and Cognitive Control. *Curr. Biol.* **25,** 1707–1716 (2015).

31. Winking, J. & Mizer, N. Natural-field dictator game shows no altruistic giving. *Evol. Hum. Behav.* **34,** 288–293 (2013).

32. Camerer, C. *Behavioral game theory: Experiments in strategic interaction.* (Princeton University Press, 2003).

33. Hartmann, M. N. *et al*. Apathy but not diminished expression in schizophrenia is associated with discounting of monetary rewards by physical effort. *Schizophr. Bull.* **41,** 503–512 (2015).
34. Marin, R. Apathy: Concept, Syndrome, Neural Mechanisms, and Treatment. in **1,** 304–314 (1996).
35. Ang, Y.-S., Lockwood, P., Apps, M. A. J., Muhammed, K. & Husain, M. Distinct Subtypes of Apathy Revealed by the Apathy Motivation Index. *PLOS ONE* **12,** e0169938 (2017).
36. Brodaty, H., Altendorf, A., Withall, A. & Sachdev, P. Do people become more apathetic as they grow older? A longitudinal study in healthy individuals. *Int. Psychogeriatr.* **22,** 426–436 (2010).
37. Clarke, D. E., Ko, J. Y., Lyketsos, C., Rebok, G. W. & Eaton, W. W. Apathy and cognitive and functional decline in community-dwelling older adults: results from the Baltimore ECA longitudinal study. *Int. Psychogeriatr.* **22,** 819–829 (2010).
38. Vansteenkiste, M., Lens, W., De Witte, S., De Witte, H. & Deci, E. L. The 'why'and 'why not'of job search behaviour: their relation to searching, unemployment experience, and well-being. *Eur. J. Soc. Psychol.* **34,** 345–363 (2004).
39. Vansteenkiste, V., Lens, W., Witte, H. & Feather, N. Understanding unemployed people's job search behaviour, unemployment experience and well-being: A comparison of expectancy-value theory and self-determination theory. *Br. J. Soc. Psychol.* **44,** 269–287 (2005).
40. Insel, T. *et al*. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. (2010).
41. FeldmanHall, O., Dalgleish, T., Evans, D. & Mobbs, D. Empathic concern drives costly altruism. *Neuroimage* **105,** 347–356 (2015).

42. Zaki, J. & Ochsner, K. N. The neuroscience of empathy: progress, pitfalls and promise. *Nat. Neurosci.* **15,** 675–680 (2012).
43. White, B. A. Who cares when nobody is watching? Psychopathic traits and empathy in prosocial behaviors. *Personal. Individ. Differ.* **56,** 116–121 (2014).
44. Foulkes, L., McCrory, E. J., Neumann, C. S. & Viding, E. Inverted Social Reward: Associations between Psychopathic Traits and Self-Report and Experimental Measures of Social Reward. *PLoS ONE* **9,** e106000 (2014).
45. Salamone, J. D., Yohn, S. E., López-Cruz, L., San Miguel, N. & Correa, M. Activational and effort-related aspects of motivation: neural mechanisms and implications for psychopathology. *Brain J. Neurol.* **139,** 1325–1347 (2016).
46. Kennerley, S. W., Behrens, T. E. J. & Wallis, J. D. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat. Neurosci.* **14,** 1581-U119 (2011).
47. Kurzban, R., DeScioli, P. & O'Brien, E. Audience effects on moralistic punishment. *Evol. Hum. Behav.* **28,** 75–84 (2007).
48. Piazza, J. & Bering, J. M. Concerns about reputation via gossip promote generous allocations in an economic game. *Evol. Hum. Behav.* **29,** 172–178 (2008).
49. Haley, K. J. & Fessler, D. M. T. Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evol. Hum. Behav.* **26,** 245–256 (2005).
50. Tversky, A. & Kahneman, D. Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertain.* **5,** 297–323 (1992).
51. Tversky, A. & Kahneman, D. Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Q. J. Econ.* **106,** 1039–1061 (1991).
52. Chong, T. T.-J. *et al*. Neurocomputational Mechanisms Underlying Valuation of Effort Costs. *Plos Biol.* (in press).

53. Blair, R. J. R. The neurobiology of psychopathic traits in youths. *Nat. Rev. Neurosci.* 786–799 (2013).

54. Cleckley, H. The mask of sanity; an attempt to reinterpret the so-called psychopathic personality. (1941).

55. Hare, R. D. & Neumann, C. N. in *Handbook of Psychopathy* (ed. Patrick, C.) 58–88 (New York: Guilford., 2006).

56. Hare, R. D. *The Hare Psychopathy Checklist-Revised: PLC-R*. (MHS, Multi-Health Systems, 1999).

57. Paulhus, D., Neumann, C. & Hare, R. *Manual for the self-report psychopathy scale*. (Toronto: Multi-Health systems, 2009).

58. Lockwood, P. L., Bird, G., Bridge, M. & Viding, E. Dissecting empathy: high levels of psychopathic and autistic traits are characterized by difficulties in different social information processing domains. *Front. Hum. Neurosci.* **7,** 1–6 (2013).

59. Seara-cardoso, A., Neumann, C., Roiser, J., Mccrory, E. & Viding, E. Investigating associations between empathy, morality and psychopathic personality traits in the general population. *Personal. Individ. Differ.* **52,** 67–71 (2012).

60. Porter, S. & Woodworth, M. Psychopathy and aggression. *Handb. Psychopathy* 481–494 (2006).

61. Blair, R. Psychopathic traits from an RDoC perspective. *SI Neuropsychiatry* **30,** 79–84 (2015).

62. Decety, J., Bartal, I. B.-A., Uzefovsky, F. & Knafo-Noam, A. Empathy as a driver of prosocial behaviour: highly conserved neurobehavioural mechanisms across species. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **371,** (2016).

63. Ruff, C. C. & Fehr, E. The neurobiology of rewards and values in social decision making. *Nat. Rev. Neurosci.* **15,** 549–562 (2014).

64. Lee, D. & Seo, H. Neural Basis of Strategic Decision Making. *Trends Neurosci.* **39,** 40–48 (2016).

65. Sui, J. & Humphreys, G. W. The Integrative Self: How Self-Reference Integrates Perception and Memory. *Trends Cogn. Sci.* **19,** 719–728 (2015).

66. Floresco, S. B. Noradrenaline and dopamine: sharing the Workload. *Trends Neurosci.* **38,** 465–467 (2015).

67. Andreoni, J. & Miller, J. Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica* **70,** 737–753 (2002).

68. Braams, B. R. *et al.* Reward-related neural responses are dependent on the beneficiary. *Soc. Cogn. Affect. Neurosci.* **9,** 1030–1037 (2014).

69. Lockwood, P. L. The anatomy of empathy: Vicarious experience and disorders of social cognition. *Behav. Brain Res.* **311,** 255–266 (2016).

70. Mobbs, D. *et al.* A Key Role for Similarity in Vicarious Reward. *Science* **324,** 900–900 (2009).

71. Morelli, S. A., Sacchet, M. D. & Zaki, J. Common and distinct neural correlates of personal and vicarious reward: A quantitative meta-analysis. *NeuroImage* **112,** 244–253 (2015).

72. Apps, M. A. J., Rushworth, M. F. S. & Chang, S. W. C. The Anterior Cingulate Gyrus and Social Cognition: Tracking the Motivation of Others. *Neuron* **90,** 692–707 (2016).

73. Lamm, C. & Majdandžić, J. The role of shared neural activations, mirror neurons, and morality in empathy--a critical comment. *Neurosci. Res.* **90,** 15–24 (2015).

74. Chang, S. W. C., Gariépy, J.-F. & Platt, M. L. Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* **16,** 243–250 (2013).

75. Lockwood, P. L., Apps, M. A. J., Roiser, J. P. & Viding, E. Encoding of Vicarious Reward Prediction in Anterior Cingulate Cortex and Relationship with Trait Empathy. *J. Neurosci.* **35,** 13720–13727 (2015).

76. Balsters, J. H. *et al.* Disrupted prediction errors index social deficits in autism spectrum disorder. *Brain* **140,** 235–246 (2016).

77. Le Bouc, R. & Pessiglione, M. Imaging Social Motivation: Distinct Brain Mechanisms Drive Effort Production during Collaboration versus Competition. *J. Neurosci.* **33,** 15894–15902 (2013).

78. Green, L. & Myerson, J. A discounting framework for choice with delayed and probabilistic rewards. *Psychol. Bull.* **130,** 769 (2004).

79. Kahneman, D. & Tversky, A. Prospect theory: An analysis of decision under risk. *Econom. J. Econom. Soc.* 263–291 (1979).

80. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 289–300 (1995).

Correspondence: All correspondence and requests for materials should be addressed to Dr. Patricia Lockwood (patricia.lockwood@psy.ox.ac.uk)

Acknowledgements

We would like to thank Prof. Craig Neumann for his help and advice with regard to the Self-Report Psychopathy Scale. We would also like to thank all members of the Cognitive Neurology group for assistance as experimental confederates. This work was supported by a BBSRC Anniversary Future Leader Fellowship (BB/M013596/1) to MAJA and a Wellcome Trust Principal Fellowship to MH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

P.L.L, M.A.J.A., M.HU, designed study, P.L.L., M.HA., S.H.Z., A.R, F. S. & M.A.J.A. collected data, P.L.L, M.A.J.A., analysed data, P.L.L, M.A.J.A., & M.HU wrote paper.

Competing interest statement

The authors declare that the research was conducted in the absence of any commercial, non-financial or financial relationships that could be construed as a potential conflict of interest.