To give or not to take – insular cortex subserves framing-induced boosts in generosity during social discounting

Manuela Sellitto[a,1], Adam Schweda[a], Bernd Weber[b], Tobias Kalenscher[a]

[a]Comparative Psychology, Institute of Experimental Psychology, Heinrich-Heine University of Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

[b]Institute of Experimental Epileptology and Cognition Research, Sigmund-Freud Straße 25, 53127 Bonn, Germany

[1]To whom correspondence should be addressed.
Department of Comparative Psychology
Institute of Experimental Psychology
Heinrich-Heine University Düsseldorf
Universitätsstraße 1
40225 Düsseldorf
Germany
Phone +49-211-81-12360
Fax +49-211-81-11629
E-mail: manuela.sellitto@hhu.de

Author contributions: M.S. and T.K. conceived the research project; M.S., A.S., B.W., and T.K designed research; M.S. and A.S. performed research; M.S. analyzed data; M.S. and T.K. wrote the paper; and M.S., A.S., B.W., and T.K revised versions of the manuscript.
The authors declare no conflict of interest.

Acknowledgments: This work was supported by Deutsche Forschungsgemeinschaft Grant (DFG) KA 2675/4-3 (to T.K.).

Abstract

Sacrificing own resources for the benefit of others is a prerequisite for society to function well. The willingness to do so, however, greatly declines as the perceived social distance from the other persons increases. Here, we asked if describing a generous choice as preventing a loss to others rather than granting them a gain would promote generosity, even towards socially remote strangers. We adapted a social discounting task where participants chose between a selfish and a costly generous option, yielding either advantageously unequal or fair financial outcomes. We manipulated the description of the decision problem in two economically equivalent frame-conditions: in the gain frame, a generous choice yielded a gain to the other, while, in the loss frame, it entailed preventing the loss of a previous endowment to the other. In four behavioral replications, we found that social discounting was strongly reduced in the loss than the gain frame, implying that participants were more generous toward remote strangers. We hypothesized that the motives underlying generosity in social discounting are frame-dependent and dissociable on the neural level. Using functional magnetic resonance imaging, we found that temporoparietal junction and ventromedial prefrontal cortex subserved generosity in the gain frame, replicating our previous finding, while anterior insular cortex was selectively recruited during generous choices in the loss frame when participants prevented the other-loss. This suggests that anterior insula promoted generosity toward strangers by encoding an anticipated harm prevention signal. These results extend our understanding of the role of insula in nudging prosocial behavior.

Keywords in alphabetical order: decision making, fMRI, framing effect, harm prevention, insular cortex, prosocial choice, social discounting, TPJ, VMPFC

Introduction

Most human societies are collaborative and offer benefits to their members that they would not be able to achieve individually. However, societies can only function efficiently if their members are willing to contribute to causes whose beneficiaries are abstract and anonymous, such as public goods, and/or to causes whose beneficiaries are socially remote, as it is often the case with wealth redistribution for social welfare, public health insurance, or state pension systems. The acceptance and support of the principle of a caring society, and the attitude towards the welfare of socially remote strangers, is therefore central for a society to function well; it also seems vital for societies to successfully meet current challenges, such as integrating refugees, addressing economic inequality, or acceding the trials and promises of a globalized world (1). Most people are indeed willing to sacrifice own resources for the welfare of others (2, 3). However, their generosity typically declines steeply with social distance between them and the recipients of help, a phenomenon called social discounting (4, 5). Hence, while people are ready to provide costly support to friends, relatives, and acquaintances, they are less inclined to help remote strangers.

The social discount function is idiosyncratic (6–9), and it is far from stable within and across individuals. For instance, we and others have shown that participants from individualistic or collectivistic cultures (10) differ in their attitude towards the welfare of socially close peers; that psychosocial stress (11) and neurohormonal action (12, 13) can increase generosity towards socially close friends and acquaintances; and that the level of prosociality towards socially close others depends on gender and cognitive load (14, 15). We further showed that disrupting the temporoparietal junction (TPJ) – a brain region we recently identified as a central hub orchestrating the balance between egocentric and other-regarding preferences in social discounting (5) – by the means of transcranial magnetic stimulation increases the steepness of social discounting (16), thus lowering the willingness to support socially remote strangers.

This body of evidence strongly suggests that the degree at which individuals value socially close and distant others' well-being is highly malleable; namely, cultural, cognitive, and neurobiological factors can enhance the consideration of the well-being of socially close relatives, friends, and acquaintances, but they can at the same time lower the attitude towards the welfare of socially remote strangers. However, despite its paramount theoretical and societal significance, means to *increase* the inclination for costly support of socially remote beneficiaries are elusive.

Here, we provide behavioral and neural evidence for a simple manipulation that aims at significantly increasing individuals' willingness to costly support socially remote others. We make use of the observation that people are more sensitive to others' losses than gains (17–26), and are consequently strongly reluctant to increase their own payoff at the expense of others' welfare (27–31). We hypothesized that participants would be more altruistic towards

others, including socially remote strangers, if a costly generous choice was framed as preventing a monetary loss to others rather than granting them a gain, even if actual economic outcomes were equivalent. In other words, we expected that the way a prosocial decision problem was framed affected the shape of the social discount function.

To test this hypothesis, we elicited social preferences in a standard version of the social discounting task (gain frame; (5)) as well as in a loss frame variant (Fig. 1). In each trial, participants decided to share money with other individuals on variable social distance levels. They chose between a selfish option, yielding high own-payoff and zero other-payoff, and a generous option, yielding a lower own-payoff and a non-zero other-payoff. The main difference between conditions was the way the decision problem was described: in the gain frame, a costly generous choice would yield a gain to the other player, while, in the loss frame, it would imply preventing the loss of a previous endowment to the other player. Importantly, the payoff structure was mathematically identical across frame conditions, i.e., the choice alternatives in the loss frame yielded identical own- and other-payoffs to those in the gain frame. Participants were explicitly instructed that the other persons would only be informed about the final outcome, but not about their endowment, or the loss of it; hence, they knew about the economic equivalence across frames.

We show in a series of independent studies that participants were more reluctant to make a selfish choice if this implied a loss of the endowment to the other, resulting in a substantially flatter social discount function in the loss than the gain frame (Fig. 2 and Fig. S1) and, hence, higher generosity towards socially remote others.

To obtain further insights into the psychological and neural mechanisms underlying this frame effect on social discounting, we measured blood oxygen level-dependent (BOLD) responses while participants performed both frame variants of the social discounting task. We hypothesized that the psychological motives underlying generosity were frame-dependent and dissociable on the neural level. Specifically, consistent with our previous work (5), we predicted that generosity in the gain frame was vicariously rewarding and the result of the resolution of the conflict of selfish and altruistic motives. In line with (5), we therefore predicted that generous choices in the gain frame would go along with activations in ventromedial prefrontal cortex (VMPFC), a brain structure known to represent own and vicarious reward value (32, 33), as well as TPJ, which has been associated with perspective taking (34) and overcoming egoism bias (5, 16). Conversely, in the loss frame, we expected that the disinclination to maximize own-gain at the expense of other-loss was motivated by the desire to comply to social norms, such as the respect of others' property rights, or the do-no-harm principle. We therefore hypothesized increased activity in brain regions associated with guilt and negative emotions of social norm transgressions, mainly the anterior insula (22, 35–42). Our results support our main hypothesis that frame-dependent choice motives were associated with

distinct neural signatures. In the gain frame, we replicate VMPFC and TPJ involvement in generous choice (5, 43), while in the loss frame, we identify anterior insula as the core component in processing altruistic motives, prompting generosity by biasing people away from selfish desires if these caused harm to others.

Results

Social discounting is flatter in the loss than the gain frame

First, in a set of behavioral experiments, we established that our framing manipulation affected generosity towards socially distant others. We elicited social preferences in a standard version of the social discounting task (the gain frame; (5)) as well as in a loss frame variant (Fig. 1). In the gain frame, participants played with other persons at variable social distance levels, and made choices between a selfish alternative, yielding higher monetary payoff to the participant and zero payoff to the other, and a generous alternative, always yielding a lower own-payoff of €75 along with a payoff of €75 to the other. In the loss frame, participants were first informed that the other person had received an initial endowment of €75. The selfish alternative yielded a variable, higher own-payoff as well as the loss of the €75 endowment to the other person, hence, resulting in a zero payoff to the other; the generous alternative yielded a fixed €75 payoff to the participant, and no further consequence to the other, thus leaving her with the initial €75 endowment. Importantly, the payoff structure was mathematically equivalent across both frame conditions, i.e., the choice alternatives in the loss frame yielded identical own- and other-payoffs to those in the gain frame. The main difference between conditions was that, in the gain frame, a generous choice would imply a gain of €75 to the other, while in the loss frame, a generous choice would imply preventing the loss of the previous €75 endowment. Importantly, participants were repeatedly instructed that the other person was unaware of her initial endowment, or the loss of it, and that she would only be informed about the final outcome of the payoff after implementing the participant's choice at the end of the experiment. Task comprehension, in particular regarding the participants' understanding that the other person would only be informed about the final outcome, but not about her endowment, or loss of it, was further stressed during the explanation of the incentivization procedure (see Materials and methods) as well as assessed in post-hoc structured interviews. All participants understood the task well.

Study 1. In a first study, data collection was done online and the task was not incentive-compatible; participants (n = 52) were paid a fixed allowance of €8.5. To reconstruct the individual social discount functions, separately for the two frame conditions, we fit a standard hyperbolic model (see Eq. 1 (4, 5); Materials and methods) to trial-by-trial choices via a softmax function to estimate the parameter k, a measure of the steepness of the social discount

function. Additionally, we determined, for each participant and each social distance level, the point at which the participant was indifferent between the selfish and the generous alternative using logistic regression (5). The difference in reward magnitudes for the participant between the two alternatives at the indifference points (see Supplemental results in Supplementary Information) represented the amount of money a subject was willing to forego (i.e., reward amount foregone) to increase the wealth of another person at a given social distance, and could be construed as a social premium, that is, the price tag participants put on increasing the wealth of the other. We took the parameter V, an indicator of generosity towards socially close others (5, 16), as the intercept at social distance 0, thus the origin of the social discount function (4, 5, 11, 13, 16).

Participants' generosity dropped much less steeply in the loss- compared to the gain-frame (median k_{gain} = 0.011 vs. median k_{loss} = 0.004; Wilcoxon matched pairs test: Z = 4.03, $p < 0.001$; $\eta_p^2 = 0.57$; Fig. S1a). The difference in social discount functions between frames was most pronounced at high social distance levels, indicating that participants were substantially more generous towards strangers in the loss than the gain frame. We also found a significant difference in V between frame conditions (median V_{gain} = 89 vs. median V_{loss} = 80; Z = 3.36, $p < 0.001$; $\eta_p^2 = 0.47$) that, however, disappeared when removing all participants with zero discounting from the analysis (see Supplemental results in Supplementary information). These data suggest that participants were strongly more generous towards socially distant others in the loss- than the gain-frame.

Study 2. In a second study, we replicated the results of our first experiment. Data collection was done online and participants (*n* = 30) were reimbursed for their time with a fixed amount of university credits. We again found that participants had flatter social discounting in the loss- than the gain-frame (median k_{gain} = 0.017 vs. median k_{loss} = 0.008; Z = 2.89, $p < 0.004$; $\eta_p^2 = 0.52$; Fig. S1b), and we found no difference in V between frame conditions (median V_{gain} = 80 vs. median V_{loss} = 80; Z = 0.21, p = 0.83). Again, these results held when excluding participants with null discounting.

Study 3. Studies 1 and 2 were not incentive-compatible. To determine whether hypothetical versus real payoffs made a difference in the frame effect on social discounting (44), we ran a third fully incentive-compatible study in a laboratory setting (*n* = 31). Payoff was contingent on the participants' choices, and was paid out to self and other, identical to (5) and to the fMRI study 4 (see next paragraph and Materials and methods). Once again, we could replicate the frame effect on k (median k_{gain} = 0.021 vs. median k_{loss} = 0.012; Z = 3.16, $p < 0.001$; $\eta_p^2 = 0.57$; Fig. S1c), and there was no difference in the V parameter between frame conditions (median

V_{gain} = 81 vs. median V_{loss} = 80; Z = 1.43, p = 0.15). These results held when excluding participants with null discounting.

Overall these results suggest that, while generosity to socially close others was comparable between frame conditions, it decayed significantly less steeply in the loss than in the gain frame, indicating that participants were considerably more generous towards socially distant others in the loss frame compared to the gain frame.

Neural mechanisms underlying the frame effect on social discounting

To obtain more substantial insights into the psychological and neural mechanisms underlying this framing effect on social discounting, in study 4, we measured BOLD responses while participants performed both frame variants of the social discounting task. The fundamental premise of our study is that the decision motives and their neural correlates differ between gain and loss frames. Specifically, we reasoned that generosity in the gain frame was mainly stimulated by other-regarding considerations. Conversely, we predicted that generous decisions in the loss frame were motivated by the desire to comply to social norms, such as the *do-no-harm* principle, or the respect of others' property rights (45), infringements of which are associated with negative social sentiments of guilt and shame. To test this idea, we focused on three main hypotheses. We, first, expected that selfish choices in both frame conditions were motivated by material self-interest, as reflected by activation in the brain's valuation network, mainly VMPFC (32, 33). Second, we predicted that generosity in the gain frame recruited a network of structures, including VMPFC and TPJ (5, 43), known to represent vicarious reward value and overcoming egoism bias. Finally, we hypothesized that brain areas implicated in negative social sentiments of social norm transgressions, such as anterior insula (e.g., (28, 29, 40–42, 46–49)), would be selectively recruited during generous choices in the loss, but not the gain frame.

We first replicated, once more, the behavioral framing effect on social discounting (*n* = 30; two participants were excluded from the behavioral analyses due to bad curve fitting or being outlier; see Materials and methods). As before, the drop in generosity across social distance was pronouncedly flatter in the loss- than the gain-frame (median k_{gain} = 0.019 vs. median k_{loss} = 0.009; Wilcoxon matched pairs test: Z = 3.46, p < 0.001; η_p^2 = 0.65; Fig. 2), but, again, there was no difference in the V parameter between conditions (median V_{gain} = 80 vs. median V_{loss} = 80; Z = 1.00, p > 0.31; the results remained identical when excluding participants with null discounting).

To address our first hypothesis that selfish behavior is motivated by material self-interest and would, hence, recruit structures of the brain's valuation network, we analyzed the neural activity during selfish choices in both frames (GLM1; see Materials and methods). In support of our hypothesis, we found that selfish choices, pooled across frame conditions, correlated

with activity in VMPFC (MNI peak coordinates at 2, 36, 22, $p_{FWE\text{-}corr} < 0.001$) as well as in other prefrontal valuation regions (Fig. 3; see Table S1 for a complete list of activations).

To test our second hypothesis, we aimed to replicate our previous finding of brain structures known to represent vicarious reward value and overcoming egoism bias in the gain frame (5, 16). Our results (GLM2; see Materials and methods) indeed revealed clusters located in VMPFC (0, 54, 14, $p_{FWE\text{-}corr} < 0.001$) and caudate (-10, 16, 10, $p_{FWE\text{-}corr} < 0.02$) to be selectively activated when participants made generous choices in the gain frame relative to generous choices in the loss frame. Additionally, a region of interest (ROI) analysis showed that right TPJ (54, -60, 34, $p_{FWE\text{-}corr} < 0.02$) was also activated in this contrast. Thus, consistent with (5, 43), a network comprising VMPFC and TPJ seems to underlie the motivation for costly generosity in the gain frame (Fig. 4; see Table S2 for a complete list of activations).

Our third hypothesis predicted that generosity in the loss frame is motivated by social norm compliance rather than other-regarding considerations; generosity would, consequently, go along with a different activation pattern in the loss than the gain frame. Accordingly, the right anterior insula (42, 4, -4, ROI analysis, $p_{FWE\text{-}corr} < 0.02$; GLM2, see Materials and methods), was selectively activated during generous choices in the loss frame relative to generous choices in the gain frame, thus in the reverse contrast than the one used to test our second hypothesis (Fig. 5a; see Table S2 for a complete list of activations).

To further illustrate this result, we quantified the extent at which insula activation correlated with the individual propensity to make generous choices in the gain and the loss frame. To this end, we extracted the parameter estimates at the individual level from an ROI with a seed in the anterior insula (coordinates obtained from the results of GLM2, see above [42, 4, -4]). Parameter estimates were extracted separately for the gain and the loss frame, after pooling all choices within each frame (GLM3; see Materials and methods). We then correlated, for each frame separately, the extracted betas with the percentage of generous choices in the respective frame. While insula activation did not significantly covary with generous choices in the gain frame ($r = -0.01$, $p = 0.48$; Fig. 5b), it did correlate positively with generous choices in the loss frame ($r = 0.034$, $p = 0.03$, one-tailed, $d_{Cohen} = 0.72$; Fig. 5c), thus corroborating the idea that anterior insula played a different role in promoting generosity in the loss than in the gain frame.

Our analysis so far suggests that insula activation reflects the psychological motives underlying generous choice in the loss frame. However, other explanations of our insula finding are conceivable, too. For instance, participants made more generous choices overall in the loss than the gain frame; i.e., they forewent more own-payoff in the loss- than the gain-frame, and insula activation might reflect the higher level of reward foregone in the loss frame. Yet, the trial-by-trial regressor of reward amount foregone (GLM2; see Materials and methods) revealed no parametric modulation of insula activity, nor of activity in any other brain region,

during generous choices in either frame condition. Additionally, insula activity is unlikely to reflect the own-reward component of the generous alternative because it was fixed (always €75) and, thus, invariant across trials in both frames. In conclusion, we maintain that the most parsimonious explanation of insula activation is indeed a frame-dependent coding of generosity in social discounting.

Discussion

We provide behavioral and neural evidence for a simple nudge that aims at increasing individuals' willingness to costly support socially remote others. We adapted a social discounting task where participants chose between a selfish option – a high gain to the self and zero-gain to the other – and a generous option – a lower gain to the self and equal gain to the other (5, 16). Based on previous evidence that people are strongly reluctant to increase their own payoff at the expense of others' welfare (27–30), we framed the generous option either as a monetary gain to the other (gain frame), or as the prevention of the loss of a previous monetary endowment to the other (loss frame) (see (19–25, 31)). Crucially, between frames, the choice alternatives differed only in the description of the decision problem, but not with regard to their actual economic consequences. In a series of four studies, we show that the social discount function was significantly flatter in the loss- than the gain-frame, indicating that participants were more generous towards socially remote others if a personal gain implied the other's loss of their previous endowment. Notably, our incentivization procedure made it logically impossible for the other person to know about her endowment, or the potential loss of it, and participants were explicitly instructed about this; all that mattered was the final positive payoff to self and others. Yet, the fact that our participants were still reluctant to inflict losses to others suggests that they had internalized the social norm of not taking away money from others to such a degree that it prevailed even in the absence of any real economic consequences for others.

We further hypothesized that the frame-dependent motives underlying generosity are dissociable on the neural level. Consistent with our previous work (5, 16), we found that generosity in the gain frame recruited a network of structures, including VMPFC and TPJ, known to represent vicarious reward value and overcoming egoism bias. By contrast, in the loss frame, we expected that the reluctance to maximize own-gain at the expense of a loss to others would be mediated by social norm compliance and associated social sentiments. We therefore hypothesized that increased activity in brain regions associated with such social sentiments, specifically the insular cortex, would be associated with generous choice in the loss frame specifically (22, 35, 47, 49–55, 36–42, 46). We indeed found that the anterior insula was significantly more activated when participants made generous choices in the loss frame, relative to the gain frame, and our analysis further showed that the degree of insula activation

across trials correlated with the individual propensity to make generous choices in the loss, but not the gain frame. Overall, these results call for the idea that the difference in generosity between frames is *one of kind, and not of degree*, that is, that the mental processes behind generosity are likely qualitatively different in the gain and the loss frame.

Our findings expand on previous evidence that preventing harm to others is a great motivator of prosocial performance (19, 22–24, 56). However, while others have found that harm prevention was particularly pronounced in a public context (19) and dependent on social feedback (20, 21), we show that similar cognitive mechanisms can strongly boost generosity even in a private context and in the absence of social feedback, thus independent of reputational concerns, judgment by social peers, or third-party punishment threats. This suggests that other-harm prevention might be an internalized motive that works unconditionally and universally across contexts, regardless of social consequences. In addition, previous experiments on harm prevention did not manipulate, or provide information on, social distance between donor and recipient (19, 23, 25, 26, 30). Hence, while the effects of the resource allocation mode on social discounting were elusive so far, our findings imply that it matters: harm-prevention motives in the loss frame were less dependent on social distance than other-regarding considerations in the gain frame, thus resulting in flatter social discounting.

Our results are consistent with the idea that certain costly altruistic behaviors are not motivated by genuinely other-regarding considerations, but instead by compliance to internalized social norms. But what impels participants to comply to social norms? Here, we propose, along with the evidence in (28), that compliance to social norms – transgression of which, in the present experiment, is attached to the act of inflicting the loss of the other person's initial endowment if a selfish choice is made in the loss frame – might be linked to anticipated feelings of guilt, shame, and remorse, and accompanying insula activation (see also (57, 58)), which ultimately sustain prosocial behavior.

In conclusion, we present a simple behavioral framing manipulation that boosts generosity towards socially remote others: framing a selfish choice as a loss to others can result in a negative affective state, eventually motivating prosocial behavior, even if the framing of the choice options is irrelevant for the actual payoff to others. Our neuroimaging data identify insula as the core component in processing loss prevention motives, prompting generosity by biasing people away from selfish desires if these caused harm to others, including strangers. Our results imply that prosocial attitudes towards others are highly malleable and strongly depend on the architecture of the decision problem. The insights gained in this study might, thus, help in designing policies aimed at increasing the acceptance and support of the principle of a caring society, and to change the attitude towards the welfare of socially remote strangers.

Materials and methods

Participants

Studies 1-3. Three separate behavioral studies were carried out to test the validity of our paradigm in different settings and with different compensation procedures. For these studies we did not calculate the sample size in advance as we were not aware of any previous similar manipulation of social discounting. Study 1 was run online (*n* = 61; nine participants later excluded from the hyperbolic model fitting due to bad fitting or being outlier; 27 females; mean age = 36 years, ±10.5 standard deviation) and participants were paid a fixed allowance of €8.5. Study 2 was run online (*n* = 36; six participants later excluded from the hyperbolic model fitting; 27 females; mean age = 21 years, ±2.7) and participants, all psychology students on campus, were reimbursed for their time with a fixed amount of university credits. Study 3 (*n* = 39; eight participants later excluded from the hyperbolic model fitting; 24 females; mean age = 25 years, ±5.7) was run in the laboratory and participants were paid out with the same fully incentive-compatible procedure as in the fMRI study 4 (see below). All participants performed a social discounting task and, at the end, they completed a questionnaire assessing social desirability (see Supplementary information). Although the social discounting task of studies 1 and 2 was not incentivized, participants were strongly encouraged to think as if they were making decisions for real. All three studies were conducted according to the Declaration of Helsinki and they were approved by the local ethics review board of the Heinrich-Heine University Düsseldorf. For studies 1 and 2 we did not collect informed consent, as this is allowed by the local ethics committee for online studies, which were completely anonymized, whereas we collected written informed consent in study 3, in the laboratory.

Study 4. After having replicated our results across the three behavioral studies with a within-subject design (see Results), for the fMRI study we estimated, via G*Power, assuming a medium-to-large effect size, that the sample size necessary to achieve a power of 0.95 was *n* = 23. Considering frequent participants' drop out during long scanning sessions as our, or due to excessive movement, we opted for *n* = 40. Forty healthy young volunteers were therefore recruited at the Life&Brain Research Center in Bonn for an fMRI study. All participants met MR-compatible inclusion criteria in addition to no self-reported current or history of neurological or psychiatric disorder, as well as no current use of medication affecting the central nervous system. Due to excessive head motion during measurements (>4mm, >4° rotation, as computed through Artrepair Toolbox; Stanford Psychiatric Neuroimaging Laboratory, see (5, 59)), 10 participants were excluded from all analyses. Thus, the final sample included 30 subjects (21 females; mean age = 25 years, ±4.6, range: 19-35 years) with high-education level (mean education = 14 years, ±1.9, range: 12-18 years, from high school to university master degree). Fifteen participants had a net monthly income between €0 and €499, eight

between €500 and €999, five between €1000 and €1499, one between €1599 and €1999, and one larger than €2500.

All participants were fluent German speakers, right-handed, and had normal or corrected-to-normal vision. As reimbursement, they were paid €20 as participation fee, plus earnings from the social discounting task. Therefore, participants' payoff ranged from €27.5 up to €35.5 (see Social discounting task).

The study was conducted according to the Declaration of Helsinki and it was approved by the local ethics review board of the Universitätsklinikum Bonn. All volunteers gave written informed consent to participate in the study.

Social discounting task

In this task (adapted from (5)), participants were first asked to imagine people from their social environment represented on a scale ranging from 1 (the person socially closest to them) to 100 (a random stranger), where a person at rank 50 was described as a person that the subject had seen several times without knowing the name. They were instructed to select six real persons located at social distances of 1, 5, 10, 20, 50, and 100 (with no need of specifying the name and their social relationship for the social distances 50 and 100). Participants were encouraged to avoid thinking of people that they felt negatively toward and people they shared a bank account or household with. Each trial began with the display of the social distance level of the partner the participant was playing with. Social distance was represented with a ruler scale consisting of 101 icons. The left-most icon, highlighted in purple, depicted the participant. One of the remaining 100 other icons was highlighted in yellow, indicating the social distance of the partner. Furthermore, social distance information was additionally indicated as a number on top of the highlighted yellow icon to prevent perceptual inaccuracies in estimating social distance (cf. Fig. 1 for an example on a partner on social distance 100).

We included two experimental conditions, a *gain frame* and a *loss frame*. The *gain frame* manipulation was near-identical to the task used by (5). Briefly, after presenting the social distance information as described above, participants were instructed that, in this trial, the experimenter gave an initial endowment of €0 the other person ("This person has €0", cf. Fig. 1A). Participants were explicitly and repeatedly instructed that the other person was not aware of her zero endowment, she would only be informed of the final payoff after implementing the participant's choice. Then, two monetary options appeared, a selfish and a generous option. The selfish option (on the left in Fig. 1A, in purple letters) indicated the reward magnitude for the participant, if chosen (e.g., €115 to the participant and no other-reward to other). The generous option contained a smaller own-reward to the participant (€75) and an other-reward to the other person (€75). Own-rewards were always indicated in purple and other-rewards

were always indicated in yellow. Participants indicated their choice of the selfish or the generous alternative by a left or right button press.

In the *loss frame* (Fig. 1B), participants were informed, after the social distance presentation, that the other person has received an initial endowment of €75. As before, participants were explicitly and repeatedly instructed that the other person was not aware of her initial endowment, or the potential loss of it. On the next screen, a selfish (left in Fig. 1B) and a generous alternative (right in Fig. 1B) appeared. When choosing the selfish alternative, the participant received the own-reward amount indicated in purple (here, €115), and the other person lost her initial endowment, as indicated in yellow (-€75), thus leaving her empty-handed. When choosing the generous alternative, the participant received a smaller own-reward indicated in purple (€75), implying that the other person would keep her endowment.

In addition to the framing (gain frame, loss frame) and the social distance levels of the other (1, 5, 10, 20, 50, 100), in each condition, we manipulated the magnitude of the own-reward across trials: we used nine selfish reward amounts per frame condition, ranging from €75 to €155 in steps of €10. The generous alternative's payoff was invariant, always yielding €75 own-reward and €75 other-reward in all conditions and trials.

Thus, in the gain frame condition, the other person always had a €0 endowment, the selfish alternative always yielded a variable own-reward, and no reward for the other, and the generous alternative invariantly yielded an equal €75/€75 split between participant and other person. In the loss frame condition, the other-endowment was always €75, the selfish alternative yielded a variable own-reward accompanied by the loss of the €75 endowment to the other, and the generous alternative always yielded €75 own-reward and had no financial consequences for the other, i.e., she could keep her initial endowment of €75.

To summarize the logic of the task, both frames were mathematically equivalent, i.e., they yielded identical final payoff states to the participant and the other person (in the example in Fig. 1: both frames yield an own-reward gain of €115 to the participant and €0 gain to the other person after a selfish choice, or €75 own-reward and €75 other-reward after a generous choice). The only difference between conditions was that a €0 other-reward outcome was framed as a loss of the initial endowment in the *loss frame* vs. a null-gain in the *gain frame,* and a €75 other-reward was framed as keep-endowment in the *loss frame* vs. a €75 gain in the *gain frame*.

The order of frame conditions, selfish-reward presentations, as well as the left or right screen-position of the selfish and generous alternative were randomized and counterbalanced across trials. The task of studies 1-3 had a total of 108 self-paced trials. The task of study 4 had a total of 216 trials as each trial type was repeated twice to allow for full left/right position counterbalancing. For events duration of study 4, please refer to Fig. 1.

Incentivization procedure. In studies 3 and 4 the social discounting task was fully incentive-compatible. At the end of the session, one of the participant's choices was randomly drawn and 10% of the own-reward amount was paid out, as well as 10% of the other-reward amount was paid out to the other person in that trial, either via cheque, for the other-persons indicated by the participants at social distance 1, 5, 10, 20, or in cash to a random person on site in the case of other-persons at social distance 50 and 100. Note that the recipients of other-reward were only notified in case of a positive payoff, but not in case of a zero payoff or in case a trial was randomly chosen that did not consider them; in addition, they were not informed beforehand about this experiment, and, thus, had no prior outcome expectations. Hence, our incentivization procedure made it logically impossible for the other persons to know about their endowment, or the loss of it.

General procedure

Studies 1-3. In studies 1 and 2, participants were instructed about the social discounting task, and then, after answering comprehension questions, they assigned other persons (i.e., name and personal relationship with them) from their social environment to the social distances 1, 5, 10, and 20, completed the task (see Social discounting task), and finally filled out a questionnaire (see Supplementary information) through Unipark online survey software (Unipark questback). Participants were provided with a web link to do so, after being recruited via flyers and advertisements on social platforms. Monetary payment for study 1 was implemented via Clickworker (clickworker GmbH), whereas university credits reimbursement was carried out on campus for study 2. In study 3, after recruitment, participants were invited to the laboratory, they were instructed on the social discounting task along with the comprehension questions. They then completed the task, implemented in Matlab R2016a (MathWorks) and Cogent toolbox (www.vislab.ucl.ac.uk), and filled out a questionnaire on a laptop. Finally, they were reimbursed for participation contingent on their choices, identical to the incentivization procedure in study 4.

Study 4. Upon arrival, participants received instructions about the social discounting task and then, after applying comprehension questions to check for full understanding of the task, they assigned other persons (i.e., name and personal relationship with them) from their social environment to the social distances 1, 5, 10, and 20 via paper and pencil. Afterwards, participants performed a few sample trials to familiarize with the task structure and they were subsequently cleared for the scanning session. At the end of the scanning session, they answered control questions concerning the social discounting task, and filled out a demographic questionnaire as well as questionnaires assessing social desirability and

empathy (see Supplementary information). Finally, participants were debriefed and received their monetary allowance.

Behavioral data analysis

Hyperbolic model. Similar to previous studies (4, 5, 11, 13, 16), we approximated the participants' decay in generosity across social distance with a hyperbolic function:

$$v = \frac{V}{(1+k*SD)} \qquad \textit{Eq. 1}$$

where v represents the discounted value of generosity, SD represents social distance, k represents the degree of discounting, and V is the intercept at social distance 0, thus the origin of the social discount function (4, 5, 11, 16). While V can be considered an indicator of generosity towards socially close others (11–13), k describes the discount rate, i.e., the steepness by which the social discount function decays across social distance. We estimated k and V for each participant separately, depending on her individual choice pattern.

To estimate V, we titrated the selfish amount to determine, at each social distance, the point at which the subject was indifferent between the selfish and generous options (i.e., indifference point; see also Supplemental results in Supplementary information). Logistic regression, implemented in Matlab R2016a (MathWorks), was used to determine the indifference points where the likelihood of choosing the selfish and the generous options was 50% (5, 16).

To fit Eq. 1 and estimate k, we modeled trial-by-trial choices via a softmax function to compute the probability P of choosing the selected option o_i on a given trial:

$$Poi = \frac{exp\ (voi*ß)}{exp\ (vo1*ß) + exp\ (vo2*ß)} \qquad \textit{Eq. 2}$$

given the subjective values v (based on the current selfish amount and social distance) of the current available options o_1 (vo_1) and o_2 (vo_2) as in Eq. 1. ß is a nuisance parameter reflecting the stochasticity of individual performance. The larger ß, the less noisy the choice pattern. Individual discount rates were defined by the respective k-value that yielded the best prediction of the observed choice probabilities by applying maximum-likelihood estimation using nonlinear optimization procedures (fminsearch function), implemented in Matlab R2016a (MathWorks). To this end, we minimized the log-likelihood of the choice probabilities to obtain the best-fitting k and ß parameter estimates, by summing across trials, given a specific set of model parameters k and ß, the logarithm of $P(o_i)$.

Both variables V and k were analyzed using non-parametric statistics as they were, in most of the cases across the four studies, not normally distributed (Kolmogorov-Smirnov, all ds > 0.20, all ps < 0.05). When participants did not discount at all (i.e., they always chose the generous option), k was set to 0 and V was set to 80 (i.e., maximum reward amount foregone = maximum

selfish amount 155 – generous amount 75). All behavioral analyses were run in Statistica 12 (StatSoft).*

Study 4

fMRI procedures

Magnetic resonance images were collected on a 3T whole-body scanner (Magnetom Trio, Siemens Medical Systems, Erlangen, Germany) with an 8-channel head coil. For functional imaging, gradient-echo echo-planar images (EPI) were acquired at TR = 2500 ms (TE = 30 ms; number of slices = 37; slice thickness = 3 mm; distance factor = 10%; FoV = 192 mm × 192 mm; matrix size = 64 x 64; flip angle = 90°). Slices (voxel size = 2 × 2 × 3 mm) were sampled in transversal orientation covering all of the brain, including the midbrain. The scanning session started with a brief localizer acquisition. Afterwards, functional data were acquired in 3 separate runs of ~421 volumes each, to allow for brief resting periods in between. In order to get information for B_0 distortion correction of the acquired EPI images, a gradient echo field map sequence (TR = 392 ms; TE 1 = 4.92 ms; TE 2 = 7.38 ms; number of slices = 37; voxel size 3 × 3 × 3 mm) was recorded before each functional run. Structural images were collected at the end (~5 min), using a T1-weighted sequence (rapid acquisition gradient echo sequence; 208 sagittal images; voxel size = 0.8 × 0.8 × 0.8 mm; 0.80 mm slice thickness). Head movements were minimized by the use of foam pads and scanner noise was reduced with earplugs. When necessary, vision was corrected-to-normal via fMRI compatible goggles. The social discounting task was programmed via an in-house software and presented via a mirror that projected a screen lying behind the participant, who made their choices via a left and a right button boxes.

fMRI preprocessing

Imaging data were preprocessed and analyzed with Statistical Parametric Mapping (SPM12, Wellcome Trust Centre for Neuroimaging, University College London, UK) implemented in Matlab R2016a (MathWorks). After checking raw data quality for each participant using SPM Check Reg function (Stanford Psychiatric Neuroimaging Laboratory), all images were preprocessed by reorienting them according to the EPI SPM template and coregistered to the fieldmap via FieldMap toolbox. After the functional images were realigned und unwarped to the middle volume and all volumes for participants' motion correction by using phase correction, ArtRepair toolbox (60) was run in order to identify bad volumes. Bad volumes of

* For additional analyses (i.e., indifference points, area under the curve, reaction times, and questionnaires) see Supplementary information.

participants included in the final sample were not repaired. However, we modeled these bad volumes as regressors of no-interest in the statistical analyses (see fMRI analyses). Finally, functional and structural images were coregistered and the images were spatially normalized based on segmentation of the anatomical image with resampling to 2 x 2 x 2 mm, and spatially smoothed using a 6 mm FWHM Gaussian kernel. High-pass temporal filtering (using a filter width of 128 s) was also applied to the data.

fMRI analyses

At the first-level analysis, trial-related activity for each participant (n = 30) was modeled by delta functions convolved with a canonical hemodynamic response function to model the effects of interest, as well as six covariates capturing residual motion-related artifacts, and a temporal derivative for each regressor of interest to account for slice timing differences.

For each participant, relevant contrasts were computed for each general linear model (GLM) and entered into second-level random effect analysis. Whole-brain level results were initially thresholded at $p < 0.001$ (uncorrected), minimum cluster size = 5 voxels, and then corrected for multiple comparisons ($p < 0.05$, family-wise error rate [FWE]). Bad volume onsets (as measured via ArtRepair toolbox; (60)), modeled with a stick function, were included as regressors of no-interest.

The first GLM (GLM1) included an unmodulated regressor of all selfish choices, pooled across frame conditions, and the selfish amount magnitude (see Social discounting task) as trial-by-trial parametric modulator. Four participants could not be included in this analysis as they did not make any selfish choice in one or more runs of the task. Reaction times (RTs) were used as duration to account for differences between gain and loss frames (see Supplemental behavioral results in Supplementary information). Additionally, missed trials were included as regressors of no-interest.

The second GLM (GLM2) searched for differences in BOLD activations between frame conditions during generous choices. It included an unmodulated regressor of all generous choices made in the loss frame and an unmodulated regressor of all generous choices made in the gain frame. Additionally, the selfish amount magnitude (see Social discounting task) was included as trial-by-trial parametric modulator of all main regressors. In the main manuscript, we additionally considered reward foregone as a parametric trial-by-trial regressor. Note that reward foregone is a linear transformation of, and thus collinear with, the selfish reward magnitude; neural activations identified by this parametric regressor might therefore reflect selfish amount or reward foregone (see main text). RTs were used as duration to account for differences between gain and loss frames (see Supplemental behavioral results in Supplementary information). Additionally, missed trials were included as regressors of no-interest.

Furthermore, a third GLM (GLM3) that included an unmodulated regressor of all choices (pooled across generous and selfish) made in the loss frame and an unmodulated regressor of all choices (pooled across generous and selfish) made in the gain frame was built, with the purpose of extracting, at the individual level, beta estimates from an ROI with seed in the insula peak [42, 4, -4] obtained from the previous GLM2, via MarsBaR region of interest toolbox for SPM (61).

We additionally conducted ROI analyses by applying small volume correction (SVC), where relevant, using a 10-mm sphere around the peak coordinates, to check specifically for, and be able to report, the following ROIs: VMPFC [-2, 40, -4] as based on (5, 62); the right TPJ [51, -49, 34; 60, -58, 31; 48, -52, 37] (5) and its homologous left coordinates (63), and the left TPJ [-24, -79, 52] (5); the right insula [42, 8, 0; 40 0 -14; 46, 14, -2; 42, 18, -6; 44, 4, -2] and the left insula [-38, 14, -2; -42, -2, 6] as in (28), and the bilateral insula as based on the Harvard-Oxford Atlas (64).

Results are displayed in the figures overlaid on the averaged normalized structural images of the 30 participants included in the fMRI analyses. The probability maps of the SPM Anatomy Toolbox, Version 3.0 (65), Neurosynth (http://neurosynth.org), and BioImage Suite *Web* (https://bioimagesuiteweb.github.io/webapp/) were used for double checking region localization.

Data availability

Digital Imaging and Communications in Medicine (DICOM) images reported in this paper have been deposited in XNAT Central (https://central.xnat.org/) under the project name ‘Framesocdisc’. Behavioral datasets have been supplied in Figshare (https://figshare.com/) under the doi: 10.6084/m9.figshare.10265309.

References

1. Kalenscher T (2014) Attitude Toward Health Insurance in Developing Countries From a Decision-Making Perspective. *J Neurosci, Psych, Econ* 7(3):174–193.
2. Nowak MA (2006) Five Rules for the Evolution of Cooperation. *Science* 314(5805):1560–1563.
3. Rilling JK, Sanfey AG (2011) The Neuroscience of Social. *Annu Rev Psychol*. 62:23–48.
4. Jones B, Rachlin H (2006) Social discounting. *Psychol Sci* 17(4):283–286.
5. Strombach T, et al. (2015) Social discounting involves modulation of neural value signals by temporoparietal junction. *Proc Natl Acad Sci* 112(5):1619–1624.
6. Vekaria KM, Brethel-Haurwitz KM, Cardinale EM, Stoycos SA, Marsh AA (2017) Social discounting and distance perceptions in costly altruism. *Nat Hum Behav* 1(April):0100.
7. Kalenscher T (2017) Social psychology: Love thy stranger as thyself. *Nat Hum Behav* 1(April):0108.
8. Archambault, C, Kalenscher, T, De Laat J (2019) Generosity and livelihoods: experimental evidence on the multidimensional nature of sharing among the Kenyan Maasai. *J Behav Decis Mak* 1-12.
9. Tiokhin L, Hackman J, Munira S, Jesmin K, Hruschka D (2019) Generalizability is not optional: insights from a cross-cultural study of social discounting. *R Soc Open Sci* 6:181386
10. Strombach T, et al. (2014) Charity begins at home: Cultural differences in social discounting and generosity. *J Behav Decis Mak* 27(3):235–245.
11. Margittai Z, et al. (2015) A friend in need: Time-dependent effects of stress on social discounting in men. *Horm Behav* 73:75–82.
12. Strang S, et al. (2017) Psychoneuroendocrinology Short Communication A matter of distance — The effect of oxytocin on social discounting is. *Psychoneuroendocrinol* 78:229–232.
13. Margittai AZ, Wingerden M Van, Schnitzler A, Joëlsc M, Kalenscher T (2018) Dissociable roles of glucocorticoid and noradrenergic activation on social discounting. *Psychoneuroendocrinol* 90:22-28.
14. Strombach T, Margittai Z, Gorczyca B, Kalenscher T (2016) Gender-specific effects of cognitive load on social discounting. *PLoS One* 11(10):1–15.
15. Soutschek A, et al. (2017) Binding oneself to the mast: stimulating frontopolar cortex enhances precommitment. *Soc Cogn Affect Neurosci* 1–8.
16. Soutschek A, Ruff CC, Strombach T, Kalenscher T, Tobler PN (2016) Brain stimulation reveals crucial role of overcoming self-centeredness in self-control. *Sci*

Adv 2(10):e1600992–e1600992.

17. de Dreu CKW, McCusker C. (1997) Gain-loss frames and cooperation in two-person social dilemmas: A transformational analysis. *J Person Soc Psych* 72(5), 1093-1106.
18. Evans AM, Beest I Van (2017) Journal of Experimental Social Psychology Gain-loss framing e ff ects in dilemmas of trust and reciprocity. *J Exp Soc Psychol* 73(July):151–163.
19. Everett JAC, Faber NS, Crockett MJ (2015) The influence of social preferences and reputational concerns on intergroup prosocial behaviour in gains and losses contexts. *R Soc Open Sci* 2(12):150546.
20. Sip KE, Smith D V., Porcelli AJ, Kar K, Delgado MR (2015) Social closeness and feedback modulate susceptibility to the framing effect. *Soc Neurosci* 10(1):35–45.
21. Smith D V., Sip KE, Delgado MR (2015) Functional connectivity with distinct neural networks tracks fluctuations in gain/loss framing susceptibility. *Hum Brain Mapp* 36(7):2743–2755.
22. Wang XT, Rao L-L, Zheng H (2016) Neural substrates of framing effects in social contexts: A meta-analytical approach. *Soc Neurosci* (March):1–12.
23. Xiao F, et al. (2016) Who are you more likely to help? the effects of expected outcomes and regulatory focus on prosocial performance. *PLoS One* 11(11):1–15.
24. Zheng H, Wang XT, Zhu L (2010) Framing effects: Behavioral dynamics and neural basis. *Neuropsychologia* 48(11):3198–3204.
25. Li R, et al. (2017) Reason's Enemy Is Not Emotion: Engagement of Cognitive Control Networks Explains Biases in Gain/Loss Framing. *J Neurosci* 37(13):3486–16.
26. Bardsley N (2008) Dictator game giving: Altruism or artefact? *Exp Econ* 11(2):122–133.
27. Baumeister RF, Stillwell AM, Heatherton TF (1994) Guilt: An Interpersonal Approach. *Psych Bulletin* 115(2):243–267.
28. Chang LJ, Smith A, Dufwenberg M, Sanfey AG (2011) Article Triangulating the Neural, Psychological, and Economic Bases of Guilt Aversion. *Neuron* 70(3):560–572.
29. Chang LJ, Sanfey AG (2013) Great expectations : neural computations underlying the use of social norms in decision-making. *Neuron* 70, 560–572
30. Crockett MJ, et al. (2014) Harm to others outweighs harm to self in moral decision making. *Proc Natl Acad Sci* 112(4).
31. List J (2007) On the Interpretation of Giving in Dictator Games. *J Polit Econ* 115(3):482–493.
32. Mobbs D, et al. (2009) A key role for similarity in vicarious reward. *Science (80)* 324(5929):900.
33. Bartra O, Mcguire JT, Kable JW (2013) NeuroImage The valuation system : A

coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 76:412–427.

34. Tusche XA, Bo A, Kanske XP, Trautwein XF, Singer T (2016) Decoding the Charitable Brain : Empathy , Perspective Taking , and Attention Shifts Differentially Predict Altruistic Giving. *J Neurosci* 36(17):4719–4732.

35. Gu X, et al. (2015) Necessary , Yet Dissociable Contributions of the Insular and Ventromedial Prefrontal Cortices to Norm Adaptation : Computational and Lesion Evidence in Humans. *J Neurosci* 5(2):467–473.

36. Lallement JH, et al. (2013) Effort increases sensitivity to reward and loss magnitude in the human brain. *Soc Cogn Affect Neurosci.*

37. Oldham S, Murawski C, Fornito A, Youssef G, Lorenzetti V (2018) The anticipation and outcome phases of reward and loss processing : A neuroimaging meta-analysis of the monetary incentive delay task. *Hum Brain Map* 3398–3418.

38. Samanez-larkin GR, Hollon NG, Carstensen LL, Knutson B (2008) Individual Differences in Insular Sensitivity During Loss Anticipation Predict Avoidance Learning. *Psychol Sci* 19(4):320–323.

39. Siebenthal Z Von, et al. (2017) Decision-making impairments following insular and medial temporal lobe resection for drug-resistant epilepsy. *Soc Cogn Affect Neurosci* 128–137.

40. Yu H, Hu J, Hu L, Zhou X (2014) The voice of conscience : neural bases of interpersonal guilt and compensation. *Soc Cogn Affect Neurosci* 1150–1158.

41. Bellucci G, Feng C, Camilleri J, Eickhoff SB, Krueger F (2018) The role of the anterior insula in social norm compliance and enforcement: Evidence from coordinate-based and functional connectivity meta-analyses. *Neurosci Biobehav Rev* 92(March):378–389.

42. Montague PR, Lohrenz T, Humphrey H (2007) Minireview To Detect and Correct : Norm Violations and Their Enforcement Minireview. *Neuron* 56, 14–18.

43. Hutcherson CA, Bushong B, Rangel A (2015) A Neurocomputational Model of Altruistic Choice and Its Implications. *Neuron* 87(2):451–463.

44. Vlaev I (2012) How different are real and hypothetical decisions ? Overestimation , contrast and assimilation in social interaction. *J Econ Psychol* 33(5):963–972.

45. Sthi R, Somanathan E (2016) American Economic Association The Evolution of Social Norms in Common Property Resource Use. *The American Economic Review* 86(4):766–788.

46. Paulus MP, Rogalsky C, Simmons A, Feinstein JS, Stein MB (2003) Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism. *Neuroimage* 19:1439–1448.

47. Huggins AA, Belleau EL, Miskovich TA, Pedersen WS, Larson CL (2018) Moderating Effects of Harm Avoidance on Resting-State Functional Connectivity of the Anterior Insula. *Front Hum Neurosci* 12(November):1–9.
48. Seara-cardoso A, et al. (2016) Anticipation of guilt for everyday moral transgressions : The role of the anterior insula and the influence of interpersonal psychopathic traits. *Sci Rep* 6:36273.
49. Wagner U, N'Diaye K, Ethofer T, Vuilleumier P (2011) Guilt-specific processing in the prefrontal cortex. *Cereb Cortex* 21(11):2461–2470.
50. Singer T, et al. (2006) Empathic neural responses are modulated by the perceived fairness of others. *Nat Letters* 439(January):10–13.
51. Canessa N, et al. (2017) Neural markers of loss aversion in resting-state brain activity. *Neuroimage* 146(September):257–265.
52. Canessa N, et al. (2013) The Functional and Structural Neural Basis of Individual Differences in Loss Aversion. *J Neurosci* 33(36):14307–14317.
53. Sokol-hessner P, Camerer CF, Phelps EA (2013) Emotion regulation reduces loss aversion and decreases amygdala responses to losses. *Soc Cogn Affect Neurosci* 8, 341:350
54. Sokol-hessner P, Rutledge RB (2019) The Psychological and Neural Basis of Loss Aversion. *Curr Dir Psych Sci* 8(1) 20–27.
55. Lamm C, Decety J, Singer T (2011) NeuroImage Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* 54(3):2492–2502.
56. Zhang X, Liu Y, Chen X, Shang X, Liu Y (2017) Decisions for others are less risk-averse in the gain frame and less risk-seeking in the loss frame than decisions for the self. *Front Psychol* 8(SEP):1–10.
57. Sellitto M, Ciaramelli E, Mattioli F, di Pellegrino G (2016) Reduced sensitivity to sooner reward during intertemporal decision-making following insula damage in humans. *Front Behav Neurosci* 9:367.
58. Belfi AM, Timothy R. Koscik, Tranel D (2015) Damage to the insula is associated with abnormal interpersonal trust. *Neuropsychologia* 71:165–172.
59. Wendelken C, O'Hare ED, Whitaker KJ, Ferrer E, Bunge SA (2011) Increased Functional Selectivity over Development in Rostrolateral Prefrontal Cortex. *J Neurosci* 31(47):17260–17268.
60. Mazaika PK, Hoeft F, Glover GH, Reiss AL. Methods and Software for fMRI Analysis for Clinical Subjects. 77309.
61. Brett M, Anton JL, Valabregue R, Pauline J-B (2002) Region of interest analysis using an SPM toolbox. *8th International Conferance on Functional Mapping of the Human*

Brain.

62. Clithero JA, Rangel A (2013) Informatic parcellation of the network involved in the computation of subjective value. *Soc Cogn Affect Neurosci* 9(9):1289–1302.

63. 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.

64. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) Fsl. *Neuroimage* 62(2):782–790.

65. Eickhoff SB, et al. (2007) Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *Neuroimage* 36(3):511–521.

Figures

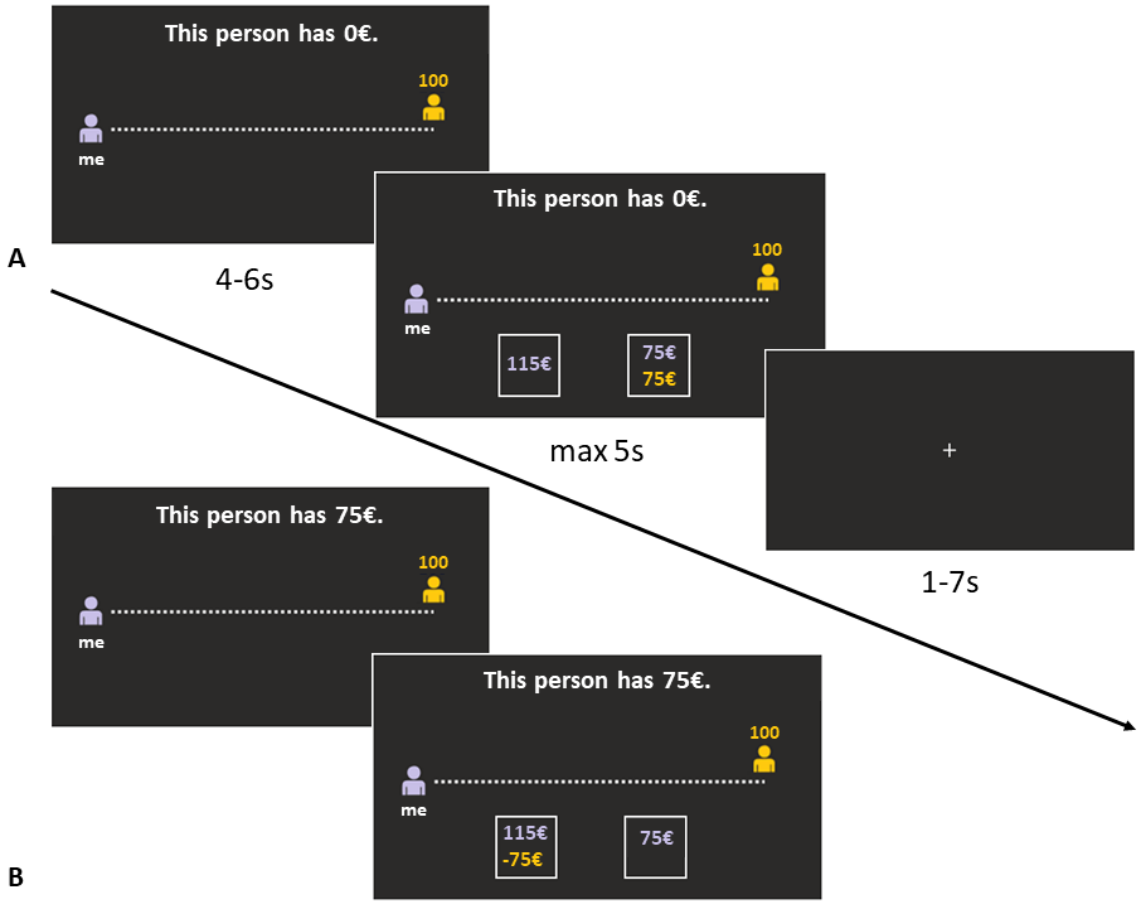

Figure 1. Social discounting task (fMRI study 4). A. Trial example of the *gain frame*. **B.** Trial example of the *loss frame*. Each trial started with the presentation of a ruler-based representation of social distance to the other-person, with a left-most purple icon representing the participant and a yellow icon indicating the social distance of the other-person in the current trial (100 in this example). Additionally, participants received information as to the endowment of the other person, i.e., "This person has €0" for the gain frame (**A**), or "This person has €75" for the loss frame (**B**) (4-6s). Afterwards, the two choice options appeared. The selfish alternative was displayed in purple fonts, indicating the own-reward magnitude to the participant (here: €115). Selfish choices implied a null gain for the partner in the gain frame (**A**) or the loss of the initial €75-endowment (in yellow) for the other person in the loss frame (**B**). The generous alternative was displayed in yellow fonts, and always yielded an equal €75 own-reward and €75 other-reward split in the gain frame **(A)**, or a €75 own-reward gain and the possibility to keep the €75 other-endowment in the loss frame **(B)**. As soon as the two choice options appeared, participants had 5s to choose one of the two alternatives. After a choice was made, or after the 5s were passed, a blank screen with a fixation cross appeared (1-7s), and then a new trial started.

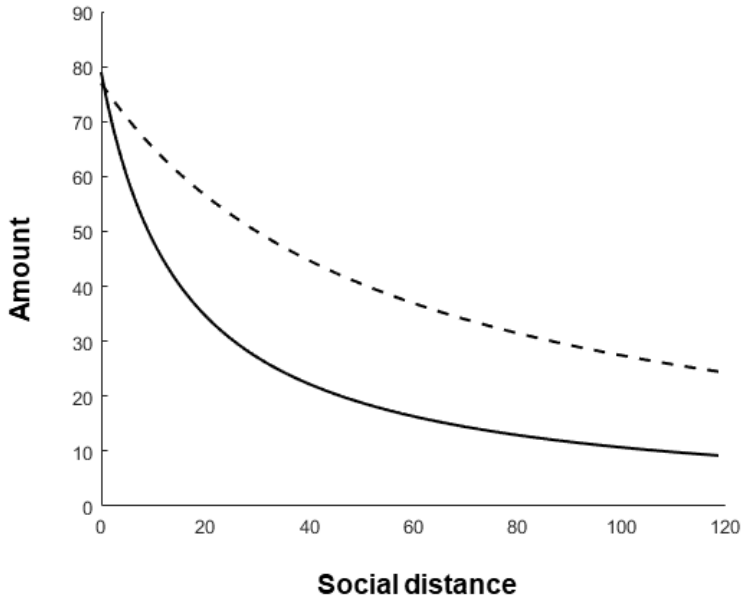

Figure 2. Hyperbolic discounting function (fMRI study 4). The change in generosity across social distances, measured as the social-distance-dependent reward amount that participants were willing to pay to increase the wealth of another person, was captured by a hyperbolic discount model (see main text for details). The figure shows the mean best-fitting hyperbolic functions, averaged across participants, computed separately for the gain frame and the loss frame. The social discounting curve for the loss frame (**dashed line**) was significantly flatter than the social discounting curve for the gain frame (**solid line**).

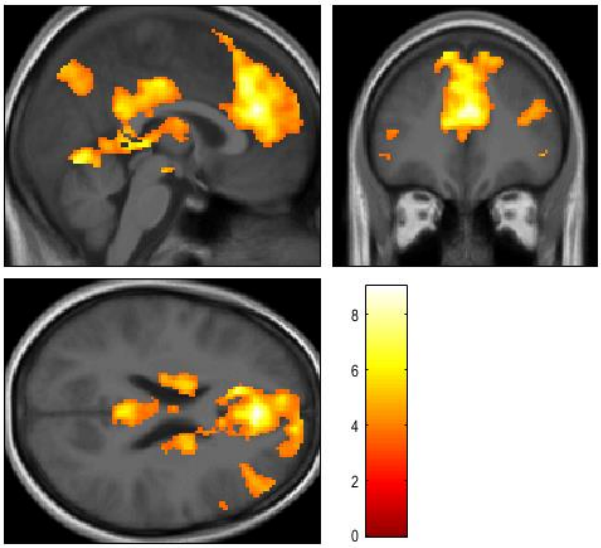

Figure 3. Selfish choices recruit the valuation network independent of frame. Selfish choices, pooled across gain and loss frames, activated the valuation network ($p < 0.05$ FWE whole-brain corrected; for illustration purposes, activations are displayed at $p < 0.001$, uncorrected, minimum cluster size ≥ 5).

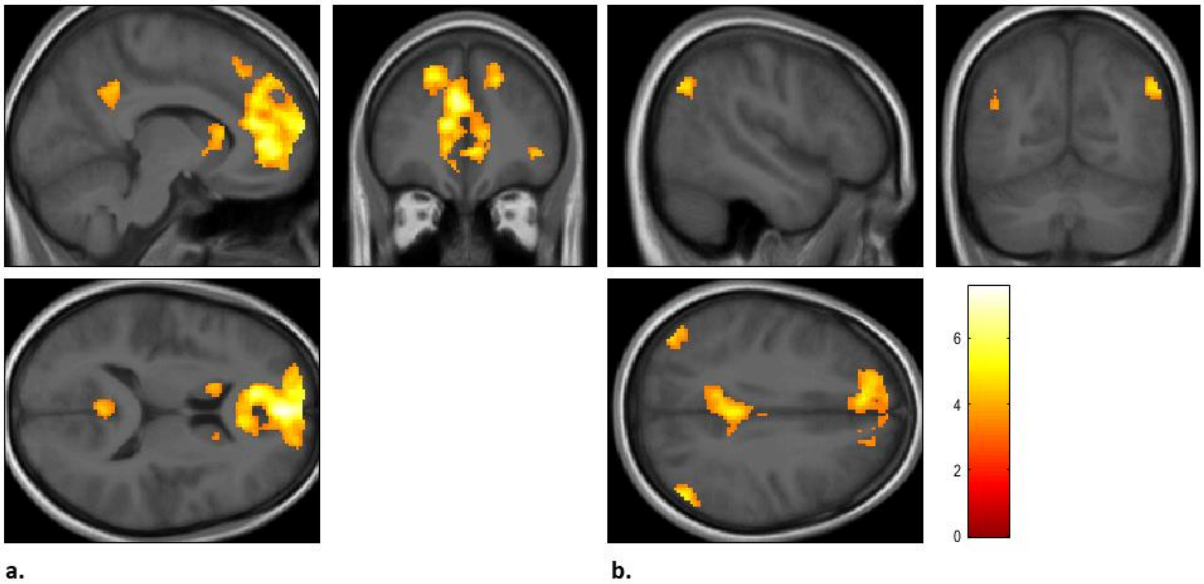

Figure 4. Generous choices in the gain frame correlate with VMPFC and TPJ activity. VMPFC and caudate (**a**) as well as right TPJ (**b**) were selectively activated during [generous choice in gain frame > generous choice in loss frame] ($p < 0.05$ FWE whole-brain corrected; for illustration purposes, activations are displayed at $p < 0.001$, uncorrected, minimum cluster size ≥ 5).

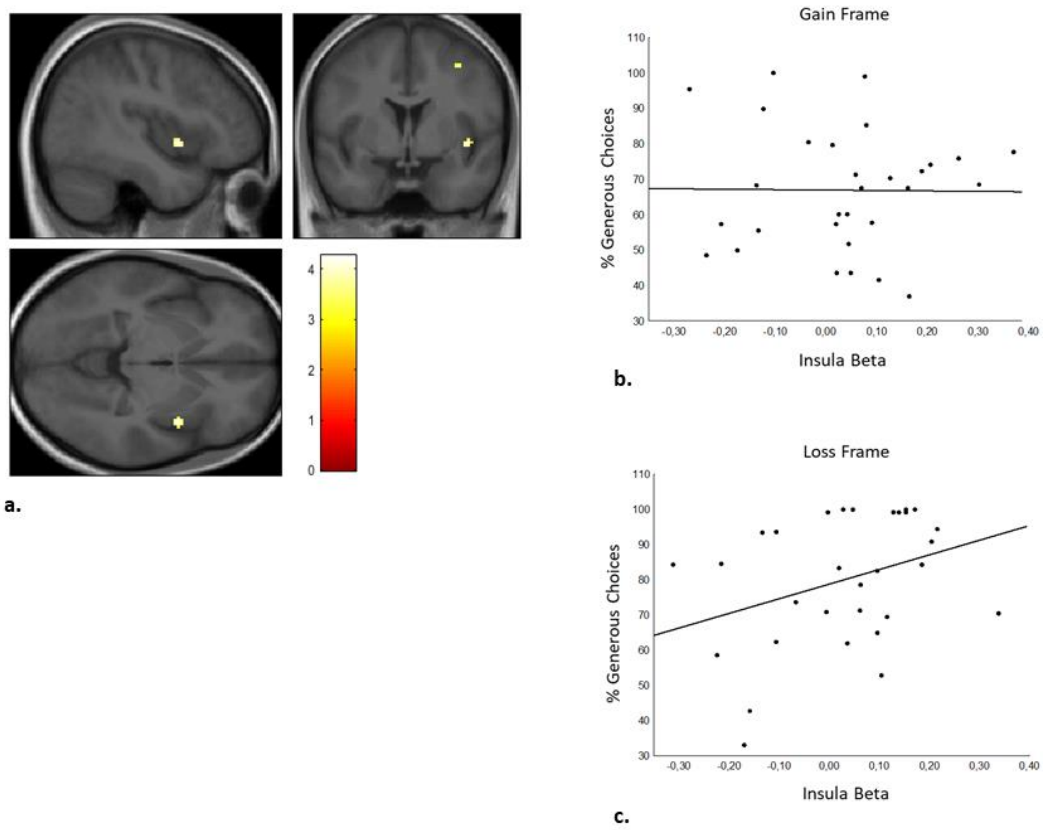

Figure 5. Insula activation underlies generous choice in the loss frame. (**a**) Insula was selectively activated during [generous choice in loss frame > generous choice in gain frame] ($p < 0.05$ ROI FEW-corrected; for illustration purposes, activations are displayed at $p < 0.001$, uncorrected, minimum cluster size ≥ 5). Parameter estimates of insula activation [42, 4, -4] were extracted separately for the gain and the loss frame, after pooling all choices (i.e., both generous and selfish) within each frame. Insula activation did not correlate significantly with the percentage of generous choices made in the gain frame (**b**) but in the loss frame (**c**) only.

Supplementary information for

To give or not to take – insular cortex subserves framing-induced boosts in generosity during social discounting

Manuela Sellitto, Adam Schweda, Bernd Weber, Tobias Kalenscher

Corresponding author: Manuela Sellitto

E-mail: manuela.sellitto@hhu.de

This PDF file includes:

Supplemental text
Figure S1
Tables S1 and S2

Supplemental materials and methods

Studies 1-3

Questionnaire. At the end of the social discounting task, participants completed the Social Desirability Scale (SDS-17; (1)), a 16-item self-report questionnaire measuring the tendency to describe oneself with socially desirable attributes.

Study 4

Control questions. At the end of the scanning session, we asked participants to indicate, on a 7-point Likert scale (1 = not at all; 7 = definitely yes), their relationship with the persons they assigned to the various social distance levels during the social discounting task. These items included:

Item 1: “I feel very close to (person at social distance X)”;

Item 2: “In general, I have a very good relation with (person at social distance X)”;

Item 3: “I am very familiar with (person at social distance X)”.

Questionnaires. At the end of the scanning session, in addition to the SDS-17 (1), participants completed the Interpersonal Reactivity Index (IRI (2, 3)), a 16-item self-report questionnaire measuring the ability to empathize with others.

Supplemental behavioral data analysis

Studies 1-4

Area under the curve (AUC). We also characterized social discounting behavior using a model-free measure, the AUC of the empirically derived discount function, i.e., the curve plotting the indifference points against social distance (see (57, 66)). The larger the AUC values, the larger the preference for generous over selfish options.

Supplemental behavioral results

Study 1

Hyperbolic model. It is not possible to model the behavior of participants with null discounting, i.e., participants who always chose the generous option across all social distance levels. In those cases, we replaced the k value with 0 and the V value with 80 (for justification and reasoning, see Behavioral data analysis in the main text). However, this practice might have biased our results in favor of our hypothesis. We thus repeated the analyses on the differences in the k and V parameters between frame conditions, but this time treating those k and V parameters of participants with zero discounting as missing values. There was still a significant difference in the k parameter between frame conditions (median k_{gain} = 0.020 vs. median k_{loss} = 0.010; Z = 2.59, $p < 0.009$; $\eta_p^2 = 0.49$) but no difference in the V parameter (median V_{gain} =

87 vs. median V_{loss} = 83; Z = 1.10, p = 0.27), supporting our conclusion that social discounting was flatter in the loss frame than the gain frame. The same pattern of results was found in studies 2-4.

Studies 1-3

Questionnaire. We pooled the SDS-17 scores across studies 1-3, and tested whether social desirability explained V nor k across frame conditions. However, the correlation did not reach significance (both ps > 0.25).

Studies 1-4

Indifference points. Our behavioral results can be seen even more clearly when comparing the indifference points between frame conditions –the selfish amount at which participant is indifferent between the selfish option and the generous option– at each social distance (see the Behavioral data analysis in the main manuscript for calculation details). In all four studies, indifference points were significantly higher in the loss compared to the gain frame for social distance 10, 20, 50, and 100 (Wilcoxon matched pairs test: all Zs > 2.61, all ps < 0.008, Bonferroni corrected), with the exception of social distance 10 in study 3 (p = 0.07), confirming the increased tendency in the loss frame to act generously toward socially remote individuals.

AUC. AUCs calculated for both frame conditions were significantly larger in the loss than in the gain frame across all four studies (Wilcoxon matched pairs test: all Zs > 4.16, all ps < 0.0001), further corroborating the conclusion of a frame-dependent difference in generosity.

Study 4

RTs. Mean RTs were normally distributed (Kolmogorov-Smirnov test, all ds < 0.15, all ps > 0.20). An ANOVA on RTs with frame (gain, loss) and choice (selfish, generous) as within subject factors yielded a significant effect of frame ($F_{1,25}$ = 13.14, p < 0.002, η_p^2 = 0.34), a significant effect of choice ($F_{1,25}$ = 15.58, p < 0.001, η_p^2 = 0.38), but no significant interaction frame × choice ($F_{1,25}$ = 1.18, p = 0.29). Bonferroni-corrected post hoc comparisons showed that participants had slower responses in the loss frame compared to the gain frame (2064.5 ms vs. 1894.0 ms), and that they had slower RTs when making selfish compared to generous decisions (2117.4 ms vs. 1841.1 ms). These results indicate that it was generally more difficult to make a selfish choice, and that it was generally more difficult to make decisions in the loss than the gain frame (see also (26) for a similar pattern of results). We considered RTs as events duration in our fMRI analyses to, thus, avoid potential choice-related RT confounds.

Control questions. We included a range of control questions in study 4 that probed the nature of the relationship of the participants to the respective people assigned to the various social distance levels. As social distance (SD) 1, ten participants indicated their mother/father, nine indicated their sibling, seven indicated their partner/husband/wife/fiancé(e), three indicated a close friend (non-sexual), and one defined the relationship to SD1 as 'other' (i.e., not included in the given options). We further asked participants to indicate their relationship with the person on SD 1 on a 7-point Likert scale (1 = not at all; 7 = definitely yes). Participants rated item 1 ("I feel very close to (person at SD1)") with an average score of 6.0 (±1.9 standard deviation), item 2 ("In general, I have a very good relation with (person at SD1)") with 6.3 (±1.7), and item 3 ("I am very familiar with (person at SD1)") with 6.0 (±1.9).

As SD5, eighteen participants indicated a close friend (non-sexual), four indicated their mother/father, three indicated their sibling, four indicated their partner (two of which reported to be in an open relationship), and one indicated his/her uncle/aunt. Participants rated item 1 with a mean score of 5.7 (±1.8), item 2 with 6.2 (±1.5), and item 3 with 5.6 (±1.5).

As SD10, nine participants indicated a close friend (non-sexual), nine indicated their partner, two indicated their sibling, two indicated an acquaintance, two indicated their grandparent, two indicated a colleague, one indicated his/her roommate, and one indicated his/her father/mother. Two participants defined the relationship to SD10 as other. Participants rated item 1 with a mean score of 4.3 (±1.5), item 2 with 5.6 (±1.5), and item 3 with 5.0 (±1.7).

As SD20, twelve participants indicated a colleague, six indicated their partner, five indicated an acquaintance, two indicated their mother/father, one participant indicated his/her cousin, one indicated a close friend (non-sexual), and one indicated his/her roommate. Two participants defined the relationship to SD20 as other. Participants rated on average item 1 with 3.6 (±1.5), item 2 with 5.3 (±1.3), and item 3 with 4.0 (±1.7).

Bonferroni-corrected comparisons revealed that the ratings of all three items significantly decreased from each social distance level to the next (all ps < 0.0001), with the exception of SD1 and SD5 (no item-score was significantly different; all ps > 0.13), and SD10 and SD20 (item 2 only reached trend level with p = 0.06). These results confirm that participants understood the social distance concept, followed the instructions of the social discounting task, and were sensitive to the differences between the SD levels.

Questionnaires. Social desirability (as measured via the SDS-17 (1)), and perspective taking, empathic concern, personal distress, and fantasy (as measured via the IRI (2, 3)) did neither correlate with the V nor with the k parameters in neither frame condition (all ps > 0.05).

Supplemental references

1. Stöber J (2001) The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *Eur J Psychol Assess* 17(3):222–232.

2. Paulus C (2006) Der Saarbrücker Persönlichkeitsfragebogen (SPF-IRI) zur Messung von Empathie. 1–11.

3. Davis MH (1983) A Mulitdimensional Approach to Individual Differences in Empathy. *J Pers Soc Psychol* 44(1):113–126.

4. Sellitto M, Ciaramelli E, Di Pellegrino G (2010) Myopic discounting of future rewards after medial orbitofrontal damage in humans. *J Neurosci* 30(49):16429 –16436.

Supplemental figure

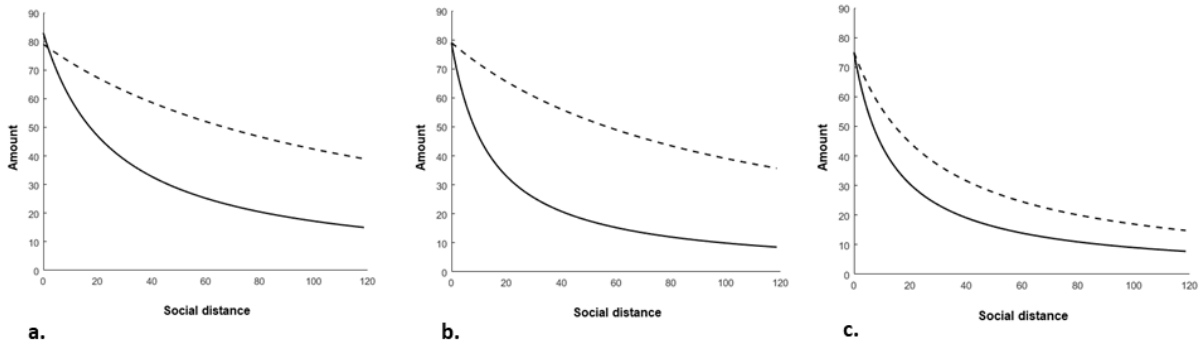

Figure S1. Hyperbolic discounting parameter k. The figure shows the mean best-fitting hyperbolic function across all participants of Study 1 (**a**), Study 2 (**b**), and Study 3 (**c**), separately for the gain frame (**solid line**) and the loss frame (**dashed line**). In all three studies, the social discount function was consistently flatter in the loss frame than the gain frame.

Supplemental tables

GLM1							MNI coordinates		
	Region	Side	$p_{FWE\text{-}corr}$	Cluster size	p(unc.)	Z-score	x	y	z
	Midbrain	L	0.000	16664	0.000	5.96	-8	-12	-14
	VMPFC	L/R				5,87	2	36	22
	Thalamus	R				5,80	22	-24	12
	Parietal cortex	L	0.000	1287	0.000	5,55	-50	-54	42
		L				4.68	-36	-58	36
		L				4.37	-46	-66	42
		R	0.000	1762	0.000	5.41	50	-54	50
		R				5.11	54	-56	42
		R				5.09	46	-50	42
	Frontal pole	R	0.000	993	0.000	4.99	48	42	-6
		R				4.87	36	48	4
		R				4.81	32	56	2
		L	0.000	548	0.000	4.89	-34	54	-2
		L				4.19	-42	52	10
		L				3.97	-40	46	2
	Middle frontal gyrus	L	0.003	203	0.000	4.74	-44	18	48
		L				4.12	-40	14	42
	Precuneous	L/R	0.000	449	0.000	4.14	2	-78	46
		L/R				4.12	0	-68	36
		L				3.87	-6	-62	36

Table S1. Activations at selfish choice onset (button press), pooled across loss and gain frame. Significant clusters active at $p < 0.001$ (unc.), whole-brain cluster-level-corrected $p_{FWE\text{-}corr} < 0.05$. R = right; L = left; VMPFC = ventromedial prefrontal cortex.

GLM2							MNI coordinates		
	Region	Side	$p_{FWE\text{-}corr}$	Cluster size	p(unc.)	Z-score	x	y	z
generous choice in gain frame > generous choice in loss frame	VMPFC	L/R	0.000	5424	0.000	6.59	0	54	14
	Frontal pole	L/R				6.16	0	60	6
	VMPFC	L				5.89	-4	38	28
	PCC	L/R	0.000	883	0.000	5.01	0	-40	34
		L				4.83	-4	-48	24
		L				4.24	-10	-48	36
	Parietal cortex	R	0.034	124	0.002	4.65	50	-66	36
		L	0.021	139	0.001	4.35	-44	-74	36
		L				3.59	-44	-64	26
	Caudate	L	0.012	158	0.000	4.29	-10	16	10
		L				3.94	-14	10	16
		L				3.83	-12	8	-2
	Middle temporal gyrus	L	0.028	130	0.002	4.17	-64	-20	-8
		L				3.92	-60	-12	-10
		L				3.92	-62	-28	-10
generous choice, loss frame > gain frame	Precuneous	R	0.000	284	0.000	4.05	8	-54	54
		R				3.89	10	-62	62
		R				3.63	12	-70	58
		L	0.004	194	0.000	3.94	-8	-66	58
		L				3.81	-8	-54	64
		L				3.72	-12	-50	54

Table S2. Activations at generous choice onset (button press). Significant clusters active at $p < 0.001$ (unc.), whole-brain cluster-level-corrected $p_{FWE\text{-}corr} < 0.05$. R = right; L = left; VMPFC = ventromedial prefrontal cortex; PCC = posterior cingulate cortex.