A neurocomputational account of the link between social perception and social action

Lisa M. Bas [1, *, †], Ian D. Roberts [2 *], Cendri A. Hutcherson [2, 3 ^], Anita Tusche [1 ^]

* Authors contributed equally (shared first authorship)

^ Authors contributed equally (shared last authorship)

[1] Department of Psychology, Queen's University, Kingston, Canada

[2] Department of Psychology, University of Toronto Scarborough, Toronto, Canada

[3] Department of Marketing, Rotman School of Management, University of Toronto, Toronto, Canada

† Corresponding author:

Dr. Lisa M. Bas

Queen's University, 99 University Ave, Kingston, ON K7L 3N6

lisa.bas@web.de

RUNNING HEAD: Social perceptions guide prosocial actions

Abstract

People selectively help others based on perceptions of their merit or need. Here, we develop a neurocomputational account of how these social perceptions translate into social choice. Using a novel fMRI social perception task, we show that both merit and need perceptions recruited the brain's social inference network. A behavioral computational model identified two non-exclusive mechanisms underlying variance in social perceptions: a consistent tendency to perceive others as meritorious/needy (bias) and a propensity to sample and integrate normative evidence distinguishing high from low merit/need in other people (sensitivity). Variance in people's merit (but not need) bias and sensitivity independently predicted distinct aspects of altruism in a social choice task completed months later. An individual's merit *bias* predicted *context-independent* variance in people's overall other-regard during altruistic choice, biasing people towards prosocial actions. An individual's merit *sensitivity* predicted *context-sensitive* discrimination in generosity towards high and low merit recipients by influencing other-regard and self-regard during altruistic decision-making. This context-sensitive perception-action link was associated with activation in the right temporoparietal junction. Together, these findings point towards stable, biologically based individual differences in perceptual processes related to abstract social concepts like merit, and suggest that these differences may have important behavioral implications for an individual's tendency toward favoritism or discrimination in social settings.

Introduction

Psychologists and economists have long sought to explain when and why people help. While people are generally altruistic, they also show selectivity, being more likely to assist those in need [1-5] and to withhold aid from those perceived as undeserving [6-10]. Need and deservingness (merit) are two distinct principles of morality. The need principle involves distributing resources to those who require them, irrespective of whether they have earned them, while the "merit principle" focuses on allocating resources based on individuals' deservingness, regardless of their actual need [11]. How do people assess whether others deserve or need help, and how does this influence their choices to initiate that help?

Based on an analogy to perception in basic sensory domains like vision [12,13], we hypothesized that variance in social perceptions might be driven by two discrete mechanisms: an individual's *sensitivity* to social situational cues signaling other's merit or need, and an individual's *bias* to perceive merit or need independently of specific cues. In other words, one can think of the likelihood *P* that person *J* perceives a target individual as deserving (meritorious) or in need of aid as being determined by the sum of all cues $C_1, C_2, \ldots C_N$ associated with that perceptual judgment, multiplied by a person's idiosyncratic sensitivity *S* to each cue, and added to their baseline tendency (*bias*) to perceive others as deserving or in need:

$$P(Judgment|J, C) \sim \sum_{i=1}^{N} S_{i,J} C_i + bias_J \qquad \text{(Eq. 1)}$$

Need-signaling cues could include facial, vocal, or postural cues of pain or distress [14-17], or cues implying imminent harm (e.g., a person in front of a runaway car [18,19]). Merit-relevant cues could include membership in 'good' or 'bad' groups (e.g., children [20], one's in-group [21], Nazi Party membership [22]) or information about the social normativity or benevolence of that person's actions [23]. An individual with higher average *sensitivity* to the cues associated with that judgment would

discriminate more between those who are high or low in merit or need (as signaled by available cues). An individual with a larger *bias* term would simply be more likely to judge *all* individuals as meritorious/needy irrespective of present cues. Whether processing and integrating these social cues, along with the resulting judgments, involves distinct neural modules [24] or a general-purpose mechanism for social cognition [25], such as the mentalizing network [26-29], remains unknown.

A similar question arises over *how* perceptions of need and merit influence meaningful *behavior*. While extensive research shows that they do [3-6,8-10,30-32], we know little about the precise computational mechanism underlying this perception-action link. We speculated that it might operate by changing the value that people ascribe to different social considerations during choice. More specifically, we build on a growing body of work suggesting that prosocial actions can be characterized as a value-based decision process [33-36]. In this framework, the value of acting prosocial is represented as a weighted sum of several attributes, like the cost of helping for the decision-maker, the benefits for others, or the fairness of the outcome. These three attributes have been repeatedly shown to guide social behaviors [37-43], with higher weights on prosocial attributes like benefits for others or fairness yielding more prosocial choices [35,44-49]. Within this computational framework, perceptions of others' merit or need could affect social behaviors through a simple mechanism: by altering the weights given to outcomes for self, outcomes for others, or fairness during the decision process. For example, perceiving someone as highly deserving or needy could increase helping by increasing the weight given to the recipient's benefits, decreasing the weight given to one's own benefits, or both.

This computational framework leads to a set of testable predictions about how individual differences in social perceptual sensitivity and bias (Eq. 1) might influence altruistic behavior: An individual's general tendency to perceive merit or need (i.e., their *bias* parameter) should be

correlated with a general tendency to ascribe higher weight to prosocial considerations (i.e., others' benefits or fairness). In contrast, an individual's sensitivity to cues signaling merit or need (i.e., their average *sensitivity* parameter *S*) should make them more discriminatory, increasing weights on prosocial considerations for those judged as deserving (or needy) but decreasing them for those judged as lacking these qualities. These weights should then manifest in the frequency with which an individual acts generously, either overall or as a function of the recipient's merit and/or need.

To test these hypotheses, and to uncover their neural basis, we first asked participants to complete a novel social perception task while we collected their brain responses using fMRI (functional magnetic resonance imaging). The behavior observed in this task allowed us to computationally estimate two distinct – but not mutually exclusive – processes underlying individual differences in merit/need perception: an overall propensity to perceive others as deserving/in need (i.e., an individual's *bias* parameter) and a tendency to sample and integrate cues signaling merit/need (i.e., an individual's *sensitivity* parameter). Neurally, we investigated whether merit and need are perceived through distinct or common neural circuits and reflect individual differences in people's perceptual bias and/or sensitivity. Examining need and merit concurrently in this task will also help clarify the computational and neural underpinnings of these related, but distinct concepts, distinguishing between them more effectively. Second, we used a separate altruistic choice task (completed on average ~303 days later) to computationally estimate the weights people place on themselves, others, and fairness when deciding whether to provide aid, and we examined how those weights varied as a function of recipient merit and need. Third, we tested whether individual differences in merit/need bias or sensitivity and their neural underpinnings in the social perception task predict individual differences in prosocial action in the altruism task. Our results suggest that merit (but perhaps not need) perceptions result from stable

individual differences in both bias and sensitivity and manifest in an individual's prosocial behavior months later.

Materials and Methods

Participants

This research took place in the context of a large project that recruited participants from the broader Los Angeles metropolitan area to come to the lab on multiple days to complete different behavioral and neuroimaging tasks related to social cognition [50]. We recruited 50 participants from this larger project pool to complete a newly developed fMRI social perception task. All participants were right-handed, had normal or corrected-to-normal vision, spoke English fluently, and had IQ scores in the normal range (20 females; mean age = 33 years, range = 19–49; full-scale IQ = 105.18 ± 8.04 (mean ± std), range = 87–127 [51,52]). Of these 50 participants, we excluded one who fell asleep and five for excessive movement (framewise displacement > 0.3 mm on over 30% of frames and visual spikes), yielding a total sample of n = 44 for this task. We also recruited 42 participants from the larger participant pool to complete the altruism task (13 females; mean age = 34 years, range = 19–49; full-scale IQ = 104.79 ± 7.97, range = 87–127). Of these 42 participants, we excluded one individual who fell asleep, five for excessive movement during fMRI data collection, three for invariant behavioral responses (identical left/right button press in > 90% of trials, indicating inattention to monetary offers in the altruism task), and five for manipulation check failures (e.g., failing memory checks regarding partner behavior in the altruism task, see below). This procedure yielded a sample of n = 28 for the altruism task, with an overlap of 25 individuals who successfully completed both sessions (social perception and altruism task). Thus, we report

results regarding the social perception task for 44 participants (27 males; mean age = 34 years, range = 19–49; full-scale IQ = 105.84 ± 7.85, range = 90–127), the altruism task for 28 participants (19 males; mean age = 35 years, range = 25–49; full-scale IQ = 106.36 ± 7.94, range = 94–127), and cross-session results (comparing data across both tasks) for the sample of 25 individuals with valid data in both (17 males; mean age = 35 years, range = 25–49; full-scale IQ = 106.60 ± 8.26, range = 94–127). Participants in our overlap sample (valid data in both tasks) had an average separation between both tasks of 303 days (range: 27-663). This delay minimized the risk of sequential dependencies between tasks and increased confidence in the temporal stability of relationships between social perception and social action. Participants received $20 per hour for each experimental session and an additional amount based on one randomly selected trial in the altruism task to incentivize choices and ensure that participants' responses reflected their actual preferences. All participants provided written informed consent according to a protocol approved by the Institutional Review Board of the California Institute of Technology (#12-0343).

Social Perception Task

To assess individual differences in social perceptions, we used a modified version of the established fMRI why/how task [53,54]. Participants viewed images of people in complex real-world scenes and made rapid yes/no judgments (button presses) while their brain responses were measured using fMRI. In separate blocks (Figure 1A), participants made judgments regarding others' perceived merit ("Does this person deserve help?"), need ("Does this person need help?"), or factual judgments that did not require social inferences ("Does this person use both hands?"), which served as a control condition. Before each task block, a visual prompt informed participants of the upcoming condition of the task block. Moreover, a keyword was briefly presented between images of a block as a reminder (i.e., "need", "deserve", or "both hands"; Figure 1A). Each block

consisted of 32 images, and participants completed two blocks per condition. Thus, participants viewed a total of 64 images, each presented once per condition (need, merit, and control). Some images displayed multiple people; a green arrow superimposed on the photograph indicated the target of the social perception. Images were displayed for 2 seconds with a 0.5-second inter-stimulus interval. The presentation order of images was fixed across participants to maximize efficiency. Crucially, this task did not require people to make altruistic choices or engage in meaningful social interaction. Instead, it focused solely on capturing individuals' patterns of perceiving others' perceived merit, need, and factual control inferences, which we modeled using an evidence accumulation framework (computational model of social perception, see below). The task was implemented in MATLAB (MathWorks) using the Psychophysics Toolbox extensions [55-57]. The stimuli and presentation code are available at https://osf.io/4u5vs/.

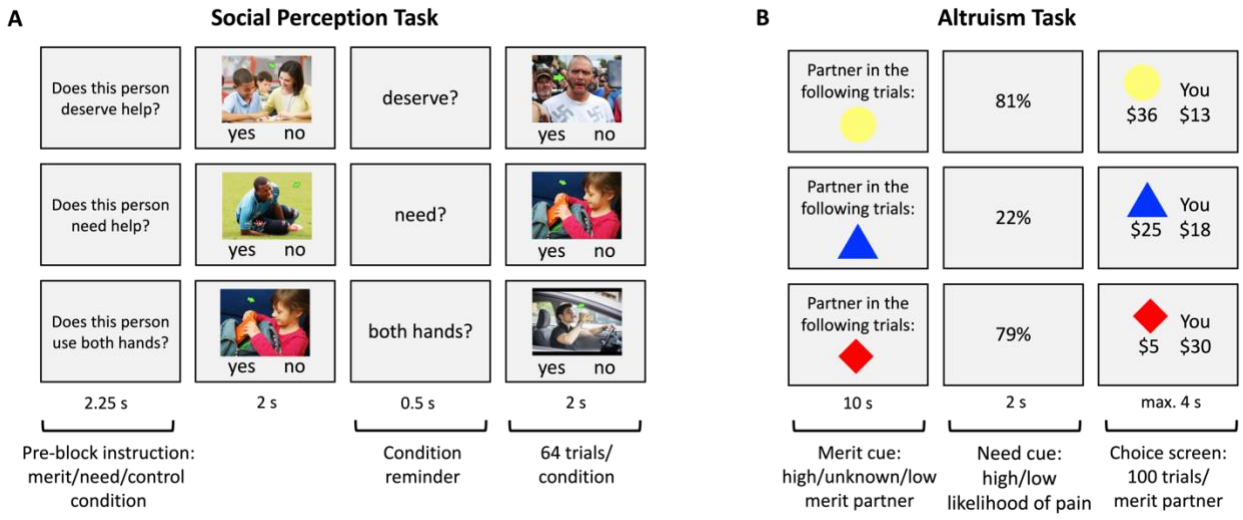

Figure 1. Experimental tasks performed on separate days. (**A)** Social Perception Task. Participants made rapid yes/no judgments regarding others' perceived deservingness (merit blocks, top row), need (need blocks, middle row), or factual inferences (control blocks, bottom row) while their brain responses were measured using fMRI. (**B**) Altruism Task. On each trial, participants accepted or rejected a monetary offer that affected the payoffs for themselves ("You") and one of three partners (choice screen; displayed offer vs. constant default of $20 for both). The three partners (identified via colored geometric shapes) differed in their associated merit (merit cue:

high/low/unknown) based on partner behavior in a separate exchange game played before the altruism task. Partner's need (need cue: high/low) was manipulated on a trial-by-trial basis, indicated by the likelihood of a painful cold pressor task (CPT) for the partner after the altruism task (high need: 80±4%; low need: 20±4%). Partners could buy out of the painful CPT using funds from one randomly selected trial at the end of the altruism task. Participants were informed that they could help their partners to avoid the painful CPT by making generous choices.

Altruism Task

Participants completed an altruism task (modified dictator game) on a different day than the social perception task (average delay of 303 days, min. = 27, max. = 663). All but one participant completed the social perception task first. On each trial, we presented participants with a proposed monetary allocation between themselves and one of three partners (e.g., $15 for themselves and $25 for their current partner; choice screen, see Figure 1B). Participants decided whether to accept or reject the proposed allocation in favor of a constant default allocation of $20 for both [45,58]. Participants indicated their choices by pressing one of four buttons ("strong no", "no", "yes", or "strong yes"). The direction of the response scale ("strong yes" to "strong no") and the presentation side of self- and other-related payoffs (left vs. right side of the screen) was counterbalanced across participants. Proposed monetary outcomes for the participant ("You") and the partner (represented by one of three colored geometric shapes) ranged from $5 to $35 (see Supplemental Figure S1). To minimize habituation and repetition effects, we randomly jittered proposal amounts by $0-$4. We informed participants that one trial would be randomly selected and implemented at the end of the experiment. In total, there were 300 trials divided across five runs (i.e., 60 trials per run, 100 trials per partner). Stimulus presentation and response collection in the altruism task were implemented using PsychoPy [59,60].

Partner's Merit (High/Low/Unknown). Participants played the altruism task with three partners that differed in their perceived merit (implemented in separate blocks; 20 consecutive

trials per partner within a functional run; counterbalanced order of the three partners across participants). The merit of the three partners in the task was manipulated before the altruism task and was based on information about partners' behavior in a sequential prisoners' dilemma task that partners played with anonymous third persons, modified from [61,62] (see Supplemental Note S1). Merit levels were manipulated such that one player was perceived as highly deserving (high merit partner), undeserving (low merit partner), or having unknown merit (no information provided before the altruism task, control condition). Partner identity was indicated in the altruism task using one of three colored geometric shapes (random combination of either a circle, diamond, or triangle, colored either red, yellow, or blue), shown at the beginning of each block of 20 trials with the same partner (merit cue, Figure 1B) and on each trial-wise offer screen indicating payoffs for the specific partner (choice screen, Figure 1B).

Partner's Need (High/Low). The altruism task also experimentally manipulated the need level of the three partners on a trial-by-trial level. Specifically, we manipulated the probability that a partner would have to complete a painful cold pressor task (CPT [63]) at the end of the experiment (outside of the scanner; hand submerged in ice water for ~2 minutes). We informed participants that each of the three partners would be given the option of using money received in the altruism task to buy out of the post-task CPT (based on a randomly selected trial at the end of the task that would be implemented according to the participants' choice on that trial). With each dollar spent, the partner could subtract 10% from their probability of having to perform the painful CPT (e.g., spending $3 would reduce an 80% chance to 50%). Thus, participants knew they could help the other player avoid the painful CPT by making generous choices.

To signal need on each trial, participants were presented with a percentage indicating the others' need (high vs. low probability of CPT) before seeing the proposed monetary allocation

(need cue, Figure 1B). Thus, in trials with high percentages, the partners had a greater need for money to avoid the painful experience. For each partner, CPT likelihood was low in half of the trials (mean CPT probability of 20%) and high in the other half (mean CPT probability of 80%; random presentation order of high vs. low need trials). As with the monetary proposal amounts, we randomly jittered need-related percentages by 0-4% to minimize habituation and repetition effects.

To ensure the saliency of the experimental need manipulation, all participants completed the painful CPT task themselves before the altruism task outside the scanner. Participants were also told that all three partners had already completed one round of the CPT task before the altruism task and were presented with their ostensible pain ratings (7 on a 7-point scale; 1 = not at all; 7 = extremely). These ratings signaled to participants that all three partners found the ice water equally and extremely painful and were highly motivated to avoid another round of the CPT. After the altruism task, participants completed a variety of computerized sanity check questions and questionnaires outside of the scanner (see Supplemental Figure S2). These sanity checks were used to verify the effectiveness of our experimental merit and need manipulations and to exclude any participants who failed to correctly remember how partners had acted in the behavioral exchange game prior to the altruism task.

Analysis

Computational Behavioral Model of Social Perception

We developed two behavioral computational models to characterize individual differences in social perceptions (social perception task) and prosocial behaviors (altruism task). Separately for each task, we modeled participants' choices and reaction times (RTs) using variants of the drift

diffusion model (DDM) [64]. This model depicts choices as the noisy accumulation of evidence until a sufficient level favoring one choice option is attained. The DDM has been used to examine processes underlying both perceptual and value-based decisions [64-67] and is being increasingly applied to studying social and affective decision-making processes [33,44,45,68-71].

To model trial-wise responses in the social perception task (Figure 1A), we assumed that when faced with the task of judging each image on a particular dimension (i.e., merit, need, or the control judgments of using both hands), a decision-maker employs the following strategy. At each moment in time, they draw noisy samples of both task-relevant and task-irrelevant evidence about *Merit*, *Need* and *Control* (both hands) from the stimuli, weighted by person- and condition-specific sensitivities S_{merit}, S_{need} and $S_{control}$. They accumulate these samples of evidence E at each timepoint t according to the following equation:

$$E(t|Condition) = Bias_{condition} + S_{merit|condition}Merit + S_{need|condition}Need + S_{control|condition}Control + \epsilon(t) \quad \text{(Eq. 2)}$$

Condition refers to the specific judgment (merit, need, control) being performed on that trial (task block). Equation 2 reflects the assumption that a person's judgment should be most sensitive to task-relevant information (e.g., cues signaling merit during the merit judgment blocks), but might also be inadvertently influenced by task-irrelevant information (e.g., cues signaling need in merit judgment blocks). For example, merit judgments should primarily reflect cues indicating merit (i.e., high sensitivity estimates of $S_{merit|merit\ condition}$) but might also reflect in part cues indicative of need (i.e., a low but non-zero $S_{need|merit\ condition}$). Thus, the model yields a total of nine sensitivity parameters estimated across the three conditions of the social perceptual task. The image-level

evidence for *Merit*, *Need* and *Control* (for the stimulus shown on that trial) was estimated using data from an independent participant sample (Supplemental Note S2). For each image, we used the mean-centered average proportion of "yes" responses that the displayed target individual deserved help, needed help, or was using both hands (based on data from the independent sample). We focus on the 'task-relevant' estimated parameters S_{merit} in the merit condition, S_{need} in the need condition, and $S_{control}$ in the control condition as indices of participants' perceptual *sensitivity* to cues suggestive of others' merit, need, or usage of both hands (but see Supplemental Note S5 for further detail on task-irrelevant sensitivity parameters). Higher parameter values suggest stronger discrimination of the normatively agreed-upon relevant cues (as captured by the independent sample). The DDM also includes three condition-wise free parameters that influence the overall drift, irrespective of the specific image ($Bias_{need}$, $Bias_{merit}$, and $Bias_{control}$; indicated as $Bias_{condition}$ in Eq. 2). These parameters allow us to capture an individual's general tendency to identify cues suggestive of others' merit and need (or control), regardless of the actual social cues present in the image. We will refer to these estimates as perceptual *bias*.

Once the momentary evidence E reaches the upper (yes) or lower (no) threshold, evidence accumulation terminates, and the corresponding choice is implemented. The difference between thresholds is estimated by a set of three parameters ($a_{\Delta need}$, $a_{\Delta merit}$, and $a_{\Delta control}$) that represent within-individual stability and difference across tasks (see Supplemental Note S3 for details). The DDM also includes three condition-wise non-decision time (*ndt*) parameters for capturing the time taken to initially encode stimuli and to implement the motor response, estimated similarly to incorporate within-individual stability and change across task conditions. Finally, the model included three condition-wise starting bias (z) parameters, which represent another potential mechanism by which biases at the onset of evidence accumulation can impact the decision process. Note that, although

we estimated these starting biases to improve model fit, our focus in all primary analyses reported in this paper is on the evidence-related *perceptual* bias for merit and need (represented by $Bias_{condition}$ in merit and need blocks, respectively), not these motor-related starting biases [72]. A detailed description of the model fitting procedure is provided in the Supplemental Note S3 and Supplemental Table S1.

Behavioral Generosity in the Altruism Task

Choices in the altruism task (Figure 1B) involved a trade-off between monetary outcomes for the self and one of three partners. Monetary proposals were selected so that a choice could benefit the partner or the participant, compared to the constant default ($20 for both players). Following previous implementations [45,58,73], we classified a choice as generous if the participant *accepted* a proposal that benefited the partner at the expense of oneself ($Self < $Other) or *rejected* a proposal that profited themselves at the partner's expense ($Self > $Other). The overall and condition-specific fraction of generous choices measured participants' generosity. Differences in generosity across conditions provided model-free estimates of the impact of social cues about others' merit and need on social behaviors. To assess whether merit and need altered generosity, we computed a mixed-effects logistic regression using the *glmer* function in R. Trial-wise information about generous choice (no = 0, yes = 1) served as the dependent variable. The model included the following fixed effects: a trial-wise indicator for the level of the partners' need (reference: low need), an indicator for the partners' merit level (reference: merit low), and their interaction. Participant id was specified as a random effect, as follows:

$$\text{Generous Choice (0/1)} \sim 1 + \text{Need} + \text{Merit} + \text{Need} \times \text{Merit} + (1|\text{Participant}) \qquad \text{(Eq. 3)}$$

Computational Behavioral Model of Altruistic Choice

To model choices and reaction time behavior in the 3 (merit: high, unknown, low) x 2 (need: high, low) design of the altruism task, we implemented a second behavioral computational model. Following previous applications [45,58], for each of the six experimental *conditions*, value-related evidence *V* at time *t* was estimated as follows:

$$V(t|condition) = w_{0|condition} + w_{self|condition}Self + w_{other|condition}Other + w_{fairness|condition}Fairness + \epsilon(t) \quad \text{(Eq. 4)}$$

Here, *Self* and *Other* refer to the monetary outcomes of the proposed allocation of a trial (minus the default outcomes of $20 for both; rescaled by dividing by 10). The *Fairness* of the proposed allocation was calculated as $-1 \times |Other - Self|$. The free parameters for each attribute weight (w_{self}, w_{other}, $w_{fairness}$) indicate variance in the degree to which evidence about outcomes to the self, other, or the fairness of the monetary offer guided choices in a particular condition of the altruism task. The value constant w_0 represents the extent to which a participant tends to prefer the proposal over the default, regardless of trial-specific values. Matching the computational model of social perception (see above), the model of altruistic choice also estimated free parameters capturing the decision threshold (*a*), non-decision time (*ndt*), and response-related starting bias (*z*) parameter (see Supplemental Note S3 and Supplemental Table S1 for details). The parameters *z*, *ndt*, and *a* were fixed across the merit x need conditions.

To capture variance in weight parameters *w* across the three levels of merit (high, low, unknown) and two levels of need (high, low), we estimated both individual-specific stability and change parameters for each of the four weights *w* (i.e., self, other, fairness, and the value constant).

A more complicated model that allowed for need × merit interactions in driving weights did not improve model fit. We thus focus on this simpler model that includes a baseline parameter for each weight, and three change parameters for the effects of increasing or decreasing merit and increasing need (see Supplemental Note S3 for details).

Correlating Social Perception and Social Action

Next, we tested for a relationship between social perception and social action. To this end, we correlated individual-specific bias and sensitivity parameters obtained from the social perception task with value-based attribute weight parameters in the altruism task. Because we found that some estimates of our computational model diverged from a normal distribution, we used non-parametric statistical tests to examine these relationships. Post-hoc tests were corrected for multiple comparisons using the fdr_bh function implemented in MatlabR2022a. All results are reported for 2-tailed statistical tests, unless reported otherwise. We removed outliers from all variables in our analyses based on values that exceeded three standard deviations from the mean.

fMRI Data

Acquisition. All neuroimaging data were acquired at the Caltech Brain Imaging Center using a Siemens Trio 3.0 Tesla scanner outfitted with a 32-channel phased-array head-coil. Functional image acquisition for the social perception task and altruism task occurred on different days in separate sessions (average delay of 303 days). For both fMRI tasks, we acquired gradient echo T2*-weighted echo-planar images (EPIs; 60 slices, voxel resolution 2.5 x 2.5 x 2.5 mm^3, TR = 700 ms, TE = 30 ms, flip angle = 53°, FOV = 200 mm, interleaved acquisition order, multi-band

acceleration factor = 6). We collected 903 volumes for the social perception task and 925 volumes for the altruism task. For all participants, we also acquired a high-resolution anatomical T1-weighted image using a MEMP-RAGE sequence (208 slices, 0.9 x 0.9 x 0.9 mm^3, matrix size 256 x 256, TR = 2.55 s, TI = 1.15 s, TE = 1.6, 3.5, 5.3, 7.1 ms with RMS echo combination, RAGE flip angle = 8˚). Distortion correction data for the fMRI EPI acquisitions employed a pair of phase-encoding polarity reversed T2w SE-EPI images with identical geometry and EPI echo train timing to the T2*w EPI images (TR 4.8 s, TE 50 ms, flip angle 90°).

Preprocessing. Preprocessing of all functional and structural brain data was performed using a standard implementation of fMRIPrep 20.2.3[75] , which is based on Nipype 1.6.1 [76,77]. A detailed description of the standardized analysis steps is provided in Supplemental Note S4.

GLM of brain responses in the social perception task. This analysis aimed to identify brain regions recruited during different judgments in the social perception task. For each participant, we estimated a general linear model (GLM) using a canonical hemodynamic response function and a 128s high-pass cut-off filter to eliminate low-frequency drifts in the data. The GLM estimated three regressors of interest corresponding to condition-wise brain responses (implemented in different blocks) during merit inferences ("deserves help?"), need inferences ("needs help?"), and factual control inferences that did not require social inferences ("used both hands?"; Figure 1A). The three condition-wise regressors were defined by the onset time of the first target image and the offset of the final image of each block (two blocks per condition). The GLM also included eight regressors of no interest: six motion regressors, the framewise displacement (estimated during preprocessing of neuroimaging data using fMRIPrep, https://fmriprep.org/), and a session constant. The GLMs were implemented in MATLAB (R2018b) using the SPM12 toolbox (http://www.fil.ion.ucl.ac.uk/spm).

For each participant, we created several contrasts of interest: First, we identified brain regions activated during merit perception (contrast [merit – control]) and need perception (contrast [need – control]). Second, we estimated brain regions that are significantly more (or less) activated for need compared to merit perception (contrasts [merit – need]; [need – merit]). The respective contrast images were then used in four group-level analyses using simple t-tests as implemented in SPM12. Whole-brain maps were thresholded at $p < 0.001$ (voxel level) and family-wise error (FWE) corrected at a cluster level of $p < 0.05$. Finally, we tested for brain regions that were reliably activated during both need and merit perceptions, by estimating a formal conjunction of both inference-specific brain maps ([need - control] ∩ [merit - control]) at the group level. Both brain maps used in this conjunction analysis were thresholded at the cluster-forming voxel level of $p < 0.001$, FWE cluster-level corrected at $p < 0.05$.

We also estimated a supplemental GLM for fMRI data from the altruism task. Due to COVID-19-related interruptions, only 25 participants from the sample that performed the social perception task also completed the fMRI altruism task. Given the limited sample size and noise level of fMRI data, we decided to focus solely on the behavior in the altruism task to address our research objectives. For completeness, we report the complementary GLM for the fMRI altruism data and its results in Supplemental Note S7.

Results

This section is structured as follows. First, we report the results of our novel fMRI social perception task to describe how people perceive others' merit and need. Behaviorally, we used parameter estimates from our computational model of social perception to characterize individuals' perceptual biases and sensitivities driving variance in social perceptions across people. Neurally,

we localized brain regions recruited during merit and need inferences and asked whether neural computations underlying merit and need perceptions are supported by distinct neural circuits or a general-purpose network for social inference processing. Second, we used data from a separate altruism task (collected on average 303 days later) to describe the effects of an interaction partner's perceived merit and need on meaningful social behavior. Third, we explored if individuals' perceptual biases and sensitivities (as captured in the social perception task) predicted variance in actual social behavior (observed in the altruism task). We examined this core question about the social perception-action link across time and contexts at both the behavioral and neural levels.

Social Perception Task: Behavior and Neural Underpinnings of Need and Merit Inferences

Despite viewing the same stimuli, participants differed dramatically in their judgments of others' merit and need. The percentage of trials perceived by participants as depicting someone as deserving (merit blocks) ranged from 33% to 81%. Similarly, the range for perceiving someone in need (need blocks) was 28% to 73% (Supplemental Figure S3). What mechanisms drive these profound differences in social perceptions?

Computational behavioral model of social perception.

To address this question, we examined two distinct computational mechanisms: first, people may differ in their general tendency to perceive others as deserving or in need (bias hypothesis). Second, people may differ in their perceptual sensitivity to merit- or need-related evidence in their choice environment and/or their ability to use this information to guide their judgments (sensitivity hypothesis). Importantly, these mechanisms are not mutually exclusive, but may vary in their relative impact on social behaviors across people and contexts. We tested the contribution of these

two potential mechanisms using our behavioral computational model of social perception (see Methods).

We first verified that the model adequately accounted for the differences in participants' choices and response times in all three task conditions (merit, need, and control blocks; for an illustration of the model fit, see Supplemental Figure S4).

Next, we examined the estimated parameters of our hierarchical model at the individual participant level. Positive average sensitivity parameters for merit (S_{merit} in merit blocks = 2.84 ± 1.02, mean ± std, significantly different from zero, $p < 0.001$, FDR corrected), need (S_{need} in need blocks = 3.28 ± 0.75, $p < 0.001$, FDR corrected), and control ($S_{control}$ in control blocks = 4.40 ± 0.69, $p < 0.001$, FDR corrected) verified that participants accurately distinguished between targets who deserve help, need help, and use both hands, and used this perceptual evidence to guide their judgments (Figure 2A). For the estimated perceptual bias parameters, people tended to perceive others as deserving (indicated by an average positive merit bias, $Bias_{merit}$ = 0.33 ± 0.47, $p < 0.001$, FDR corrected) and not in need of help (indicated by an average negative need bias, $Bias_{need}$ = -0.17 ± 0.39, $p = 0.009$, FDR corrected). In the control condition, the estimated bias showed a small positive but non-significant trend ($Bias_{control}$ = 0.09 ± 0.36, $p = 0.099$, FDR corrected; Figure 2B). It is worth noting that the *Bias* parameters are strongly associated with (though not the sole determinant of) the mean response rate. For a description of the estimated values for the hyper-mean parameters in our model and further sanity checks, see Supplemental Table S2 and Supplemental Note S5.

Notably, individuals' bias and sensitivity estimates were uncorrelated for merit (Spearman's $r = 0.10$, $p = 0.73$, FDR corrected; S_{merit} and $Bias_{merit}$) and need judgments

(Spearman's r = -0.10, p = 0.73, FDR corrected; S_{need} and $Bias_{need}$), suggesting two distinct mechanisms driving variance in social perceptions across people. Likewise, we found no link between need and merit sensitivity (Spearman's r = -0.22, p = 0.60, FDR corrected; S_{merit} and S_{need}), indicating that people might be sensitive to cues signaling merit but not need, and vice versa. However, estimates of the need bias and merit bias were correlated (Spearman's r = 0.73, $p < 0.001$, FDR corrected; $Bias_{merit}$ and $Bias_{need}$), suggesting that people who tend to perceive others as needy might also tend to perceive others as deserving. See Supplemental Figure S5 for full details on intercorrelations of estimates of the computational model. These findings highlight the benefits of formal computational models, which can capture and quantify distinct processes that can be hard to distinguish based on observed behaviors alone. It also raises the interesting question of whether the specificity of need and merit observed at the behavioral level gives rise to inference specificity at the neural level.

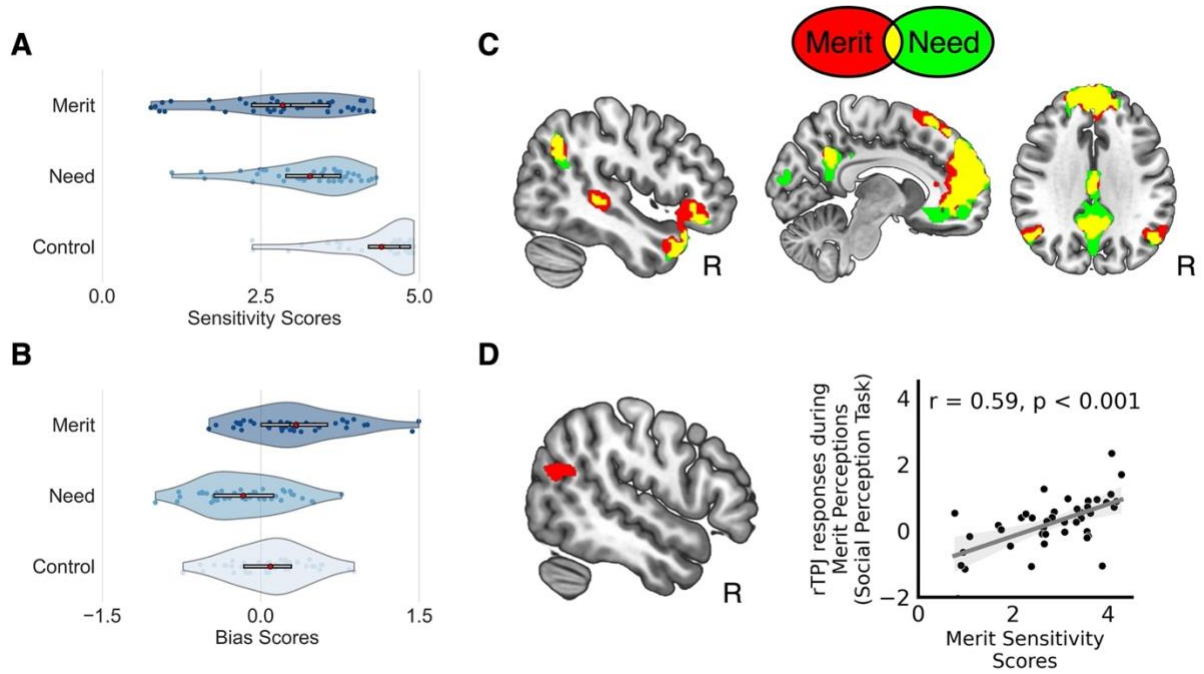

Figure 2. (**A**) Distribution of participant-specific model-based estimates of perceptual sensitivity and (**B**) perceptual bias for each inference condition in the social perception task (merit, need, and

control). (**C**) Need and merit inferences activate the mentalizing network to a similar extent (red illustrates brain regions activated for [merit - control], green illustrates brain regions identified for [need - control], and yellow indicates overlap; both contrast maps thresholded at $p < 0.001$ at the voxel level, FWE corrected at the cluster level at $p < 0.05$; R = right hemisphere). (**D**) Activity in the right temporoparietal junction (TPJ) during merit perceptions (merit - control) reflects individual differences in merit sensitivity scores estimated in the computational model of social perception.

Neural underpinnings of merit and need perceptions.

To address this question, we examined the neural substrates of need and merit inferences obtained in the social perception task. The task is a modified version of an established fMRI why/how task [53,54] that reliably identifies the mentalizing network [78-82]. Not surprisingly, we found that need and merit inferences also recruited the mentalizing network. The medial prefrontal cortex (MPFC), superior temporal sulcus (STS), temporal pole (TP), temporoparietal junction (TPJ), and posterior cingulate cortex (PCC) were reliably activated during both merit and need inferences, among other regions. Figure 2C illustrates the overlap of brain areas activated during merit and need judgments (Supplemental Table S3 provides the condition-specific results). Table 1 provides the results of the formal conjunction of brain maps identified for [need - control] and [merit - control] inferences, each thresholded at $p < 0.001$ at the voxel level, FWE corrected at the cluster level at $p < 0.05$ (see Supplemental Figure S6 for a visualization of the conjunction analysis). No brain region was significantly more activated during merit than need inferences [merit – need] at our omnibus threshold. However, for the reverse contrast [need – merit], need inferences showed significantly greater activation in the cuneus (MNI: [-4, -98, 18], 8741 voxels, $t = 7.64$), intraparietal sulcus (MNI: [28, -62, 56], 256 voxels, $t = 4.58$) and sensorimotor cortex (MNI: [4, -38, 64], 294 voxels, $t = 4.39$) ($p < 0.001$ at the voxel level, FWE corrected at the cluster level at $p < 0.05$), suggesting an enhanced activation of the (extended) mirror system [83,84]. Future research may reveal additional

distinctions between merit and need appraisals in trial-wise (compared to our block-wise) fMRI designs.

We also employed supplemental multivariate decoding analyses (searchlight analysis [85-87]), as commonly used in social perception and neuroscience research [7,58,82,88,89], corroborating our univariate findings (see Supplemental Note S6, Supplemental Table S10). Taken together, our results demonstrate that both need and merit inferences reliably recruited the well-established mentalizing network, and to a comparable extent. Answering our first core question: These findings are consistent with the notion that both appraisals are supported by general-purpose rather than domain-specific social cognitive mechanisms.

Table 1. Conjunction of neural activations during social need and merit inferences in the social perception task.

Brain Region	**Side**	**T**	**k**	**MNI**		
				x	**y**	**z**
[Need - Control] ∩ [Merit - Control]						
Dorsomedial prefrontal cortex	L	7.78	6250	-14	38	48
Superior temporal sulcus	L	6.61	1154	-64	-26	-14
Superior temporal sulcus	R	5.42	146	70	-32	0
Temporal pole	R	6.16	523	40	22	-32
Posterior cingulate cortex	L	5.97	671	-2	-54	32
Midcingulate cortex	L	5.17	119	-2	-10	34
Temporoparietal Junction (TPJ)	R	4.56	167	54	-66	34

Temporoparietal Junction (TPJ)/Angular Gyrus	L	4.58	141	-50	-62	32
Cerebellum	L	5.33	336	-26	-82	-42

Note. Results are reported at a statistical threshold of $p < 0.001$ at the voxel level, FWE corrected at cluster level at $p < 0.05$. Only peak activations of each cluster are shown. L = Left hemisphere, R = Right hemisphere, MNI = Montreal Neurological Institute, k = cluster size in voxels.

Model estimates of merit sensitivity modulate the neural underpinnings of merit perceptions.

So far, we showed that a general-purpose network involved in social inference processes is recruited during both merit and need inferences. Next, we ran a series of whole-brain analyses to identify brain areas for which inference-evoked brain responses covary with individual differences in estimates of the computational model of social perception. Given the evidence above that individuals activated the social cognitive brain network to make these judgments, we hypothesized that stronger activation in one or more hubs in this network should correlate with greater perceptual sensitivity and/or bias in these social inferences. Consistent with this prediction, we found that brain responses during merit inferences (merit - control) systematically covaried with participants' merit sensitivity scores in the right temporoparietal junction (rTPJ, peak at [MNI 56, -64, 22], $t = 4.24$, $k = 228$ voxels, $p < 0.001$ at the voxel level, FWE corrected at the cluster level at $p < 0.05$). In other words, participants with larger merit sensitivity scores (captured in the computational behavioral model) exhibited larger rTPJ responses when appraising someone's merit (Figure 2D). This effect was specific for brain responses obtained during merit inferences; an analogous whole-brain analysis using neural activation obtained during need inferences (need - control) to predict merit sensitivity did not yield significant effects. We also did not observe any brain regions where inference-evoked brain responses were associated with variance in participants' need sensitivity or bias scores for either merit or need at the whole brain level ($p < 0.001$ at the voxel level, FWE

corrected at the cluster level at $p < 0.05$). Thus, despite merit and need inferences both recruiting the general-purpose mentalizing network to an equivalent extent on average, we found some evidence of neuroanatomical specificity for merit inferences when considering estimates of the computational behavioral model (Figure 2D). Notably, this functional link between the rTPJ and merit sensitivity was robust when we repeated the whole-brain analysis for the reduced sample of $n = 25$ participants with overlapping altruism task data ($p < 0.001$ at the voxel level, FWE corrected at the cluster level at $p < 0.05$; for details, see Supplemental Table S4; for illustration, see Supplemental Figure S7).

Altruism Task Behavior

The results reported above relate to *perceptions* of need and merit; however, they say nothing about how such perceptions might influence decisions to help. In a separate altruism task, we examined how independently manipulating a social target's merit (high, unknown, low) and need (high, low) alter prosocial behavior. This altruism task was unrelated to the social perception task and was completed on average 303 days later. It allowed us to characterize how people *act* on perceptions of merit and need when deciding whether to give aid to another person.

Partner's need and merit alter generosity in the altruism task.

Did the experimental manipulations of another's need and merit affect people's generosity during altruistic choice? We addressed this question by fitting a mixed-effects logistic regression model to the observed generous and selfish choices (coded as 1/0) in the altruism task (Eq. 3; see Figure 3A for the average proportion of generous choices across conditions, and Supplemental Table S5 for summary statistics). The model's total explanatory power was substantial ($R^2 = 0.33$) and significantly better than a null model that assumed no effect of need or merit (χ^2 (5, N = 28) =

239.68, $p < 0.001$). Partners' merit (χ^2 (2, N = 28) = 45.68, $p < 0.001$) and need (χ^2 (1, N = 28) = 16.79, $p < 0.001$) both influenced generosity. On average, subjects were more generous to a partner in high (vs. low) need (beta = 0.40, 95% CI [0.21, 0.58], $p < 0.001$). They were also more generous to a high (vs. low) merit partner (beta = 0.60, 95% CI [0.42, 0.79], $p < 0.001$) and an unknown (vs. low) merit partner (beta = 0.52, 95% CI [0.33, 0.71], $p < 0.001$). No interactions between the level of merit and need were significant (χ^2 (2, N = 28) = 2.77, p = 0.25; need [high] × merit [unknown]: beta = 0.19, 95% CI [-0.06, 0.45], p = 0.142; need [high] × merit [high]: beta = 0.19, 95% CI [-0.07, 0.45], p = 0.147). Likewise, results of a formal model comparison revealed that adding the interaction of need and merit did not improve model fit significantly over a model that only considered the main effects (χ^2 (2, N = 28) = 2.75, p = 0.252; Supplemental Table S6). These findings suggest that need and merit inferences had fully independent effects on social choice. Consistent with this notion, we also found that merit-induced changes in generosity (high-low merit partner) and need-induced changes in generosity (high-low need) were uncorrelated (Spearman r = 0.17, p = 0.377). In other words, people who changed their generous behavior as a function of another's merit were not necessarily the same as those who changed their behavior in response to another person's need. Consequently, further analyses focused on the main effects of need and merit on altruistic choice. Overall, the experimental manipulations of another's need and merit affected people's generosity during altruistic choice.

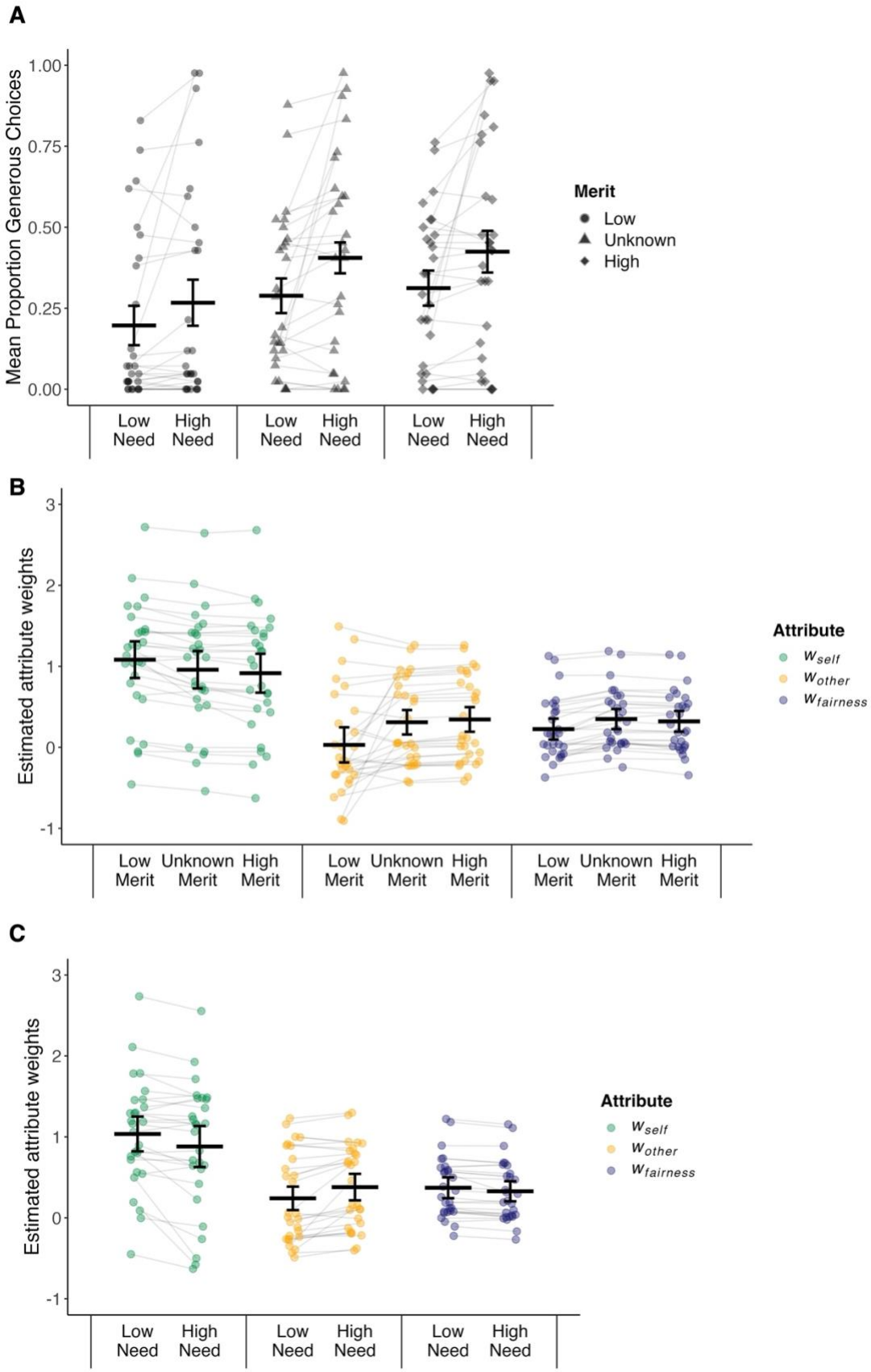
A
Mean Proportion Generous Choices
1.00
0.75
0.50
0.25
0.00
Low Need
High Need
Low Need
High Need
Low Need
High Need
Merit
Low
Unknown
High
B
Estimated attribute weights
3
2
1
0
-1
Low Merit
Unknown Merit
High Merit
Low Merit
Unknown Merit
High Merit
Low Merit
Unknown Merit
High Merit
Attribute
w_{self}
w_{other}
$w_{fairness}$
C
Estimated attribute weights
3
2
1
0
-1
Low Need
High Need
Low Need
High Need
Low Need
High Need
Attribute
w_{self}
w_{other}
$w_{fairness}$

Figure 3. (**A**) Partner's merit and need altered generosity in the altruism task. High (vs. low) need contexts elicited more generosity ($p < 0.001$). Compared to a low merit partner (circle), generosity was enhanced towards a high merit (diamond, $p < 0.001$) and an unknown merit partner (triangle, $p < 0.001$). (**B**) Condition-specific attribute weights (w_{self}, w_{other}, $w_{fairness}$) for low, unknown, and high merit partner contexts. (**C**) Condition-specific attribute weights for low and high need contexts. All p's ≤ 0.01, FDR corrected, for the six pair-wise comparisons of changes in attribute weights (high vs. low merit/need). Dots represent participant-specific estimates from the computational model of altruistic choice (see Supplemental Note S3 for detail); black lines illustrate the estimated means and 95% confidence intervals.

Computational behavioral model of altruistic choice: partners' merit and need alter social attribute weights.

Several mechanisms might drive the changes in altruistic choices we observed (see above): others' merit or need might decrease self-interest, increase other-regard, increase fairness considerations, or some combination of these. To identify the relative contribution of the processes that drive generosity, we turned toward an established behavioral computational model of altruistic choice (Eq. 4; see Methods). We first verified that our computational model of altruistic choice fit the data well by showing that choices and reaction times (RTs) were captured with high accuracy (for visualization of model fit, see Supplemental Figure S4).

Next, we examined the relative importance of the social attributes (drift weights w), irrespective of the partner's need or merit. On average (across all conditions), self-related outcomes strongly guided choices in the altruism task (indicated by the average overall positive weight on payoffs for self, $w_{self} = 0.97 \pm 0.75$, mean ± std), more so than concern for others' outcomes (overall $w_{other} = 0.22 \pm 0.49$, $p < 0.001$; FDR corrected) or fairness (overall $w_{fairness} = 0.35 \pm 0.36$, $p = 0.028$; FDR corrected). These results match the findings of prior studies [58,73]. Variance in these *overall* attribute weights across people served as an indicator of more general,

context-independent differences in (pro)social decision-making (explaining who will be more generous, on average, and why).

Finally, we tested if contextual cues about the partner's merit or need altered the degree to which benefits for oneself (w_{self}), others (w_{other}), or fairness concerns ($w_{fairness}$) guided social choices (separate Wilcoxon signed rank tests for each attribute and experimental manipulation; FDR corrected for multiple comparisons; see Supplemental Table S7 for attribute-specific estimates; see Figure 3B and C for illustration). Any changes in model estimates would indicate *context-dependent* effects (of varying levels of merit and need) of attributes' input on altruistic behaviors. When interacting with a high (vs. low) merit partner, benefits for oneself (w_{self}) guided choices *less* ($p < 0.001$, FDR corrected), whereas considerations of others' benefits (w_{other}, $p < 0.01$, FDR corrected) and fairness concerns (overall $w_{fairness}$, $p < 0.001$, FDR corrected) guided choices *more*. Mirroring this pattern of results, in high (vs. low) need settings, weights on self-related outcomes *decreased* ($p < 0.01$, FDR corrected), whereas weights on others' benefits *increased* ($p < 0.001$, FDR corrected). Unlike the merit-induced effects, a partner's high (vs. low) need *reduced* weights on fairness concerns on choices ($p < 0.001$, FDR corrected). Thus, if others faced great need, participants were more willing to ignore their fairness preferences. Together, these results suggest that merit- and need-evoked changes in generosity are driven by systematic changes in the social decision process - namely changes in attribute weights - as captured in our behavioral computational model of altruistic choice. Thus, we addressed another key question by showing that partners' merit and need levels changed how specific choice-relevant considerations (self-regard, other-regard, fairness) guided people's decisions to act prosocially.

Individual differences in effects of others' need and merit and altruistic decision-making.

Notably, people differed substantially in their overall generosity (Supplemental Table S5) and the degree to which generosity varied as a function of their partner's need and merit. Our model assumes this variation was driven by changes in social attribute weights w estimated in the behavioral computational model (see above). To quantify these individual differences and examine their relations, we calculated change scores of model-based estimates for each participant. For example, to capture the change in an individual's other-regard as a function of their partner's need, we computed the following participant-specific difference score [w_{other} in high need – w_{other} in low need]). Likewise, to assess the change in other-regard in response to their partner's merit, we estimated the change in [w_{other} high merit - w_{other} low merit]. We did this separately for each attribute weight estimated in the computational model of altruistic choices (w_{self}, w_{other}, $w_{fairness}$). These change scores reflect the impact of partners' merit and need on individuals' altruistic decision process. As a sanity check, we confirmed that changes in attribute weights reflect participant-specific changes in generosity (Supplemental Table S8). In other words, larger (smaller) changes in observed social behaviors can be explained by larger (smaller) changes in estimated attribute weights. These findings provide insight into the precise mechanism by which others' merit and need effect (pro)social behavior, namely by altering the weighting of certain choice attributes in the decision process. Below, we link these change scores in social behaviors with estimates of social perception (sensitivity, bias).

Variance in perceptual sensitivity and bias (social perception task) predict variance in prosocial behavior across people and contexts (altruism task).

The previous section contained three main takeaways regarding the separate altruism task. First, partners' merit and need independently impacted social behaviors (generosity). Second, we identified the mechanism of these context-dependent changes in social behaviors: merit and need altered the relative importance of self- and other-regard, and fairness preferences (captured in attribute weights *w* estimated in our computational model of altruistic choice). Third, we demonstrated that people differ in the degree to which they change their behaviors in response to social cues on others' merit or need. This raises an important question: what factors determine the impact of others' need or merit on behavior?

We propose that stable individual differences in social perceptions - as captured in our computational model of social perception and their neural underpinnings - can provide insights into this question. Behaviorally, the computational model decomposes individual differences in social perceptions into bias and sensitivity terms. These model estimates from the social perception task correspond to two mechanisms that drive variation in people's social perceptions of others' merit and need. We propose that these stable perceptual mechanisms can, in turn, impact social decision-making. Here, we bring together data from both tasks: the social perception task (which did not require meaningful social behavior towards others) and the altruism task (which included costly social actions). We had two specific hypotheses in mind designing the study: we speculated that merit bias (or need bias) in the social perception task should be related to *average* weights on others' outcomes in the altruism task. In contrast, estimates of an individual's sensitivity to merit (or need) should predict the extent to which a person *alters* their weight on others' outcomes as a

function of their partner's merit (or need). By combining data from both tasks, we explore a fundamental question: do people's sensitivities and biases during social perceptions translate into subsequent social action? Notably, we see the results below as evidence of *stable* individual differences, since social action (i.e., generosity) was measured on average almost 10 months (~303 days) after the social perception task.

Perceptual merit sensitivity predicts merit-related contextual changes in altruistic choice.

Estimates from our computational behavioral model map onto specific hypotheses about the relationship between behaviors observed in both tasks. Perceptual sensitivity estimates reflect individuals' tendency to sample and integrate evidence about merit (need) during the perceptual decision process. Higher sensitivity estimates will yield higher discriminability between others as a function of their merit (need). Assuming stable individual differences in sensitivity and the impact of categorizing individuals as meritorious or needy on altruistic decision-making, we hypothesized that individuals with higher perceptual merit (need) sensitivity in the social perception task would exhibit a greater change in social behavior depending on the partner's merit (need) during the altruism task. To test this hypothesis, we correlated individuals' merit sensitivity scores (social perception task) with merit-induced changes in social behavior (i.e., participant-specific changes in generosity towards [high - low merit] partners in the altruism task). We found that variance in merit sensitivity reflected merit-induced changes in generosity (Spearman's $r = 0.51$, $p = 0.010$; 4A). In other words, individuals generally sensitive to merit information during social perceptions were also more susceptible to merit information during costly social behaviors. Follow-up analyses using estimates of a computational model of the altruism task revealed that

this perception-action link was driven by merit-induced changes in other-regard and self-regard during altruistic choice (change in w_{other} for [high - low merit] partners: Spearman's r = 0.49, p = 0.045, FDR corrected; change in w_{self} for [high - low merit] partners: Spearman's r = -0.44, p = 0.047, FDR corrected; no significant link with merit-related changes in fairness weight, p = 0.784, FDR corrected) (4B and C). In other words, individuals with higher merit sensitivity showed larger discrimination in the value placed on others' wellbeing (and self-interest) when interacting with supposedly deserving and undeserving people. Next, we repeated this set of analyses for sensitivity scores estimated in the need condition of the social perception task. Here, variance in participants' need sensitivity did not reflect need-induced changes in generosity (p = 0.745) or need-induced changes in altruistic choice attributes (all p's > 0.484). However, this absence of effects for need ought to be interpreted with caution, given the comparatively small sample size. Results were qualitatively similar when statistically controlling for the delay between both tasks (partial correlations).

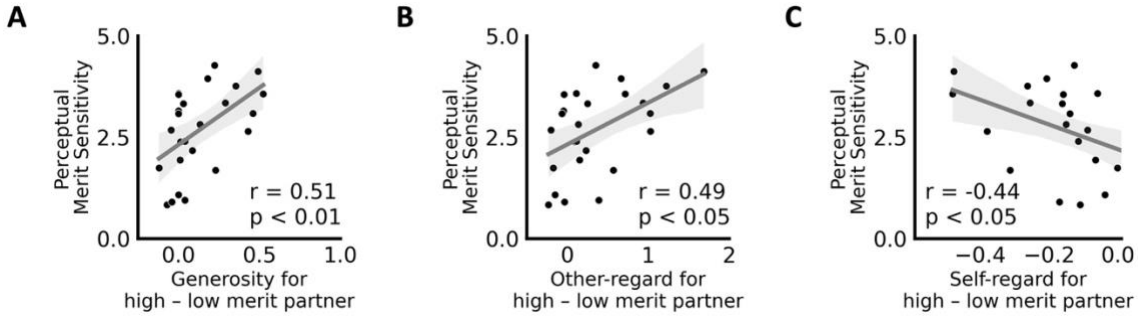

Figure 4. (**A**) Variance in individuals' general sensitivity to others' merit (social perception task) is positively linked with merit-related changes in generosity, (**B**) other-regard (w_{other}), and (**C**) self-regard (w_{self}) during altruistic choice for high vs. low merit partners. Higher values on the x-axes indicate increased generosity/other-regard and decreased self-interest when interacting with a high vs. low merit partner in the altruism task.

Perceptual bias estimates predict individuals' overall other-regard, self-regard, and fairness considerations in altruistic choice.

We also hypothesized that people's stable perceptual biases in the social perception task might translate into *context-independent* differences in social action (generosity) across people. Put differently, we assumed that individuals who have a *general* tendency to perceive others as deserving/in need (irrespective of social cues present in the environment) should be more willing to help others irrespective of contextual variation in others' merit/need. We tested this notion by correlating merit and need *bias* parameters with individuals' overall generosity in the altruism task and the overall weight of choice-relevant attributes (i.e., outcomes for self, w_{self}; outcome for other, w_{other}; and fairness, $w_{fairness}$). Contrary to our hypothesis, individuals' merit bias scores were not correlated with overall generosity (p = 0.282). However, we did find that people who tend to perceive others as deserving tended to be more other-oriented overall during altruistic choices: variance in merit bias scores in the social perception task positively correlated with the weights on others' benefits ($overall\ w_{other}$, Spearman's r = 0.50, p = 0.035, FDR corrected; marginal positive link with overall fairness concerns in the altruism task, $overall\ w_{fairness}$, Spearman's r = 0.43, p = 0.053, FDR corrected; no significant link was observed with overall self-regard, p = 0.153, FDR corrected). This finding suggests that stable differences in people's tendency to perceive others as deserving predicts people's overall other-regard across different social choice contexts (on average) ten months later. In contrast, variance in individuals' need bias (social perception task) in our sample did not correlate with overall generosity (p = 0.150) or overall attribute weights (all p's > 0.076, FDR corrected). As mentioned above, we cannot rule out the possibility that null findings may be due to the comparatively small sample size and should be interpreted cautiously

(also see discussion). Results were qualitatively similar when statistically controlling for the delay between both tasks (partial correlations).

Neural markers of merit sensitivity predict merit-related behavioral changes during altruistic choice.

So far, within the social perception task, we found that merit-evoked neural activation in the rTPJ reflects variance in people's merit sensitivity (Figure 2D). Moreover, when combining data from both tasks, we showed that individuals' merit sensitivity (social perception task) predicts merit-related changes in other-regard and self-regard that guide context-dependent changes in social behavior (altruism task) (4B and C). Considering these findings, a post-hoc test examined whether activity in the rTPJ – obtained during merit inferences (social perception task) – also predicts merit-related changes in other-regard (self-regard) in the altruism task (beyond merit sensitivity). We tested this question by using the following equation:

$$\Delta w_{other\ j} \sim 1 + S_{merit\ j} + \text{rTPJ}_{j} \qquad \text{(Eq. 5)}$$

Here, $\Delta w_{other\ j}$ represents an individual *j's* merit-related change in other-regard during altruistic choice (w_{other} for high merit - low merit partners). We used two predictor variables: participants' behavioral merit-sensitivity scores ($S_{merit\ j}$; estimated in the computational model of social perception) and participant-specific neural responses in the rTPJ obtained during merit inferences (merit - control) in the social perception task (all voxels of the rTPJ cluster, averaged across all voxels in the cluster, Supplemental Table S3). This allowed us to assess the additional predictive power of neural responses in the rTPJ, after controlling for behavioral merit sensitivity. The model's total explanatory power was substantial ($R^2 = 0.40$) and significantly better than a null model with just an intercept (χ^2 (2, N = 25) = 2.46, $p < 0.001$). We found that both model-based

estimates of individuals' merit sensitivity and rTPJ responses reliably, independently, and to an equivalent degree predict changes in other-regard during social choice approximately 10 months later (S_{merit}: beta = 0.20, SE = 0.08, 95% CI [0.04, 0.36], t(22) = 2.50, p = 0.021; rTPJ: beta = 0.29, SE = 0.13, 95% CI [0.05, 0.54], t(22) = 2.34, p = 0.029). These findings indicate that neural correlates of merit inferences – namely activity in the rTPJ – predict context-dependent changes in other-regard during social action across time and contexts, above and beyond predictive information related to perceptual merit sensitivity.

For completeness, we also estimated a modified version of Equation 5 in which we changed the dependent variables to merit-evoked changes in self-regard in the altruism task (replacing $\Delta w_{other\ j}$ with $\Delta w_{self\ j}$, defined as participant-specific change scores in self-regard when interacting with a high versus low merit partner). The model yielded an R^2 of 0.37 and the rTPJ remained a significant predictor (beta = -0.09, 95% CI [-0.17, -0.02], t(21) = -2.47, p = 0.022). However, the behavioral merit sensitivity was reduced to only marginal significance (S_{merit}: beta = -0.05, 95% CI [-0.10, 0.00], t(21) =-1.92, p = 0.069). Thus, rTPJ responses during merit inferences were tied to estimates of contextual changes in both other- and self-regard to an equivalent or possibly even greater degree than behaviour alone.

Discussion

Humans do not help others indiscriminately: they are more inclined to help people perceived as needy or deserving [4,6-9,30,31]. Using a novel fMRI social perception task, we disentangled two distinct computational mechanisms that shape variance in these social judgments: a general *bias* to perceive others as more or less deserving (in need) and a degree of discrimination or *sensitivity* to social cues signaling others' merit (need). Estimates of these two computations were

uncorrelated, suggesting they represent distinct - but not mutually exclusive - processes driving individual differences in perceptions of peoples' merit and need. We also demonstrated that these computations (for merit, if not need) might be stable and generalizable over time: individuals' perceptual merit sensitivity predicted the degree to which they discriminated between others based on merit in a separate altruism task completed from 27 to 663 days later. Moreover, their perceptual merit bias predicted a general propensity to weigh others' outcomes instead of their own during altruistic choices. Neurally, merit sensitivity (but not bias) was associated with increased activity of the TPJ during perceptual judgments, which in turn predicted merit-related discrimination in altruistic behavior. Together, our results identify a set of distinct neurocomputational mechanisms that contribute to our understanding of *when* and *how* perceptions of others translate into social actions.

Translating perception into action.

Variance in people's sensitivity in merit perceptions predicts context-specific social behaviors and discrimination. This finding contributes to a growing literature regarding parochial altruism. Parochial altruism refers to the tendency to exhibit altruistic behavior towards in-group members and to withhold it from (or even display hostility towards) out-group members [90]. Parochial altruism occurs around the world in private and public settings [91], in sports, politics, war, and religion [92-95], and has also been linked to activation patterns in the TPJ [96-98]. Although we did not use group- or membership-based cues to characterize partners in the altruism task (manipulating merit instead via partner *behaviors* in a separate task), the common locus in the TPJ might indicate shared mechanisms for both types of context-dependent social behaviors. Interestingly, our results suggest a considerable degree of stability of this idiosyncratic perceptual sensitivity across time and contexts, since the social perception and altruism tasks were completed on average ~10 months

apart, up to almost two years for some participants. This finding is consistent with research showing a hereditary component of in-group favoritism and parochial altruism [99]. Future work should examine the extent to which merit-related perceptual sensitivity represents an innate/genetic or learned quantity, and how it correlates with other types of discriminatory behaviors. Our results also speak to the literature on universal altruism: individual differences in the *bias* to perceive merit correlated with the *overall* weight an individual placed on others during altruistic choice months later. This finding supports empirical evidence that dispositionally cooperative people are more universal in their cooperation [100] and suggests that this could partly come from a generalized bias to perceive people as deserving. Although we did not find a distinct neural signature of this bias, future work on genetic or anatomical differences might yield clearer results.

The value of computational decomposition of perception and action.

The power of formal computational models to uncover patterns, principles, and dynamics in social perception and behavior [101-104] has made them an increasingly popular tool in economics, psychology, and neuroscience [105-111]. Our results contribute to this movement by showing how computational models of social perception can provide novel insights into the different computations (i.e., bias and sensitivity) underlying impression formation and its effect on behavior. Although the concept that judgments are a composite of subprocesses is not novel in itself [112], modeling allowed us to formally disentangle these different (neuro)computational mechanisms and demonstrate how they shape different aspects of meaningful social action months (or even years) later. Future research should confirm and extend these findings using other social judgments in other, ideally more ecologically relevant contexts [25]. For example, research suggests that perceptions of others' warmth and competence can impact hiring decisions [113,114]. Are

perceptions of warmth or competence likewise driven by stable individual differences in bias and sensitivity? Do stable individual differences in merit bias or sensitivity shape real-world prosociality, such as political support for social welfare programs [115-117], or the extent to which different people engage in the online posting of degrading content, harsh comments, or cyber-bullying [118,119]?

The neural bases of need and merit perception.

Determining whether socially relevant cues are processed by domain-specific or domain-general neural circuitry remains an active goal of the social and cognitive neurosciences. While considerable evidence exists for specificity in some domains (e.g., face processing [120-122] [122], emotion recognition [127-129], empathy in different modalities [89], or even aspects of moral decision-making [133]), other research points to the broad engagement of the mentalizing network across tasks [134-140]. Given their importance for decision-making, we sought to determine whether perceptions of either need or merit fall within the category of social phenomena processed by dedicated neural circuits. Our neural findings largely suggest the answer is no. Both merit and need perception engaged the mentalizing network to a similar extent and were virtually indistinguishable neurally, with some minor differences. This held true even when applying multivariate decoding approaches, which have been suggested to be more sensitive than traditional univariate analysis techniques [141]. We note, however, that perceptions of others' merit/need in this task likely represent a composite of multiple different sub-components (e.g., related to specific social cues sampled and integrated to yield the final social judgment). Thus, our results do not preclude the existence of domain-specific neural circuitry at a lower level of social appraisals (e.g., gender or age categorization, facial and postural cues that a person is in pain, etc.).

Limitations and future directions.

Our results come with some important limitations. One of the biggest concerns is the puzzling absence of any observed neural or behavioral correlates of need sensitivity, in either the social perception or altruism tasks. We primarily relied on univariate analyses to support our conclusions here. Although a supplementary multivariate pattern analysis yielded little additional insight, exploring alternative methods such as the gradient approach or functional connectivity [142,143] could prove more revealing. One other contributing factor to this null result could simply be the small sample size for our altruism task due to COVID-related delays and participant attrition. Other alternative explanations are also possible, however. While purely speculative, one possibility might lie in a limitation of the stimuli used for the social perception task. Need was generally signaled directly and concretely in each picture (i.e., someone with a knife to their throat, someone with a pained facial expression, etc.). In contrast, merit often had to be inferred indirectly from cues about the cognitive and motivational dispositions of the target (e.g., performance of unethical actions, clothing indicating group membership, other signs this person was a "good" person, etc.). The ability to discriminate based on these abstract cues might thus have correlated with TPJ response and with the use of similarly abstract cues during the altruism task to judge merit of social partners. It is possible that if we had signaled need using similarly abstract cues in the social perception task, we might have observed greater associations with mentalizing network regions and/or with altruistic behavior months later. Future work will be needed to more systematically vary factors related to concrete vs. abstract inference, in larger and more diverse samples, and with a greater range of socially relevant behaviors. Finally, another important open question concerns the origins of the identified perceptual biases and sensitivities. Future research should examine if differences in social perceptions stem from societal or religious values see [144] conveyed during individuals' upbringing, genetic factors, or a combination of both.

Data Availability Statement

Functional imaging data from the social perception task and the altruism task are available at: Conte Social Inference and Context collection (https://nda.nih.gov/edit_collection.html?id=2643). Raw behavioral data for both tasks, ROI masks, computational modeling data used for analysis and to create figures, and the newly created social perception task (including licensing details) are deposited on the Open Science Framework and available at: https://osf.io/4u5vs/.

Code Availability Statement

FMRI data were preprocessed using the open-source fMRIPrep analysis pipeline (see hyperlinks in the manuscript). Analysis scripts for the first-level GLMs and code for computational models are available on OSF. Source code underlying the supplemental MVPA analysis is openly available here [145].

Acknowledgements

This research was supported by funding from NIMH Conte Center P50 MH094258 (C.A.H., A.T.), the Natural Sciences and Engineering Research Council of Canada RGPIN-2019-04329 (A.T.), and the Social Sciences and Humanities Research Council of Canada 435-2016-1274 (C.A.H.). Computations were performed on the Niagara supercomputer at the SciNet HPC Consortium. SciNet is funded by Innovation, Science and Economic Development Canada; the Digital Research Alliance of Canada; the Ontario Research Fund: Research Excellence; and the University of Toronto. We sincerely thank Julian Michael Tyszka and Tim Armstrong for their support during data collection and Jill Jacobson for helpful comments on the manuscript.

References

1 Batson, C. D., Turk, C. L., Shaw, L. L. & Klein, T. R. Information function of empathic emotion: Learning that we value the other's welfare. *Journal of Personality and Social Psychology* **68**, 300-313, doi:10.1037/0022-3514.68.2.300 (1995).

2 Batson, C. D., Eklund, J. H., Chermok, V. L., Hoyt, J. L. & Ortiz, B. G. An additional antecedent of empathic concern: valuing the welfare of the person in need. *Journal of Personality and Social Psychology* **93**, 65-74, doi:10.1037/0022-3514.93.1.65 (2007).

3 Cappelen, A. W., Moene, K. O., Sørensen, E. Ø. & Tungodden, B. Needs Versus Entitlements-an International Fairness Experiment. *Journal of the European Economic Association* **11**, 574-598, doi:10.1111/jeea.12000 (2013).

4 Molouki, S. & Bartels, D. M. Are future selves treated like others? Comparing determinants and levels of intrapersonal and interpersonal allocations. *Cognition* **196**, 104150, doi:10.1016/j.cognition.2019.104150 (2020).

5 Batson, C. D., Bolen, M. H., Cross, J. A. & Neuringer-Benefiel, H. E. Where is the altruism in the altruistic personality? *Journal of Personality and Social Psychology* **50**, 212–220, doi:10.1037/0022-3514.50.1.212 (1986).

6 Hare, T. A., Camerer, C. F., Knoepfle, D. T., O'Doherty, J. P. & Rangel, A. Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience* **30**, 583-590, doi:10.1523/JNEUROSCI.4089-09.2010 (2010).

7 Tusche, A., Bockler, A., Kanske, P., Trautwein, F. M. & Singer, T. Decoding the Charitable Brain: Empathy, Perspective Taking, and Attention Shifts Differentially Predict Altruistic Giving. *Journal of Neuroscience* **36**, 4719-4732, doi:10.1523/JNEUROSCI.3392-15.2016 (2016).

8 Fong, C. M. Evidence from an experiment on charity to welfare recipients: Reciprocity, altruism and the empathic responsiveness hypothesis. *Economic Journal* **117**, 1008-1024, doi:10.1111/j.1468-0297.2007.02076.x (2007).

9 Eckel, C. C. & Grossman, P. J. Altruism in anonymous dictator games. *Games and Economic Behavior* **16**, 181-191, doi:10.1006/game.1996.0081 (1996).

10 Van Doesum, N. J., Van Lange, P. A. M., Tybur, J. M., Leal, A. & Van Dijk, E. People from lower social classes elicit greater prosociality: Compassion and deservingness matter. *Group Processes & Intergroup Relations* **25**, 1064-1083, doi:10.1177/1368430220982072 (2021).

11 Wilson, C. The role of a merit principle in distributive justice. *The Journal of ethics* **7**, 277-314 (2003).

12 Green, D. M. & Swets, J. A. *Signal detection theory and psychophysics*. (John Wiley, 1966).

13 LaBerge, D., Legrand, R. & Hobbie, R. K. Functional identification of perceptual and response biases in choice reaction time. *Journal of Experimental Psychology* **79**, 295-299, doi:10.1037/h0026892 (1969).

14 FeldmanHall, O., Dalgleish, T., Evans, D. & Mobbs, D. Empathic concern drives costly altruism. *Neuroimage* **105**, 347-356, doi:10.1016/j.neuroimage.2014.10.043 (2015).

15 Murray, A. D. Infant crying as an elicitor of parental behavior: an examination of two models. *Psychological Bulletin* **86**, 191–215, doi:10.1037/0033-2909.86.1.191 (1979).

16 Marsh, A. A. Neural, cognitive, and evolutionary foundations of human altruism. *Wiley Interdisciplinary Reviews: Cognitive Science* **7**, 59-71, doi:10.1002/wcs.1377 (2016).
17 Small, D. A. & Verrochi, N. M. The Face of Need: Facial Emotion Expression on Charity Advertisements. *Journal of Marketing Research* **46**, 777-787, doi:10.1509/jmkr.46.6.777 (2009).
18 Rand, D. G. & Epstein, Z. G. Risking your life without a second thought: Intuitive decision-making and extreme altruism. *PloS One* **9**, e109687, doi:10.1371/journal.pone.0109687 (2014).
19 Vieira, J. B., Schellhaas, S., Enström, E. & Olsson, A. Help or flight? Increased threat imminence promotes defensive helping in humans. *Proceedings of the Royal Society B* **287**, 20201473, doi:10.1098/rspb.2020.1473 (2020).
20 Burt, C. D. & Strongman, K. Use of images in charity advertising: Improving donations and compliance rates. *International Journal of Organisational Behaviour* **8**, 571-580 (2005).
21 Stürmer, S., Snyder, M. & Omoto, A. M. Prosocial emotions and helping: the moderating role of group membership. *Journal of Personality and Social Psychology* **88**, 532–546, doi:10.1037/0022-3514.88.3.532 (2005).
22 Opotow, S. Moral exclusion and injustice: An introduction. *Journal of Social Issues* **46**, 1–20, doi:10.1111/j.1540-4560.1990.tb00268.x (1990).
23 Van de Vondervoort, J. W., Aknin, L. B., Kushnir, T., Slevinsky, J. & Hamlin, J. K. Selectivity in toddlers' behavioral and emotional reactions to prosocial and antisocial others. *Developmental Psychology* **54**, 1–14, doi:10.1037/dev0000404 (2018).
24 Santavirta, S. *et al.* Functional organization of social perception networks in the human brain. *NeuroImage* **272**, 120025, doi:10.1016/j.neuroimage.2023.120025 (2023).
25 Thornton, M. A. & Mitchell, J. P. Theories of person perception predict patterns of neural activity during mentalizing. *Cerebral Cortex* **28**, 3505-3520, doi:10.1093/cercor/bhx216 (2018).
26 Van Overwalle, F. Social cognition and the brain: a meta-analysis. *Human brain mapping* **30**, 829-858, doi:10.1002/hbm.20547 (2009).
27 Schurz, M., Radua, J., Aichhorn, M., Richlan, F. & Perner, J. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews* **42**, 9-34, doi:10.1016/j.neubiorev.2014.01.009 (2014).
28 Molenberghs, P., Johnson, H., Henry, J. D. & Mattingley, J. B. Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews* **65**, 276-291, doi:10.1016/j.neubiorev.2016.03.020 (2016).
29 Bzdok, D. *et al.* Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function* **217**, 783-796, doi:10.1007/s00429-012-0380-y (2012).
30 Engel, C. Dictator games: a meta study. *Experimental Economics* **14**, 583-610, doi:10.1007/s10683-011-9283-7 (2011).
31 Jilke, S. & Tummers, L. Which clients are deserving of help? A theoretical model and experimental test. *Journal of Public Administration Research and Theory* **28**, 226-238, doi:10.1093/jopart/muy002 (2018).
32 Batson, C. D. *et al.* " I've been there, too": Effect on empathy of prior experience with a need. *Personality and Social Psychology Bulletin* **22**, 474-482, doi:10.1177/0146167296225005 (1996).

33 Tusche, A. & Bas, L. M. Neurocomputational models of altruistic decision-making and social motives: Advances, pitfalls, and future directions. *Wiley Interdisciplinary Reviews: Cognitive Science* **12**, e1571, doi:10.1002/wcs.1571 (2021).
34 Chen, F. D., Zhu, Z., Shen, Q., Krajbich, I. & Hare, T. A. Intrachoice Dynamics Shape Social Decisions. *Management Science*, doi:10.1287/mnsc.2023.4732 (2023).
35 Teoh, Y. Y., Yao, Z., Cunningham, W. A. & Hutcherson, C. A. Attentional priorities drive effects of time pressure on altruistic choice. *Nature Communications* **11**, 3534, doi:10.1038/s41467-020-17326-x (2020).
36 Mayr, U. & Freund, A. M. Do We Become More Prosocial as We Age, and if So, Why? *Current Directions in Psychological Science* **29**, 248-254, doi:10.1177/0963721420910811 (2020).
37 Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C. & Fehr, E. Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron* **75**, 73-79, doi:10.1016/j.neuron.2012.05.021 (2012).
38 Fehr, E. & Schmidt, K. M. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* **114**, 817-868, doi:10.1162/003355399556151 (1999).
39 Charness, G. & Rabin, M. Understanding social preferences with simple tests. *Quarterly Journal of Economics* **117**, 817-869, doi:10.1162/003355302760193904 (2002).
40 Tricomi, E., Rangel, A., Camerer, C. F. & O'Doherty, J. P. Neural evidence for inequality-averse social preferences. *Nature* **463**, 1089-1091, doi:10.1038/nature08785 (2010).
41 Tabibnia, G., Satpute, A. B. & Lieberman, M. D. The sunny side of fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science* **19**, 339-347, doi:10.1111/j.1467-9280.2008.02091.x (2008).
42 Rhoads, S. A., Cutler, J. & Marsh, A. A. A feature-based network analysis and fMRI meta-analysis reveal three distinct types of prosocial decisions. *Social cognitive and affective neuroscience* **16**, 1214-1233, doi:10.1093/scan/nsab079 (2021).
43 Carlson, R. W., Maréchal, M. A., Oud, B., Fehr, E. & Crockett, M. J. Motivated misremembering of selfish decisions. *Nature communications* **11**, 2100, doi:10.1038/s41467-020-15602-4 (2020).
44 Saulin, A., Horn, U., Lotze, M., Kaiser, J. & Hein, G. The neural computation of human prosocial choices in complex motivational states. *NeuroImage* **247**, 118827, doi:10.1016/j.neuroimage.2021.118827 (2022).
45 Hutcherson, C. A., Bushong, B. & Rangel, A. A Neurocomputational Model of Altruistic Choice and Its Implications. *Neuron* **87**, 451-462, doi:10.1016/j.neuron.2015.06.031 (2015).
46 Hu, J., Konovalov, A. & Ruff, C. C. A unified neural account of contextual and individual differences in altruism. *Elife* **12**, e80667, doi:10.7554/eLife.80667 (2023).
47 Hu, J., Hu, Y., Li, Y. & Zhou, X. Computational and neurobiological substrates of cost-benefit integration in altruistic helping decision. *Journal of Neuroscience* **41**, 3545-3561, doi:10.1523/JNEUROSCI.1939-20.2021 (2021).
48 Sul, S. *et al.* Spatial gradient in value representation along the medial prefrontal cortex reflects individual differences in prosociality. *Proceedings of the National Academy of Sciences* **112**, 7851-7856, doi:10.1073/pnas.1423895112 (2015).

49 Li, Y., Hu, J., Ruff, C. C. & Zhou, X. Neurocomputational evidence that conflicting prosocial motives guide distributive justice. *Proceedings of the National Academy of Sciences* **119**, e2209078119, doi:10.1073/pnas.2209078119 (2022).
50 Kliemann, D. *et al.* Caltech Conte Center, a multimodal data resource for exploring social cognition and decision-making. *Scientific Data* **9**, 138, doi:10.1038/s41597-022-01171-2 (2022).
51 Wechsler, D. Wechsler Abbreviated Scale of Intelligence (WASI) [Database record]. APA PsycTests., doi:10.1037/t15170-000 (1999).
52 Wechsler, D. Wechsler Abbreviated Scale of Intelligence--Second Edition (WASI-II) [Database record]. APA PsycTests., doi:10.1037/t15171-000 (2011).
53 Spunt, R. P. & Adolphs, R. Validating the Why/How contrast for functional MRI studies of Theory of Mind. *Neuroimage* **99**, 301-311, doi:10.1016/j.neuroimage.2014.05.023 (2014).
54 Spunt, R. P. & Adolphs, R. Folk explanations of behavior: a specialized use of a domain-general mechanism. *Psychological Science* **26**, 724-736, doi:10.1177/0956797615569002 (2015).
55 Brainard, D. H. The Psychophysics Toolbox. *Spatial Vision* **10**, 433-436, doi:10.1163/156856897X00357 (1997).
56 Kleiner, M., Brainard, D. & Pelli, D. What is new in Psychophysics Toolbox. *Perception* **36**, 416 (2007).
57 Pelli, D. G. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* **10**, 437-442, doi:10.1163/156856897x00366 (1997).
58 Tusche, A. & Hutcherson, C. A. Cognitive regulation alters social and dietary choice by changing attribute representations in domain-general and domain-specific brain circuits. *Elife* **7**, e31185, doi:10.7554/eLife.31185 (2018).
59 Peirce, J. W. PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods* **162**, 8-13, doi:10.1016/j.jneumeth.2006.11.017 (2007).
60 Peirce, J. W. Generating Stimuli for Neuroscience Using PsychoPy. *Frontiers in Neuroinformatics* **2**, doi:10.3389/neuro.11.010.2008 (2009).
61 Singer, T. *et al.* Empathic neural responses are modulated by the perceived fairness of others. *Nature* **439**, 466-469, doi:10.1038/nature04271 (2006).
62 Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J. & Frith, C. D. Brain responses to the acquired moral status of faces. *Neuron* **41**, 653-662, doi:10.1016/s0896-6273(04)00014-5 (2004).
63 Hines, E. A. & Brown, G. E. The cold pressor test for measuring the reactibility of the blood pressure: Data concerning 571 normal and hypertensive subjects. *American Heart Journal* **11**, 1-9, doi:10.1016/s0002-8703(36)90370-8 (1936).
64 Ratcliff, R., Smith, P. L., Brown, S. D. & McKoon, G. Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences* **20**, 260-281, doi:10.1016/j.tics.2016.01.007 (2016).
65 Krajbich, I., Armel, C. & Rangel, A. Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience* **13**, 1292-1298, doi:10.1038/nn.2635 (2010).
66 Krajbich, I. & Rangel, A. Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences* **108**, 13852-13857, doi:10.1073/pnas.1101328108 (2011).

67 Towal, R. B., Mormann, M. & Koch, C. Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences* **110**, E3858-E3867, doi:10.1073/pnas.1304429110 (2013).
68 Roberts, I. D. & Hutcherson, C. A. Affect and Decision Making: Insights and Predictions from Computational Models. *Trends in Cognitive Sciences* **23**, 602-614, doi:10.1016/j.tics.2019.04.005 (2019).
69 Krajbich, I., Hare, T., Bartling, B., Morishima, Y. & Fehr, E. A common mechanism underlying food choice and social decisions. *PLoS Computational Biology* **11**, e1004371, doi:10.1371/journal.pcbi.1004371 (2015).
70 Kutlikova, H. H., Zhang, L., Eisenegger, C., van Honk, J. & Lamm, C. Testosterone eliminates strategic prosocial behavior through impacting choice consistency in healthy males. *Neuropsychopharmacology* **48**, 1541–1550, doi:10.1038/s41386-023-01570-y (2023).
71 Harris, A., Clithero, J. A. & Hutcherson, C. A. Accounting for taste: A multi-attribute neurocomputational model explains the neural dynamics of choices for self and others. *Journal of Neuroscience* **38**, 7952-7968, doi:10.1523/JNEUROSCI.3327-17.2018 (2018).
72 Witt, J. K., Taylor, J. E. T., Sugovic, M. & Wixted, J. T. Signal detection measures cannot distinguish perceptual biases from response biases. *Perception* **44**, 289-300, doi:10.1068/p7908 (2015).
73 Hutcherson, C. A. & Tusche, A. Evidence accumulation, not 'self-control', explains dorsolateral prefrontal activation during normative choice. *Elife* **11**, e65661, doi:10.7554/eLife.65661 (2022).
74 Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, doi:10.1038/s41592-018-0235-4 (2018).
75 Esteban, O., Ross Blair, Christopher J. Markiewicz, Shoshana L. Berleant, Craig Moodie, Feilong Ma, Ayse Ilkay Isik, et al. FMRIPrep. *Software* **Zenodo**, doi:10.5281/zenodo.852659 (2018).
76 Gorgolewski, K. *et al.* Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics* **5**, 13, doi:10.3389/fninf.2011.00013 (2011).
77 Gorgolewski, K. J. *et al.* Nipype. *Software* **Zenodo**, doi:10.5281/zenodo.596855 (2018).
78 Straccia, M. A. *et al.* Null results of oxytocin and vasopressin administration on mentalizing in a large fMRI sample: evidence from a randomized controlled trial. *Psychological Medicine* **53**, 2285-2295, doi:10.1017/S0033291721004104 (2021).
79 Tan, K. M., Burklund, L. J., Craske, M. G. & Lieberman, M. D. Posttraumatic stress disorder and the social brain: Affect-related disruption of the default and mirror networks. *Depression and Anxiety* **36**, 1058-1071, doi:10.1002/da.22953 (2019).
80 Thompson, E. L., Bird, G. & Catmur, C. Mirror neuron brain regions contribute to identifying actions, but not intentions. *Human Brain Mapping* **43**, 4901-4913, doi:10.1002/hbm.26036 (2022).
81 Cardenas, S. I. *et al.* Theory of mind processing in expectant fathers: Associations with prenatal oxytocin and parental attunement. *Developmental Psychobiology* **63**, 1549-1567, doi:10.1002/dev.22115 (2021).
82 Tusche, A., Spunt, R. P., Paul, L. K., Tyszka, J. M. & Adolphs, R. Neural signatures of social inferences predict the number of real-life social contacts and autism severity. *Nature Communications* **14**, 4399, doi:10.1038/s41467-023-40078-3 (2023).

83 Pineda, J. A. Sensorimotor cortex as a critical component of an 'extended' mirror neuron system: Does it solve the development, correspondence, and control problems in mirroring? *Behavioral and Brain Functions* **4**, 1-16, doi:10.1186/1744-9081-4-47 (2008).
84 Van Overwalle, F. & Baetens, K. Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage* **48**, 564-584, doi:10.1016/j.neuroimage.2009.06.009 (2009).
85 Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences* **103**, 3863-3868, doi:10.1073/pnas.0600244103 (2006).
86 Haynes, J.-D. & Rees, G. Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience* **7**, 523-534, doi:10.1038/nrn1931 (2006).
87 Heinzle, J. *et al.* Multivariate decoding of fMRI data: Towards a content-based cognitive neuroscience. *e-Neuroforum* **18**, 1-16, doi:10.1007/s13295-012-0026-9 (2012).
88 Brooks, J. A., Stolier, R. M. & Freeman, J. B. Computational approaches to the neuroscience of social perception. *Social Cognitive and Affective Neuroscience* **16**, 827-837, doi:10.1093/scan/nsaa127 (2021).
89 Corradi-Dell'Acqua, C., Tusche, A., Vuilleumier, P. & Singer, T. Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. *Nature Communications* **7**, 10904, doi:10.1038/ncomms10904 (2016).
90 Bernhard, H., Fischbacher, U. & Fehr, E. Parochial altruism in humans. *Nature* **442**, 912-915, doi:10.1038/nature04981 (2006).
91 Romano, A., Sutter, M., Liu, J. H., Yamagishi, T. & Balliet, D. National parochialism is ubiquitous across 42 nations around the world. *Nature communications* **12**, 4456, doi:10.1038/s41467-021-24787-1 (2021).
92 Hein, G., Silani, G., Preuschoff, K., Batson, C. D. & Singer, T. Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron* **68**, 149-160, doi:10.1016/j.neuron.2010.09.003 (2010).
93 Ginges, J., Hansen, I. & Norenzayan, A. Religion and support for suicide attacks. *Psychological Science* **20**, 224-230, doi:10.1111/j.1467-9280.2009.02270.x (2009).
94 Choi, J.-K. & Bowles, S. The coevolution of parochial altruism and war. *Science* **318**, 636-640, doi:10.1126/science.1144237 (2007).
95 Brewer, M. B., Buchan, N. R., Ozturk, O. D. & Grimalda, G. Parochial altruism and political ideology. *Political Psychology* **44**, 383-396, doi:10.1111/pops.12852 (2023).
96 Baumgartner, T., Götte, L., Gügler, R. & Fehr, E. The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human Brain Mapping* **33**, 1452-1469, doi:10.1002/hbm.21298 (2012).
97 Obeso, I., Moisa, M., Ruff, C. C. & Dreher, J.-C. A causal role for right temporo-parietal junction in signaling moral conflict. *Elife* **7**, e40671, doi:10.7554/eLife.40671 (2018).
98 Fujino, J. *et al.* Role of the right temporoparietal junction in intergroup bias in trust decisions. *Human brain mapping* **41**, 1677-1688, doi:10.1002/hbm.24903 (2020).
99 Lewis, G. J. & Bates, T. C. Genetic evidence for multiple biological mechanisms underlying in-group favoritism. *Psychological Science* **21**, 1623-1628, doi:10.1177/0956797610387439 (2010).
100 Romano, A., Balliet, D., Yamagishi, T. & Liu, J. H. Parochial trust and cooperation across 17 societies. *Proceedings of the National Academy of Sciences* **114**, 12702-12707, doi:10.1073/pnas.1712921114 (2017).

101 Crockett, M. J. How Formal Models Can Illuminate Mechanisms of Moral Judgment and Decision Making. *Current Directions in Psychological Science* **25**, 85-90, doi:10.1177/0963721415624012 (2016).
102 Johnson, D. J., Hopwood, C. J., Cesario, J. & Pleskac, T. J. Advancing research on cognitive processes in social and personality psychology: A hierarchical drift diffusion model primer. *Social Psychological and Personality Science* **8**, 413-423, doi:10.1177/1948550617703174 (2017).
103 Hackel, L. M. & Amodio, D. M. Computational neuroscience approaches to social cognition. *Current Opinion in Psychology* **24**, 92-97, doi:10.1016/j.copsyc.2018.09.001 (2018).
104 Freeman, J. B., Stolier, R. M., Brooks, J. A. & Stillerman, B. S. The neural representational geometry of social perception. *Current Opinion in Psychology* **24**, 83-91, doi:10.1016/j.copsyc.2018.10.003 (2018).
105 Capraro, V. & Perc, M. Mathematical foundations of moral preferences. *Journal of the Royal Society Interface* **18**, 20200880, doi:10.1098/rsif.2020.0880 (2021).
106 Kvaran, T. & Sanfey, A. G. Toward an integrated neuroscience of morality: the contribution of neuroeconomics to moral cognition. *Topics in Cognitive Science* **2**, 579-595, doi:10.1111/j.1756-8765.2010.01086.x (2010).
107 Kliemann, D. & Adolphs, R. The social neuroscience of mentalizing: challenges and recommendations. *Current Opinion in Psychology* **24**, 1-6, doi:https://doi.org/10.1016/j.copsyc.2018.02.015 (2018).
108 Charpentier, C. J. & O'Doherty, J. P. The application of computational models to social neuroscience: promises and pitfalls. *Social Neuroscience* **13**, 637-647, doi:10.1080/17470919.2018.1518834 (2018).
109 Behrens, T. E., Hunt, L. T. & Rushworth, M. F. The computation of social behavior. *Science* **324**, 1160-1164, doi:10.1126/science.1169694 (2009).
110 Freeman, J. B. & Ambady, N. A dynamic interactive theory of person construal. *Psychological Review* **118**, 247, doi:10.1037/a0022327 (2011).
111 Yu, H., Siegel, J. Z. & Crockett, M. J. Modeling Morality in 3-D: Decision-Making, Judgment, and Inference. *Topics in Cognitive Science* **11**, 409-432, doi:10.1111/tops.12382 (2019).
112 Kenny, D. A. & Albright, L. Accuracy in interpersonal perception: a social relations analysis. *Psychological Bulletin* **102**, 390-402, doi:10.1037/0033-2909.102.3.390 (1987).
113 Cuddy, A. J. C., Glick, P. & Beninger, A. The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in Organizational Behavior* **31**, 73-98, doi:10.1016/j.riob.2011.10.004 (2011).
114 Louvet, E. Social judgment toward job applicants with disabilities: Perception of personal qualities and competences. *Rehabilitation Psychology* **52**, 297-303, doi:10.1037/0090-5550.52.3.297 (2007).
115 Appelbaum, L. D. Who deserves help? Students' opinions about the deservingness of different groups living in Germany to receive aid. *Social Justice Research* **15**, 201-225, doi:10.1023/A:1021025528408 (2002).
116 Attewell, D. Deservingness perceptions, welfare state support and vote choice in Western Europe. *West European Politics* **44**, 611-634, doi:10.1080/01402382.2020.1715704 (2021).

117 Oorschot, W. v. Who should get what, and why? On deservingness criteria and the conditionality of solidarity among the public. *Policy & Politics* **28**, 33–48, doi:10.1332/0305573002500811 (2000).
118 Coe, K., Kenski, K. & Rains, S. A. Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication* **64**, 658-679, doi:10.1111/jcom.12104 (2014).
119 Tokunaga, R. S. Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in Human Behavior* **26**, 277-287, doi:10.1016/j.chb.2009.11.014 (2010).
120 Elbich, D. B., Molenaar, P. C. & Scherf, K. S. Evaluating the organizational structure and specificity of network topology within the face processing system. *Human brain mapping* **40**, 2581-2595, doi:10.1002/hbm.24546 (2019).
121 Itier, R. J., Alain, C., Sedore, K. & McIntosh, A. R. Early face processing specificity: it's in the eyes! *Journal of cognitive neuroscience* **19**, 1815-1826, doi:10.1162/jocn.2007.19.11.1815 (2007).
122 Kanwisher, N. Domain specificity in face perception. *Nature Neuroscience* **3**, 759-763, doi:10.1038/77664 (2000).
123 Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience* **17**, 4302-4311, doi:10.1523/JNEUROSCI.17-11-04302.1997 (1997).
124 Kanwisher, N. & Yovel, G. The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 2109-2128, doi:10.1098/rstb.2006.1934 (2006).
125 Beauchamp, M. S. The social mysteries of the superior temporal sulcus. *Trends in Cognitive Sciences* **19**, 489-490, doi:10.1016/j.tics.2015.07.002 (2015).
126 Deen, B., Koldewyn, K., Kanwisher, N. & Saxe, R. Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebral Cortex* **25**, 4596-4609, doi:10.1093/cercor/bhv111 (2015).
127 Ekman, P., Levenson, R. W. & Friesen, W. V. Autonomic nervous system activity distinguishes among emotions. *Science* **221**, 1208-1210, doi:10.1126/science.6612338 (1983).
128 Adolphs, R., Tranel, D., Damasio, H. & Damasio, A. Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* **372**, 669-672, doi:10.1038/372669a0 (1994).
129 Morris, J. S. *et al.* A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* **383**, 812-815, doi:10.1038/383812a0 (1996).
130 Calder, A. J. Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology* **13**, 699-745, doi:10.1080/026432996381890 (1996).
131 Whalen, P. J. *et al.* A functional MRI study of human amygdala responses to facial expressions of fear versus anger. *Emotion* **1**, 70–83, doi:10.1037/1528-3542.1.1.70 (2001).
132 Ashley, V., Vuilleumier, P. & Swick, D. Time course and specificity of event-related potentials to emotional expressions. *Neuroreport* **15**, 211-216, doi:10.1097/00001756-200401190-00041 (2004).

133 Hutcherson, C. A., Montaser-Kouhsari, L., Woodward, J. & Rangel, A. Emotional and Utilitarian Appraisals of Moral Dilemmas Are Encoded in Separate Areas and Integrated in Ventromedial Prefrontal Cortex. *Journal of Neuroscience* **35**, 12593-12605, doi:10.1523/JNEUROSCI.3402-14.2015 (2015).
134 Barrett, L. F. & Satpute, A. B. Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology* **23**, 361-372, doi:10.1016/j.conb.2012.12.012 (2013).
135 Schuwerk, T., Schurz, M., Müller, F., Rupprecht, R. & Sommer, M. The rTPJ's overarching cognitive function in networks for attention and theory of mind. *Social cognitive and affective neuroscience* **12**, 157-168, doi:10.1093/scan/nsw163 (2017).
136 Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N. & Saxe, R. Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PloS one* **4**, e4869, doi:10.1371/journal.pone.0004869 (2009).
137 Mitchell, J. P. Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex* **18**, 262-271, doi:10.1093/cercor/bhm051 (2008).
138 Baetens, K., Vandekerckhove, M. & Van Overwalle, F. Social Versus Nonsocial Reasoning. *Brain Mapping: An Encyclopedic Reference* **3**, 227–230, doi:10.1016/B978-0-12-397025-1.00184-6 (2015).
139 Carter, R. M. & Huettel, S. A. A nexus model of the temporal–parietal junction. *Trends in Cognitive Sciences* **17**, 328-336, doi:10.1016/j.tics.2013.05.007 (2013).
140 Lugrin, C., Konovalov, A. & Ruff, C. Salience-based information integration: An overarching function of the "social brain"? *bioRxiv*, 2023.2001.2030.525877, doi:10.1101/2023.01.30.525877 (2023).
141 Kriegeskorte, N. & Bandettini, P. Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage* **38**, 649-662, doi:10.1016/j.neuroimage.2007.02.022 (2007).
142 Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences* **113**, 12574-12579, doi:10.1073/pnas.1608282113 (2016).
143 Vos de Wael, R. *et al.* BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Communications Biology* **3**, 103, doi:10.1038/s42003-020-0794-7 (2020).
144 Amormino, P. *et al.* Beliefs about humanity, not higher power, predict extraordinary altruism. *Journal of Research in Personality* **101**, 104313, doi:10.1016/j.jrp.2022.104313 (2022).
145 Hebart, M. N., Görgen, K. & Haynes, J.-D. The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in neuroinformatics* **8**, 88, doi:10.3389/fninf.2014.00088 (2015).