



The behavioral and neural signatures of distinct conceptions of fairness

Laura Niemi, Emily Wasserman & Liane Young

To cite this article: Laura Niemi, Emily Wasserman & Liane Young (2018) The behavioral and neural signatures of distinct conceptions of fairness, *Social Neuroscience*, 13:4, 399-415, DOI: [10.1080/17470919.2017.1333452](https://doi.org/10.1080/17470919.2017.1333452)

To link to this article: <https://doi.org/10.1080/17470919.2017.1333452>



View supplementary material [↗](#)



Accepted author version posted online: 19 May 2017.
Published online: 06 Jun 2017.



Submit your article to this journal [↗](#)



Article views: 217



View Crossmark data [↗](#)



Citing articles: 4 View citing articles [↗](#)

ARTICLE



The behavioral and neural signatures of distinct conceptions of fairness

Laura Niemi^a, Emily Wasserman^b and Liane Young^b

^aDepartment of Psychology, Harvard University, Cambridge, MA, USA; ^bDepartment of Psychology, Boston College, Chestnut Hill, MA, USA

ABSTRACT

Adhering to standard procedures (impartiality), returning favors (reciprocity) or giving based on individuals' needs (charity) may all be considered moral and/or fair ways to allocate resources. However, these allocation behaviors may be perceived as differently motivated, and their moral evaluation may make different demands on theory of mind (ToM) – the capacity to process information about mental states, including motives. In Studies 1 and 2, we examined participants' moral judgments of allocations based on (1) impartiality, (2) reciprocity, (3) charity and (4) unspecified criteria as depicted in vignettes, as well as participants' perceptions of allocators' motivations. In Study 3, we used functional magnetic resonance imaging to investigate how brain regions for ToM were recruited during moral evaluation of the same vignettes. Reciprocity and charity were processed similarly, in that they recruited ToM regions to the same extent, i.e., precuneus, dorsal and ventral medial prefrontal cortex and left temporoparietal junction (LTPJ). In turn, impartiality and the unspecified condition were processed similarly, recruiting the same ToM regions to a lesser extent. Nevertheless, reciprocity elicited greater activity relative to impartiality and unspecified in the ToM regions of interest. Overall, evaluations of different allocation behaviors depend differently on ToM, with charity and reciprocity eliciting greater attention to individuals' unique states and motivations.

ARTICLE HISTORY

Received 20 September 2016
Revised 15 April 2017
Published online 6 June 2017

KEYWORDS

Fairness; resource allocation; theory of mind; moral judgment; moral cognition; fMRI

Introduction

Much of the work in moral psychology has focused on how people weigh competing moral values, such as deontological versus utilitarian concerns (Gleichgerrcht & Young, 2013; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Koenigs et al., 2007) and concerns for loyalty versus fairness (Waytz, Dungan, & Young, 2013). In this research, we examine possibly distinct values *within* the moral domain of fairness. On the one hand, research examining individual differences in moral values suggests that fairness, along with concern about harm, is universally endorsed across cultures and across the political spectrum (e.g., Graham et al., 2011; Gray, Schein, & Ward, 2014; Gray, Young, & Waytz, 2012; Niemi & Young, 2013; Rozin, Lowery, Imada, & Haidt, 1999). On the other hand, fairness can be difficult to pin down. Fairness is prone to being defined tautologically and flexibly invoked in motivated reasoning: “Fair is fair!” (Rasinski, 1987; Valdesolo & DeSteno, 2007, 2008; Wierzbicka, 2006). Democratic U.S. President Barack Obama has stressed the need for “a fair shot and a fair shake”, and at the same time the most reliable source for a conservative take on political issues brandishes the slogan “Fair & Balanced” (FoxNews).

One possibility considered in the present work is that fairness values appear to be among the most widely shared moral values precisely because people are able to endorse multiple and at times opposing conceptions of fairness.

In contrast to extensive prior work examining people's responses to fair or unfair outcomes, including work revealing reward-related neural activity for equal outcomes even in the absence of material utility (Tabibnia, Satpute, & Lieberman, 2008), the current research focuses on people's inferences about the intentions and motivations underlying resource allocation methods. Prior work has revealed a network of brain regions involved in social cognition and, more specifically, theory of mind (ToM), the capacity to represent and reason about mental states (e.g., intentions, motivations), including the right and left temporoparietal junction (RTPJ and LTPJ), precuneus, and dorsal and ventral medial prefrontal cortex (DMPFC and VMPFC) (e.g., Castelli, Happe, Frith, & Frith, 2000; Fletcher et al., 1995; Gobbini, Koralek, Bryan, Montgomery, & Haxby, 2007; Jenkins & Mitchell, 2010; Saxe & Kanwisher, 2003; Spunt & Adolphs, 2015; Völlm et al., 2005). These regions are recruited for calculations of moral blame and praise, when these calculations

require representing agents' motivational states, e.g., did she mean to poison his coffee, or did she do it by accident? (Buckholtz & Marois, 2012; Greene & Haidt, 2002; Moll, De Oliveira-Souza, Bramati, & Grafman, 2002; Yoder & Decety, 2014; Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010; Young, Cushman, Hauser, & Saxe, 2007). An open question concerns the role of ToM for processing distinct forms of fairness. We propose that, although considerations of impartiality, reciprocity and charity are all recognized on some level as relevant to the concept of fairness, an important distinction may be rooted in assessments of agents' underlying motivations; therefore, moral judgments of different allocation methods may rely differently on ToM. Later, we review literatures on impartiality, reciprocity and charity, to situate our own work.

Impartiality

The human preference for impartiality emerges early (Fehr, Bernhard, & Rockenbach, 2008; Geraci & Surian, 2011; Schmidt & Sommerville, 2011; Sloane, Baillargeon, & Premack, 2012; Sommerville, Schmidt, Yun, & Burns, 2013) and is widely observed to endure into adulthood (Dawes, Fowler, Johnson, McElreath & Smirnov, 2007; Loewenstein, Thompson, & Bazerman, 1989; Shaw & Knobe, 2013; Van Lange, 1999; Van Lange, Otten, De Bruin, & Joireman, 1997; Van Lange, Schippers, & Balliet, 2011). Does moral evaluation of impartial allocation require rich representation of allocators' mental states? On the one hand, recent work suggests that preferences for impartiality reflect people's motivations to avoid either being or appearing biased (Shaw, 2013). Evaluating impartial allocations may therefore involve inferring and evaluating such motivations. On the other hand, impartiality by its nature involves using rule-based procedures that allow the allocator to sidestep social demands, including the pull of individual people's needs (as in the case of charity) and/or alliances (as in the case of reciprocity). When an allocator uses a standardized procedure to allocate resources, people may be less likely to consider the allocator's intentions. To the extent that people are convinced someone is "playing by the rules," they might assume the individual is not acting according to his or her own personal interests. Thus, we hypothesize that moral evaluation of impartial allocators depends less on ToM.

Reciprocity

Prominent theories of moral psychology such as Moral Foundations Theory have identified reciprocity and fairness as belonging to the same moral domain (e.g.,

Graham et al., 2011; Haidt, 2007). Like impartiality, reciprocity is a fundamental social norm (Axelrod & Hamilton, 1981; Baumann, Andre, & Sperber, 2013; Gurven, 2006; Hill & Kaplan, 1993; Rand & Nowak, 2013; Trivers, 1971; Wedekind & Milinski, 2000) that emerges early (Harris, 1970; Kenward, Hellmer, Söderström-Winter, & Eriksson, 2015; Olson & Spelke, 2008) and over a wide range of contexts, from patterns of food-sharing in anthropological research (Gurven, 2006; Hill & Kaplan, 1993) to the Golden Rule ("Do unto others as you would have them do unto you") (Batson et al., 2003). A number of features of reciprocity in particular suggest that its moral evaluation may robustly trigger ToM. Reciprocity relies on the ability of an allocator to "keep score," or to distinguish among individuals and remember past behaviors (Axelrod & Hamilton, 1981; Gurven, 2006; Trivers, 1971), as opposed to adopting a "veil of ignorance" (Rawls, 1971, 2001). Although reciprocity may appear appropriate within a dyadic interaction, it may also be viewed as self-serving in a group context insofar as reciprocity privileges an individual who contributed to the allocator's own personal benefit (Elster, 2006; Gurven, 2006). Here we directly examine whether people infer self-serving motives when evaluating allocators who operate on reciprocity-based criteria.

Charity

Like impartiality, charity may be rooted in equality because need-based allocations have the effect of "leveling the playing field" (Deutsch, 1975; Meindl, Iyer, & Graham, 2017; Shaw & Olson, 2012). Nevertheless, there is a way in which charitable allocations could be seen as unfair, since, like reciprocity, charity involves distinguishing among individuals. Furthermore, charity has been likened to reciprocity extended over a longer period of time (Gurven, 2006; Trivers, 1971). For example, food-sharing in forager societies that favors lower-producing families, as opposed to those contributing the most, is hypothesized to function as "social insurance" – the benefits of such a system to high-producing allocators emerge when they are down on their luck (Gurven, 2006). According to Trivers (1971), giving to the recipient most in need is *most* personally advantageous. This recipient is maximally grateful and most tightly bound to reciprocity norms that favor the allocator in the future. In other words, need-based allocations need not stem from an equality preference, but from a desire to build reciprocal social ties. Moral evaluation of charity may therefore elicit attention to allocators' motives.

Overview of the present research

Prior studies on fairness have largely focused on contrasting equal or fair outcomes with unequal or unfair outcomes. The aim of the current work is to characterize, in behavioral and neural detail, people's processing of distinct forms of fairness when outcomes are held constant (one person/group always benefits more than others), and allocation criteria are varied. In all studies, participants read vignettes featuring protagonists who allocated resources in the context of work, school, family or recreational situations based on criteria representative of impartiality, reciprocity or charity. In two initial behavioral studies, participants delivered ratings of fairness and moral praiseworthiness (Study 1) and protagonists' motivations (Study 2). Study 3 used functional magnetic resonance imaging (fMRI) to investigate whether morally evaluating allocations based on impartiality, reciprocity and charity differentially recruits ToM. We leveraged the independent ratings of protagonists' motivations from Study 2 to determine whether perceptions of particular motives (e.g., protagonists' personal goals, emotion) were associated with observed neural patterns.

Study 1: moral judgments of impartiality, reciprocity and charity

Across all three studies, we used scenarios constructed to target different conceptions of fairness across contexts, i.e., workplace, school, family and recreational situations. In four versions of each of 24 scenarios (96 total stories), a protagonist allocated resources based on (1) reciprocity, (2) impartiality, (3) charity or (4) unspecified criteria (see

Figure 1). Participants were presented with 24 scenarios (six per condition in one of eight counterbalanced orders; see Appendix for full text of scenarios; see Supplementary Material for pre-ratings by an independent sample validating the vignettes). In this study, we investigated participants' judgments of the fairness and moral praiseworthiness of protagonists who allocated according to criteria based on impartiality, reciprocity and charity. Because reciprocity and charity might be construed as forms of favoritism, when compared to impartiality, we aimed to determine whether participants considered impartial allocators to be most fair and moral.

Study 1: method

Participants were 110 individuals on Amazon Mechanical Turk ($M(SD)_{age} = 33.13(11.02)$; 50 female, 60 male) who each read 24 vignettes. After each vignette, participants answered the following questions: "Did [protagonist] act fairly? (7-point scale anchored at 1 = Not at all, 7 = Very much) and "Was [protagonist]'s action morally blameworthy or morally praiseworthy?" (7-point scale anchored at 1 = Completely Morally Blameworthy, 4 = Neither Blameworthy nor Praiseworthy and 7 = Completely Morally Praiseworthy).¹

Study 1: results

Fairness

Fairness ratings differed across allocation types, as shown in Figure 2 ($F(3,327) = 82.82, p < .001$).

Background (A, B)

- (A) Sasha is a manager at a large factory. She is in charge of scheduling shifts for all the managers to complete safety trainings.
(B) Today Sasha has to assign shifts, and she knows afternoon shifts are always preferred to morning shifts.

Method of Allocation (C)

- | | | | |
|---|--|--|---|
| <p>(1) <i>Reciprocity</i>
Sasha thinks about some managers who recently were a great help to her during the planning of the safety training curriculum.</p> | <p>(2) <i>Impartiality</i>
Sasha thinks about which managers had the morning shifts last week, since she trades off shifts week to week.</p> | <p>(3) <i>Charity</i>
Sasha thinks about a couple managers who were struggling to adjust to having newborns at home.</p> | <p>(4) <i>Unspecified</i>
Sasha thinks about the managers and the available shifts. She opens the scheduling document and selects some managers' names.</p> |
|---|--|--|---|

Outcome (D)

- Sasha assigns those managers the better afternoon shifts.

Figure 1. Composition of a sample fairness scenario (see full text of all scenarios in Appendix) with examples of the variations for conditions (1–4). Scenarios were presented as paragraphs for pre-ratings and in Studies 1 and 2; and in four segments (A–D) in Study 3.

¹Participants also completed four items not analyzed here assessing affinity for the protagonist (i.e., "Is [protagonist] someone that you think you would get along with?", "Is [protagonist] someone you would like to be friends with?", "How much do you like [protagonist]?" and perceived similarity to the protagonist ("Do you think you would have made the same decision as [protagonist]?").

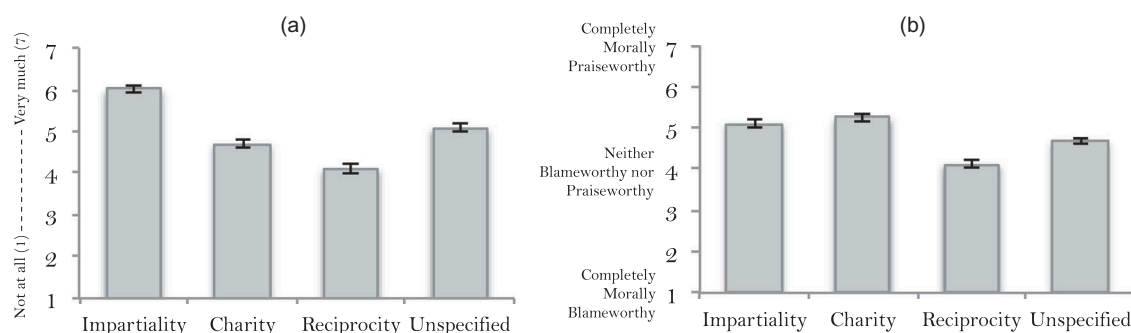


Figure 2. Participants' ratings of the fairness (a) and moral praiseworthiness (b) of the protagonist in Study 1. Error bars indicate SEM.

Bonferroni-corrected pairwise comparisons revealed significant differences between all conditions ($p < .001$; charity and unspecified, $p = .04$). The pattern of results indicated that protagonists in the impartiality scenarios were rated as having acted the most fairly, significantly more so than the protagonists in all other scenarios.

Moral praiseworthiness

Moral praiseworthiness ratings differed across allocation types, as shown in Figure 2 ($F(3,327) = 45.53$, $p < .001$). In Bonferroni-corrected pairwise comparisons, protagonists in the impartiality and charity scenarios were rated as similarly morally praiseworthy ($p = .179$); both impartiality and charity were rated as more morally praiseworthy than reciprocity and unspecified ($p < .001$; reciprocity and unspecified, $p < .001$).

Correlations

Across all allocation scenarios, fairness and moral praiseworthiness were correlated ($r = .65$, $p < .001$). While all significant, correlations of fairness and moral praiseworthiness ranged in strength across scenario types: impartiality ($r = .49$, $p < .001$); charity ($r = .69$, $p < .001$); reciprocity ($r = .82$, $p < .001$); unspecified ($r = .66$, $p < .001$).

Study 1: summary

In Study 1, we found that participants rated impartial allocators as more fair than allocators who used criteria based on reciprocity and charity. Yet reciprocity and charity were not viewed identically: participants considered charitable allocators to be as morally praiseworthy as impartial allocators, and both impartial and

charitable allocators to be more morally praiseworthy than reciprocity-based allocators.² Ratings of fairness and moral praiseworthiness were correlated across allocation types; correlations were stronger for the allocation types rated as less fair, i.e., reciprocity and charity. These results suggest that, moral praiseworthiness and fairness, while correlated overall, are distinct forms of evaluation in this context. To further explore how judgments of different allocation types may rely on mental state representation, in Study 2, we investigated participants' perceptions of the allocators' underlying motivations.

Study 2: perceived motivations of impartiality, reciprocity and charity

Study 2 focused on people's perceptions of the allocators' underlying motivations. We measured the extent to which the protagonists' decisions were perceived as deriving from consideration of (1) the unique states of individuals versus the overall state of the group, (2) their own emotions, (3) personal goals and (4) standard procedures. We aimed to determine whether participants inferred favoritism from reciprocity scenarios: i.e., whether participants would see reciprocity-based allocators as basing their decisions on the unique states of individuals as well as their own personal goals and possibly emotions, rather than on standard procedures. We hypothesized that charity would diverge from reciprocity in that participants would not perceive charity-based allocators as motivated by personal goals. Finally, we hypothesized that impartial allocations would be perceived as basing their decisions on the overall group not the unique states of individuals, and on

²Aligning with results from pre-ratings (see Supplementary Material) showing that participants rated unspecified allocations as reflecting impartiality more than reciprocity and charity, participants also rated unspecified allocations as substantially fair (see subsection "Fairness" under "Study 1: results"). However, although participants seemed to infer that allocations were carried out fairly when criteria were unspecified, they weren't willing to grant as much moral praise in the absence of evidence (see subsection "Moral praiseworthiness" under "Study 1: results").

standard procedures rather than personal goals or emotions. We also measured (e) how hard it was for participants to judge whether the protagonist 'did the right thing' in order to track potential differences in participants' uncertainty in their moral evaluations.

Study 2: method

Participants were 100 individuals on Amazon Mechanical Turk ($M(SD)_{age} = 39.00(13.36)$; 52 female, 48 male) who each read 24 scenarios as in Study 1. After each scenario, participants answered the questions in Table 1.

Study 2: results

Individuals versus group

Ratings of how much the protagonists based their decisions on the *unique states of individuals* versus the *overall group* varied by condition ($F(3,297) = 158.92$, $p < .001$; see Figure 3). As hypothesized, protagonists in the reciprocity and charity scenarios were rated as basing their decisions more on the unique states of individuals and to the same degree ($p > .45$; all comparisons Bonferroni-corrected); protagonists in the impartiality scenarios were rated as basing their decisions more on the overall group, compared to all other conditions ($ps < .001$); and protagonists in the unspecified condition were rated as basing their decisions on "both individuals and the group" (ratings significantly different from all other conditions; $ps < .001$).

Emotion and standard procedures

Ratings of how much the protagonists based their decisions on their own *emotion* ($F(3,297) = 598.11$, $p < .001$) and *standard procedures* ($F(3,297) = 407.05$, $p < .001$) varied by condition. All conditions were

significantly different from one another ($ps < .001$; all comparisons Bonferroni-corrected). Protagonists in the reciprocity and charity conditions were rated as basing their decisions more on *emotion* and less on *standard procedures* compared to impartiality and the unspecified condition (see Figure 3). One-sample t -tests revealed that for *emotion*, reciprocity and charity were higher than the midpoint (4 = "Somewhat"), and impartiality and unspecified were lower than the midpoint ($ps < .001$); for *standard procedures*, reciprocity and charity were lower than the midpoint, and impartiality and unspecified were higher than the midpoint ($ps < .001$).

Personal goals

Ratings of how much the protagonists based their decisions on *personal goals* varied by condition ($F(3,297) = 72.29$, $p < .001$). In the reciprocity condition, protagonists were rated as basing their decisions on *personal goals* more than all other conditions ($ps < .001$; all comparisons Bonferroni-corrected; see Figure 3). Charity and impartiality were indistinguishable ($p = 1$); and the unspecified condition was significantly different from all other conditions ($ps < .002$). One-sample t -tests revealed that perceived personal goal motivation was higher than the midpoint (4 = "Somewhat") only in the reciprocity condition; all other conditions were lower than the midpoint ($p < .001$).

Difficulty in moral evaluation

Ratings of difficulty in moral evaluation (how hard participants thought it was to judge the allocator as "doing the right thing") varied by condition $F(3,297) = 27.49$, $p < .001$; see Figure 3). Participants considered impartial allocations easier to judge as "doing the right thing" compared to all other conditions ($ps < .001$; all comparisons Bonferroni-corrected), which did not differ from each other; ($ps = 1$). Overall, participants did not consider the vignettes difficult to morally evaluate: in one-sample t -tests, all means were lower than the midpoint; $ps < .001$.

Study 2 summary

As in Study 1, impartiality diverged from charity and reciprocity: participants saw it as relatively easy to judge impartiality as "doing the right thing"; participants also rated impartiality as based more on standard procedures, and based less on allocators' emotions, compared to charity and reciprocity. We also found that impartiality and charity diverged from reciprocity in that reciprocity was perceived as more motivated by allocators' own personal goals. Congruently, in Study 1, reciprocity was rated as significantly less fair and

Table 1. Items in Study 2.

Motivations (A–D)	Item
A. Individual versus group ^a	How much was [protagonist]'s decision based on the unique states of individuals versus the overall group?
B. Emotion	How much did [protagonist]'s emotions guide [his/her] decision-making?
C. Personal goals	How much did personal goals guide [his/her] decision-making?
D. Standard procedures	How much did standard procedures guide [his/her] decision-making?
E. Difficulty	How hard is it for you to decide if [protagonist] did the right thing?

^a(A) used a 7-point response scale with the following anchors: 1 = Entirely individuals, 2 = Mostly individuals, 3 = Somewhat more individuals, 4 = Both individuals & the group, 5 = Somewhat more the group, 6 = Mostly the group, 7 = Entirely the group. All other items (B–E) used a 7-point response scale with the anchors: 1 = Not at all, 4 = Somewhat, 7 = Very much.

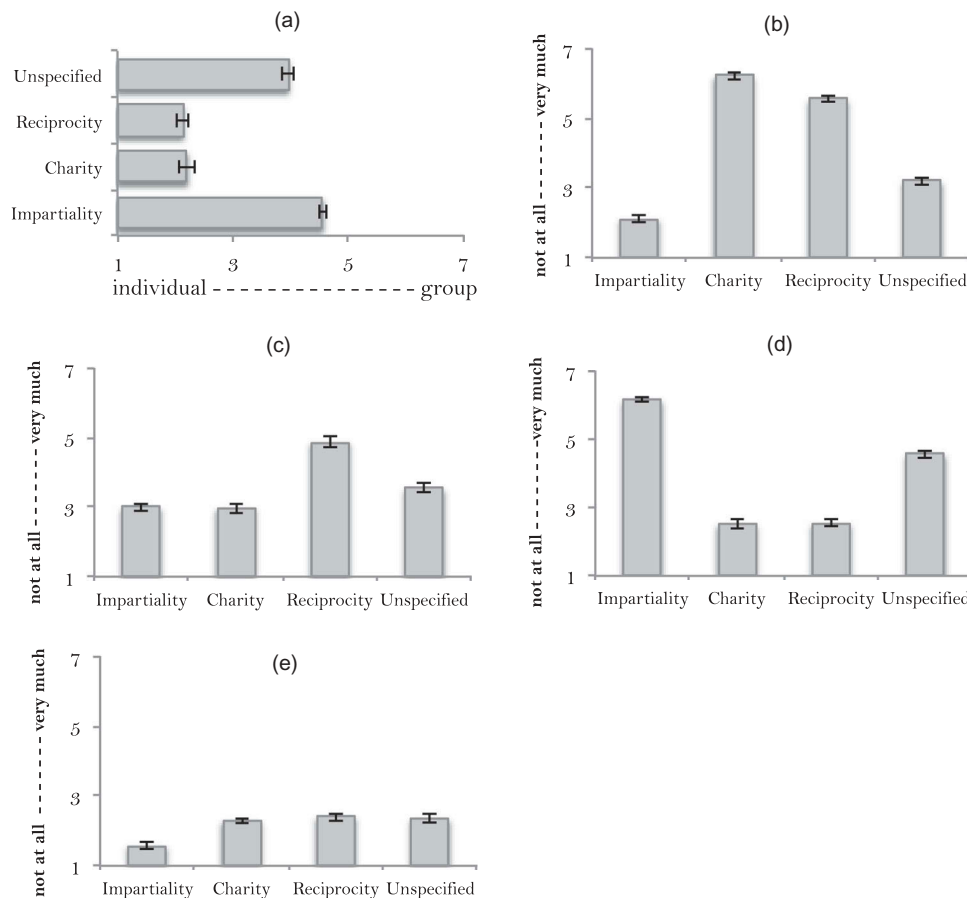


Figure 3. Ratings of scenarios. (a) *Individual versus Group focus*: charity and reciprocity significantly lower than impartiality and unspecified, but not different from each other; (b) *Emotion*: All differences significant; (c) *Personal goals*: impartiality and charity significantly lower than reciprocity and unspecified, but not different from each other. (d) *Standard procedures*: impartiality significantly higher than charity, reciprocity and unspecified; charity and reciprocity not significantly different from each other. (e) *Difficulty of judging as "doing the right thing"*: impartiality was significantly lower than all other conditions, which did not differ from each other.

morally praiseworthy than both charity and impartiality, suggesting that perceived self-serving motives, when detected, drive down both fairness and moral praiseworthiness evaluations.

Study 3: fMRI

Study 3 aimed to distinguish between competing hypotheses: (1) whether morally evaluating reciprocity and charity – both of which may be considered forms of partiality involving individuation of recipients – would elicit more activity in brain regions for ToM than impartiality, or (2) whether morally evaluating reciprocity would diverge from charity and impartiality and elicit more ToM activity, given that reciprocity-based allocators are seen as selfishly motivated. We also considered the possibility that evaluations of impartiality, reciprocity and charity might *all* require robust recruitment of ToM, given prior work revealing a critical role for ToM in

moral cognition (Decety & Cacioppo, 2012; Koster, Hale, Saxe, Dungan, & Young, 2013; Young et al., 2010; see reviews: Young & Dungan, 2007, 2012; Young & Tsoi, 2013), even in the absence of explicit mental state information (Young & Saxe, 2009).

Study 3: method

Participants and procedure

Sixteen naïve right-handed participants (aged 18–27 years, $M(SD)_{age} = 22.44(2.66)$, eight female, eight male) participated for payment. Participants were native English speakers and had normal or corrected-to-normal vision. Participants provided written informed consent in accordance with the internal review board at Boston College.

Participants read 24 scenarios as in Studies 1 and 2 (six per condition: impartiality, reciprocity, charity, unspecified; in one of eight counterbalanced orders). The

scenarios were presented in six runs, each containing four scenarios (one scenario per condition) in a pseudorandom order. At the beginning of each run, participants were instructed to read the scenarios and rate the protagonists.³ Scenarios were presented in four 6-s segments (see segments A–D in Figure 1), for a total presentation time of 24 s per scenario. Following each scenario, participants were presented with the question “How morally praiseworthy?” which remained onscreen for 4 s. Participants entered their responses (ranging from 1 = not at all, to 4 = very) using a button box. Fixation blocks of 14 s were interleaved between the scenarios.

In the same scan session, participants completed a functional localizer task (presented in two runs interleaved with the main experiment runs) that contrasted 10 scenarios involving false beliefs with 10 scenarios involving false photographs to identify brain regions for ToM (see Dodell-Feder, Koster-Hale, Bedny, & Saxe, 2011 ; Saxe & Kanwisher, 2003). Participants completed a series of post-scan survey measures not analyzed here.⁴

Analyses

We conducted whole-brain random-effects analyses and regions of interest (ROIs) analyses targeting regions identified by the ToM localizer (RTPJ, LTPJ, precuneus, DMPFC and VMPFC). ROIs were selected for each participant and were defined as contiguous voxels with a 9-mm radius of the peak voxel that passed the threshold of $p < .001$. The average percent signal change (PSC) relative to baseline ($PSC = 100 \times \text{raw BOLD magnitude for (condition–baseline)}/\text{raw BOLD magnitude for baseline}$) was calculated for each condition at each time point (averaging across all voxels in the ROI and all runs of the same condition) within each of the five ROIs. Analyses were conducted on the outcome segments of the scenarios (identical across all conditions), which began 18 s into each fairness run and lasted for 6 s (see segment D in Figure 1 for example). In addition, we assessed correlations among neural responses, in-scanner moral praiseworthiness ratings and item-level ratings of the perceived motivations behind allocations, provided by the independent participant sample in Study 2.

Imaging protocol

Participants were scanned at the Harvard Center for Brain Science, Cambridge, MA, USA, on a 3.0 T

Siemens Tim Trio fMRI scanner using $36 \times 3 \times 3$ mm near axial slices (0.54 mm gap) covering the whole brain. Standard gradient echo planar imaging (EPI) procedures were used (repetition time [TR] = 2 s, echo time [TE] = 30 ms, flip angle = 90° , field of view [FOV] = 216×216 mm, interleaved acquisition). Anatomical data were collected with T1-weighted multi-echo magnetization-prepared rapid acquisition gradient echo image sequences (TR = 2530 ms; TE = 1.64 ms; FA = 7° ; 1-mm isotropic voxels; 0.5 mm gap between slices; FOV = 256×256). Functional MRI data preprocessing and analyses were performed using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>) and custom software. The data of each participant were corrected for slice timing, realigned to the first EPI, spatially normalized onto a common brain space (Montreal Neurological Institute), spatially smoothed using a Gaussian filter (full-width half-maximum = 8 mm kernel), and high-pass filtered (128 Hz). The experimental task was modeled using a boxcar regressor convolved with a canonical hemodynamic response function. The general linear model included movement parameters as nuisance regressors.

Study 3: results

In-scanner ratings of moral praiseworthiness

Differences in moral praiseworthiness ratings were observed, as shown in Figure 4 ($F(3,45) = 13.85$, $p < .001$). As in Study 1, Bonferroni-corrected pairwise comparisons revealed that charity was rated as more morally praiseworthy than reciprocity and unspecified ($ps < .004$). Charity and impartiality were not significantly different from each other ($p = .14$). Reciprocity and unspecified did not differ ($p = 1$). Impartiality was rated as more morally praiseworthy than unspecified ($p = .03$); the contrast between impartiality and reciprocity did not reach significance ($p = .056$), likely due to insufficient statistical power to detect the effect, compared to Study 1, where participants rated impartiality and charity as both more morally praiseworthy than reciprocity.

Functional localizer results

Whole-brain random-effects analyses revealed greater activation in regions for ToM for scenarios describing mental states over scenarios describing physical states as in prior studies (Dodell-Feder et al., 2011 ; Saxe &

³Instruction wording: “Read a series of scenarios describing a fictional character. Rate how morally praiseworthy the behavior of the character is.”

⁴Post-scan survey measures included the Machiavellian Personality Scale (Dahling, Whitaker, & Levy, 2009), the Interpersonal Reactivity Index (Davis, 1980), a point allocation task (Van Lange et al., 1997) and the Moral Values Questionnaire (Graham et al., 2011).

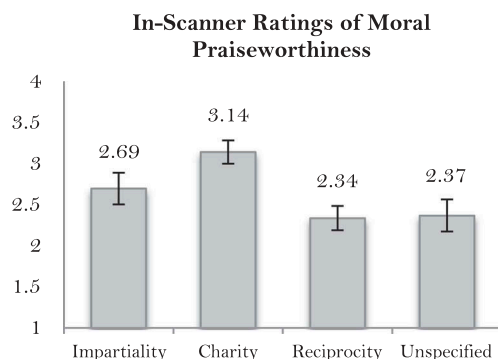


Figure 4. Mean moral praiseworthiness ratings of impartiality and charity (which were not significantly different) were significantly higher than reciprocity and unspecified (which were not significantly different). Error bars indicate SEM.

Kanwisher, 2003). We localized the following regions in participants: RTPJ (16/16 participants), LTPJ (16/16 participants), precuneus (16/16 participants), DMPFC (12/16 participants), VMPFC (12/16 participants) (Table S3).

Whole-brain analyses

We contrasted each type of fairness (reciprocity, charity and impartiality) with unspecified in whole-brain random-effects analyses (threshold $p < .001$ uncorrected, $k > 10$ voxels) (see Table S2 for additional contrasts). No clusters passed threshold for the contrast of impartiality > unspecified; that is, impartiality was undifferentiable from unspecified in neural activity in whole-brain analyses. Reciprocity > unspecified revealed clusters in DMPFC (see Figure 5(a,b)) and left inferior frontal gyrus (IFG; see Table S2). Charity > unspecified revealed clusters in the same location in DMPFC as the contrast of reciprocity > unspecified, and clusters in VMPFC and precuneus (see Figure 5(c,d)); Table S2). Because reciprocity and charity activated similar brain areas, we ran a conjunction analysis of reciprocity > unspecified and charity > unspecified, which revealed an

overlapping cluster in DMPFC (see Figure 5(e); Table S2). Thus, charity and reciprocity, although farthest apart in moral praiseworthiness ratings, recruited an overlapping area in DMPFC, to a similar extent. Reciprocity > charity revealed clusters near the right IFG; charity > impartiality revealed clusters in cingulate gyrus near precuneus, and VMPFC (Table S2). Finally, no clusters passed threshold for contrasts of impartiality > charity or reciprocity, or for reciprocity or charity > impartiality.

ROI analyses

ROI analyses examined differences in neural activation while participants read the outcome of the allocation (identical in the stimuli across conditions) (see Figure 6; see Appendix for full text of scenarios). Activation in precuneus (see means in Figure 6) differed across the four conditions ($F(3,45) = 13.61$, $p < .001$). Bonferroni-corrected planned contrasts revealed that activity in precuneus did not differ between reciprocity and charity ($p = 1$) and that both reciprocity and charity recruited greater activity compared to impartiality and unspecified, which did not differ from each other ($p = 1$); reciprocity versus impartiality: $p < .001$, reciprocity versus unspecified: $p < .001$; charity versus impartiality: $p = .03$, charity versus unspecified: $p = .05$.

Activation in DMPFC (see means in Figure 6) differed across the four conditions ($F(3,33) = 5.96$, $p < .002$). Bonferroni-corrected planned contrasts revealed that activity levels did not differ between reciprocity and charity in DMPFC ($p = .72$), and that reciprocity recruited greater DMPFC activity compared to impartiality ($p = .02$) and unspecified ($p = .01$), which did not differ from each other ($p = 1$). Charity did not significantly differ from impartiality or unspecified ($ps > .6$).

Activation in VMPFC (see means in Figure 6) also differed across the four conditions ($F(3,33) = 5.96$, $p < .04$). Bonferroni-corrected planned contrasts revealed no difference between reciprocity and charity

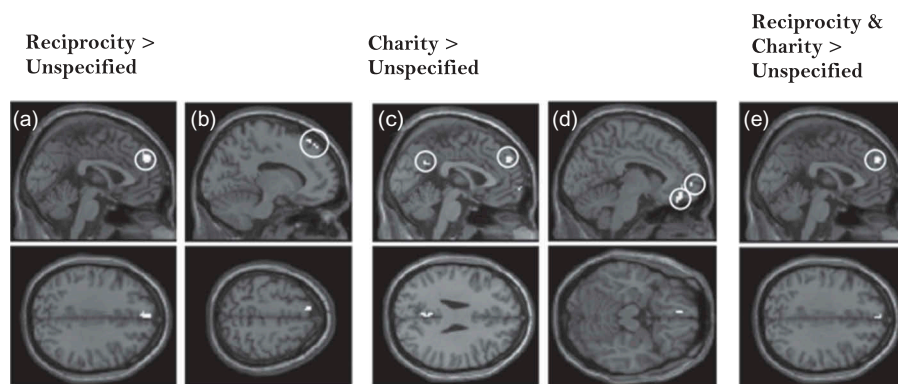


Figure 5. Clusters identified in whole-brain analyses (threshold $p < .001$ uncorrected, $k > 10$ voxels). Reciprocity > Unspecified: (a, b) DMPFC. Charity > Unspecified: (c) precuneus and DMPFC, (d) VMPFC. Reciprocity and Charity > Unspecified: (e) DMPFC.

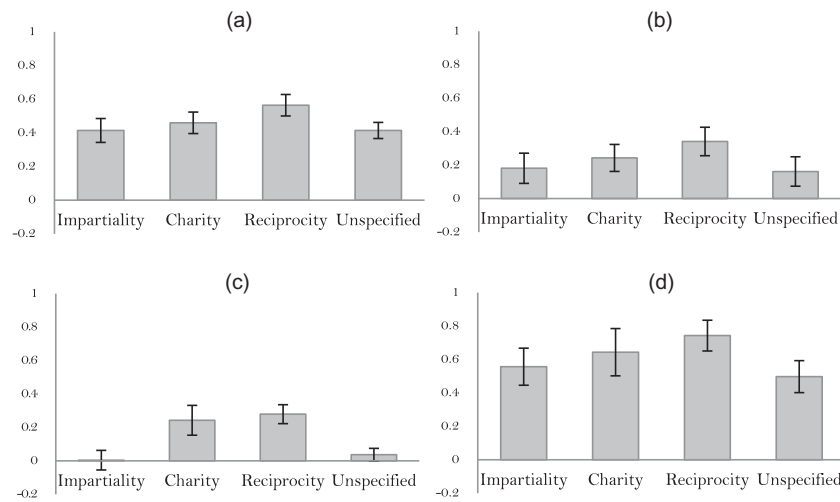


Figure 6. PSC in ROIs in which activation significantly differed across the impartiality, charity, reciprocity and unspecified conditions. (a) LTPJ, (b) DMPFC, (c) precuneus, (d) VMPFC. Errors bars indicate SEM.

($p = 1$), but greater VMPFC activity for reciprocity compared to unspecified ($p = .03$). Impartiality, charity and unspecified did not differ from each other ($ps > .5$).

Activation in LTPJ (see means in Figure 6) differed across the four conditions ($F(3,45) = 4.134$, $p = .011$). Bonferroni-corrected planned contrasts revealed that activity was greater for reciprocity compared to the other conditions (impartiality ($p = .03$) and unspecified ($p = .02$). Impartiality, charity and unspecified did not differ from each other ($ps > .36$). Finally, the difference in activation in RTPJ across the four conditions did not reach significance ($F(3,45) = 2.03$, $p = .12$).

Neural activation, moral praiseworthiness and item-level ratings

Moral praiseworthiness ratings collected in the scanner, as well as the independent ratings of the perceived motivations behind protagonists' decisions from Study 2 were entered into correlational analyses with activity in the ROIs (LTPJ, RTPJ, precuneus, DMPFC, VMPFC), for each condition.⁵ Moral praiseworthiness ratings were uncorrelated with activation within ROIs.

For impartiality, activity levels in precuneus were related to item-level ratings of difficulty of judging the protagonist as having done "the right thing" ($r = .786$, $p < .001$). This correlation suggests that the cognitive difficulty of evaluating impartiality, rather than underlying motivations, predicted precuneus activity. For reciprocity, activity levels in LTPJ were related to item-level ratings of motivation by personal goals ($r = .702$, $p = .002$). That is, the extent to which self-serving motives were detectable predicted the extent to which LTPJ was

recruited during moral evaluation of reciprocity. For charity and unspecified, no correlations survived correction for multiple comparisons. These correlational results should be interpreted with caution, however, given the small sample size and limitations in statistical power.

Study 3 summary

Moral evaluations of reciprocity and charity robustly recruited activity in multiple brain regions for ToM (precuneus, DMPFC, VMPFC, LTPJ); activity in areas did not differ between reciprocity and charity. However, reciprocity elicited greater neural activity during moral evaluations, relative to both impartiality and unspecified, within ToM ROIs (precuneus, DMPFC, LTPJ). Impartiality and the unspecified condition, in turn, consistently did not differ from each other in PSC within ToM ROIs.

General discussion

While prior research documenting universal endorsement of fairness has tended to group different fairness values within a single domain (e.g., equality, reciprocity; Graham et al., 2011), the current work reveals important distinctions within the domain of fairness. In particular, the present work reveals behavioral and neural evidence for the differential reliance on ToM. Recent work has suggested that people may prefer impartial allocators because they see them as having prosocial rather than selfish motives (Shaw, 2013; in press). The current results indicate that moral judgments of impartiality do not necessarily

⁵Significance cutoff for p -value corrected for multiple comparisons = .002.

depend on the processing of complex social cues such as allocators' motives: moral evaluation of impartiality elicited less activity in brain regions for ToM (precuneus, VMPFC, DMPFC, LTPJ) compared to reciprocity and charity in Study 3. Behavioral results from Studies 1 and 2 indicated that impartiality was rated as significantly more fair, easier to judge as "doing the right thing," based more on standard procedures, and based less on allocators' emotions, compared to charity and reciprocity.

Although we found some overlapping neural patterns for reciprocity and charity, people recognized the uniquely self-serving capacity of reciprocity, which set it apart from charity and impartiality, behaviorally and neurally. While cases have been made for the personal advantageousness of charity (Gurven, 2006; Trivers, 1971) and impartiality (Baumard et al., 2013; DeScioli & Kurzban, 2009, 2013; Shaw, 2013), the present participants considered *reciprocity* in particular to be more selfishly motivated and less morally praiseworthy. Reciprocity involved greater activation relative to impartiality and unspecified in a key region for social cognition, the DMPFC. Activity within the DMPFC has been associated with attributions for social behaviors, specifically, for answering "why" (versus "how") questions about human actions, and for calculating blame and praise (Spunt & Adolphs, 2015; Yoder & Decety, 2014). DMPFC may contribute to the encoding of divergent motivations for allocations, which are integrated into moral evaluations.

Similarly, reciprocity also involved greater activation relative to impartiality and unspecified in other key regions for ToM, precuneus and LTPJ. Other work has highlighted the role of ToM in reasoning about immoral actions, including those that result in obtaining an advantage over others (FeldmanHall, Mobbs, & Dalgleish, 2014; Wright, Symmonds, Fleming & Dolan, 2011; Young & Dungan, 2011). Reasoning about agents' mental states may allow people to determine whether agents have violated norms – e.g., by allocating strategically to enhance personal alliances.

The current work has implications for everyday disputes around issues of fairness. Defendants of affirmative action, e.g., stress that implementation of equality-maximizing hiring quotas is fair insofar as such policies help to correct the disadvantages and systematic biases faced by people who are underrepresented in certain fields. That is, need-based allocations help level the playing field to include more players in an impartial system. In practice, proponents of affirmative action might request that policy-makers step into the shoes of individuals who have been underrepresented, determine that their current needs are the result of previous

grievous unfairness and decide that reducing disparity is a desirable goal. Presumably, such emotionally evocative exercises will increase endorsement of affirmative action policies as fair and morally praiseworthy by igniting empathy (Singer, 2007). Our results indicate that such exercises may invoke moral praise from some, but not necessarily broad agreement about fairness. Prototypical fairness, instead, was represented by impartiality, which may be signaled by the extent to which it does *not* draw upon complex social cognition, and the ease with which it can be judged as "doing the right thing". Intriguingly, prior work has shown that "high-accountability" leaders (whose behaviors would be made public) rated empathy-induced preferential allocations to be more unfair than "low-accountability" leaders (Blader & Rothman, 2014). This finding suggests that, at some level, people intuit that charity is not a simple solution to a resource allocation problem and is liable to be perceived as unfair. People may be most likely to endorse allocation systems in which needy individuals will be helped while systematic favoritism (i.e., partiality-based unfairness) will be counteracted.

Future work should investigate another kind of allocation: merit-based allocation. This is often presented as at odds with charity and impartiality, e.g., in debates about appropriate criteria for scholarships, college admissions and pay rates. The current results suggest that the extent to which people believe that merit-based allocation systems are motivated by a strategy aimed at fostering alliances (e.g., maintaining social ties within an elite ingroup), rather than standard procedures and quantifiable evidence (e.g., standardized test scores, hours clocked), may determine the extent to which merit-based allocations are perceived as moral and/or fair. According to the current work, ToM activity may be diagnostic of people's intuitive responses to merit-based allocations. Our finding of increased activation of LTPJ for reciprocity items seen as more motivated by personal goals suggests that increased deployment of ToM may co-occur with moral scrutiny of merit-based allocations. In contrast, if merit-based allocations are viewed as sourced in impartial procedures, then they may elicit less ToM.

Membership in both distally connected social groups and tight dyadic bonds allows for survival and reproduction (Greene, 2013; Nowak, 2006; Trivers, 1971). Reciprocity may be integral for dyadic cohesion: trading off who "treats" and returning help with offers of assistance likely sustain friendships and keep marriages and partnerships stable. Norms promoting need-based allocations may ensure the well-being of children and elders – protecting those who aren't necessarily able to advocate for themselves. Impartiality may be

required to keep the peace in large groups comprised of members with conflicting interests and advocacy motives. While the current work suggests distinct cognitive and neural signatures of distinct forms of fairness, it is for future work to investigate the functions of distinct fairness norms across social and motivational contexts.

Acknowledgments

We gratefully acknowledge the comments, suggestions and assistance of Peter Blake, Paul Bloom, Amelia Brown, Alek Chakroff, Sara Cordes, Fiery Cushman, James Dungan, Michelle Hurst, Katie McAuliffe, Parisa Oviedo, Steven Pinker, Josh Rottman, Jordan Theriault, Lily Tsoi, Patricia Tueme, the Harvard Center for Brain Science and the attendants of the 2016 meeting of the Social and Affective Neuroscience Society.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Alfred P. Sloan Foundation; Dana Foundation;

References

- Axelrod, R., & Hamilton, W. (1981). The evolution of cooperation. *Science*, 211(4489), 1390–1396. doi:10.1126/science.7466396
- Batson, C. D., Lishner, D. A., Carpenter, A., Dulin, L., Harjusola-Webb, S., Stocks, E. L., et al. (2003). "As you would have them do unto you": Does imagining yourself in the other's place stimulate moral action? *Personality and Social Psychology Bulletin*, 29, 1190–1201. doi:10.1177/0146167203254600
- Baumard, N., Andre, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, 36, 59–122. doi:10.1017/S0140525X11002202
- Blader, S., & Rothman, N. (2014). Paving the road to preferential treatment with good intentions: Empathy, accountability and fairness. *Journal of Experimental Social Psychology*, 50, 65–81. doi:10.1016/j.jesp.2013.09.001
- Buckholtz, J. W., & Marois, R. (2012). The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience*, 15(5), 655–661. doi:10.1038/nn.3087
- Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, 12, 314–325. doi:10.1006/nimg.2000.0612
- Dahling, J. J., Whitaker, B. G., & Levy, P. E. (2009). The development and validation of a new Machiavellianism scale. *Journal of Management*, 35, 219–257. doi:10.1177/0149206308318618
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, 10, 85.
- Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature*, 446, 794–796. doi:10.1038/nature05651
- Decety, J., & Cacioppo, S. (2012). The speed of morality: A high-density electrical neuroimaging study. *Journal of Neurophysiology*, 108(11), 3068–3072. doi:10.1152/jn.00473.2012
- DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition*, 112, 281–299. doi:10.1016/j.cognition.2009.05.008
- DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, 139, 477–496. doi:10.1037/a0029065
- Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues*, 31(3), 137–149. doi:10.1111/josi.1975.31.issue-3
- Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2011). fMRI item analysis in a theory of mind task. *NeuroImage*, 55(2), 705–712. doi:10.1016/j.neuroimage.2010.12.040
- Elster, J. (2006). Fairness and norms. *Social Research*, 73(2), 365–376.
- Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in children. *Nature*, 454, 1079–1083. doi:10.1038/nature07155
- FeldmanHall, O., Mobbs, D., & Dalgleish, T. (2014). Deconstructing the brain's moral network: Dissociable functionality between the temporoparietal junction and ventromedial prefrontal cortex. *Scan*, 9, 297–306.
- Fletcher, P., Happe, F., Frith, U., Baker, S., Dolan, R., Frackowiak, R., & Frith, C. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition*, 57(2), 109–128. doi:10.1016/0010-0277(95)00692-R
- Geraci, A., & Surian, L. (2011). The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Developmental Science*, 14(5), 1012–1020. doi:10.1111/desc.2011.14.issue-5
- Gleichgerricht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PloSone*, 8(4), e60418. doi:10.1371/journal.pone.0060418
- Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social Brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience*, 19(11), 1803–1814. doi:10.1162/jocn.2007.19.11.1803
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101, 366–385. doi:10.1037/a0021847
- Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, 143(4), 1600–1615. doi:10.1037/a0036149
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23, 101–124. doi:10.1080/1047840X.2012.651387
- Greene, J. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin Press.

- Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Science*, 6, 517–523. doi:10.1016/S1364-6613(02)02011-9
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105–2108. doi:10.1126/science.1062872
- Gurven, M. (2006). The evolution of contingent cooperation. *Current Anthropology*, 47, 185–192. doi:10.1086/499552
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316(5827), 998–1002. doi:10.1126/science.1137651
- Harris, M. B. (1970). Reciprocity and generosity: Some determinants of sharing in children. *Child Development*, 41, 313–328. doi:10.2307/1127034
- Hill, K., & Kaplan, H. (1993). On why male foragers hunt and share food. *Current Anthropology*, 34(5), 701–710. doi:10.1086/204213
- Jenkins, A. C., & Mitchell, J. P. (2010). Mentalizing under uncertainty: Dissociated neural responses to ambiguous and unambiguous mental state inferences. *Cerebral Cortex*, 20(2), 404–410. doi:10.1093/cercor/bhp109
- Kenward, B., Hellmer, K., Söderström-Winter, L., & Eriksson, M. (2015). Four-year-olds' strategic allocation of resources: Attempts to elicit reciprocation correlate negatively with spontaneous helping. *Cognition*, 136, 1–8. doi:10.1016/j.cognition.2014.11.035
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, 446, 908–911. doi:10.1038/nature05631
- Koster-Hale, J., Saxe, R., Dungan, J., & Young, L. (2013). Decoding moral judgments from neural representations of intentions. *Proceedings of the National Academy of Sciences of the United States of America*, 110(14), 5648–5653. doi:10.1073/pnas.1207992110
- Loewenstein, G. F., Thompson, L., & Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology*, 57(3), 426–441. doi:10.1037/0022-3514.57.3.426
- Meindl, P., Iyer, R., & Graham, J. (2017). A pleasure/power principle of justice: Distributive justice beliefs are guided by societal concerns for pleasure and power. Manuscript submitted for publication.
- Moll, J., De Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and non-moral social judgments. *NeuroImage*, 16, 696–703. doi:10.1006/nimg.2002.1118
- Niemi, L., & Young, L. (2013). Caring across boundaries versus keeping boundaries intact: Links between moral values and interpersonal orientations. *Plos ONE*, 8(12), e81605. doi:10.1371/journal.pone.0081605
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, 314(5805), 1560–1563. doi:10.1126/science.1133755
- Olson, K. R., & Spelke, E. S. (2008). Foundations of cooperation in young children. *Cognition*, 108, 222–231. doi:10.1016/j.cognition.2007.12.003
- Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, 17(8), 413–425. doi:10.1016/j.tics.2013.06.003
- Rasinski, K. A. (1987). What's fair is fair — Or is it? Value differences underlying public views about social justice. *Journal of Personality and Social Psychology*, 51, 201–211.
- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
- Rawls, J. (2001). *Justice as fairness: A restatement*. Cambridge, MA: Harvard University Press.
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, 76(4), 574. doi:10.1037/0022-3514.76.4.574
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, 19(4), 1835–1842. doi:10.1016/S1053-8119(03)00230-1
- Schmidt, M., & Sommerville, J. (2011). Fairness expectations and altruistic sharing in 15-month-old human infants. *Plos ONE*, 6(10), e23223. doi:10.1371/journal.pone.0023223
- Shaw, A. (2013). Beyond "to share or not to share": The impartiality account of fairness. *Current Directions in Psychological Science*, 22, 413. doi:10.1177/0963721413484467
- Shaw, A., & Knobe, J. (2013). Not all mutualism is fair, and not all fairness is mutualistic. *Behavioral and Brain Sciences*, 36, 100–101. doi:10.1017/S0140525X12000878
- Shaw, A., & Olson, K. R. (2012). Children discard a resource to avoid inequity. *Journal of Experimental Psychology: General*, 141, 382–395. doi:10.1037/a0025907
- Singer, T. (2007). The neuronal basis of empathy and fairness. In *Novartis Foundation Symposium "Empathy and Fairness"* (Vol. 278, pp. 20–30).
- Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science*, 23, 196–204. doi:10.1177/0956797611422072
- Sommerville, J. A., Schmidt, M. F. H., Yun, J.-E., & Burns, M. (2013). The development of fairness expectations and prosocial behavior in the second year of life. *Infancy*, 18(1), 40–66. doi:10.1111/j.1532-7078.2012.00129.x
- Spunt, R. P., & Adolphs, R. (2015). Folk explanations of behavior. *Psychological Science*, 26 (6), 724–736.
- Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). 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
- Trivers, R. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46, 35–57. doi:10.1086/406755
- Valdesolo, P., & DeSteno, D. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science*, 18 (8), 689–690. doi:10.1111/j.1467-9280.2007.01961.x
- Valdesolo, P., & DeSteno, D. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology*, 44, 1334–1338. doi:10.1016/j.jesp.2008.03.010
- Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, 77, 337–349. doi:10.1037/0022-3514.77.2.337
- Van Lange, P. A. M., Otten, W., De Bruin, E. M. N., & Joireman, J. A. (1997). Development of prosocial, individualistic, and

- competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology*, 73, 733–746. doi:[10.1037/0022-3514.73.4.733](https://doi.org/10.1037/0022-3514.73.4.733)
- Van Lange, P. A. M., Schippers, M., & Balliet, D. (2011). Who volunteers in psychology experiments? An empirical review of pro-social motivation in volunteering. *Personality and Individual Differences*, 51, 279–284. doi:[10.1016/j.paid.2010.05.038](https://doi.org/10.1016/j.paid.2010.05.038)
- Völlm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., Deakin, J. F., & Elliott, R. (2005). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *Neuroimage*, 1(29), 90–98.
- Waytz, A., Dungan, J., & Young, L. (2013). The whistleblower's dilemma and the fairness–loyalty tradeoff. *Journal of Experimental Social Psychology*, 49(6), 1027–1033. doi:[10.1016/j.jesp.2013.07.002](https://doi.org/10.1016/j.jesp.2013.07.002)
- Wedekind, C., & Milinski, M. (2000). Cooperation through image scoring in humans. *Science*, 288, 850–852. doi:[10.1126/science.288.5467.850](https://doi.org/10.1126/science.288.5467.850)
- Wierzbicka, A. (2006). *English: Meaning and culture*. Oxford: Oxford University Press.
- Wright, N. D., Symmonds, M., Fleming, S. M., & Dolan, R. J. J. (2011). Neural segregation of objective and contextual aspects of fairness. *Journal of Neuroscience*, 31(14), 5244–5252. doi: [10.1523/JNEUROSCI.3138-10.2011](https://doi.org/10.1523/JNEUROSCI.3138-10.2011)
- Yoder, K. J., & Decety, J. (2014). The good, the bad, and the just: Justice sensitivity predicts neural response during moral evaluation of actions performed by others. *The Journal of Neuroscience*, 34(12), 4161–4166. doi:[10.1523/JNEUROSCI.4648-13.2014](https://doi.org/10.1523/JNEUROSCI.4648-13.2014)
- Young, L., Camprodon, J., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporo-parietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 6753–6758.
- Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, 104(20), 8235–8240. doi:[10.1073/pnas.0701408104](https://doi.org/10.1073/pnas.0701408104)
- Young, L., & Dungan, J. (2011). Where in the brain is morality? Everywhere and maybe nowhere. *Social Neuroscience*, 7(1), 1–10. doi:[10.1080/17470919.2011.569146](https://doi.org/10.1080/17470919.2011.569146)
- Young, L., & Dungan, J. (2012). Where in the brain is morality? Everywhere and maybe nowhere. *Social Neuroscience*, 7(1), 1–10. doi:[10.1016/j.neuropsychologia.2009.03.020](https://doi.org/10.1016/j.neuropsychologia.2009.03.020)
- Young, L., & Saxe, R. (2009). An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, 21(7), 1396–1405. doi:[10.2307/1131635](https://doi.org/10.2307/1131635)
- Young, L., & Tsoi, L. (2013). When mental states matter, when they don't, and what that means for morality. *Social and Personality Psychology Compass*, 7(8), 585–604. doi:[10.1111/spc3.v7.8](https://doi.org/10.1111/spc3.v7.8)

Appendix

Scenarios

Participants received one version of each of the 24 scenarios in all studies; each condition (*Reciprocity*, *Impartiality*, *Charity*, *Unspecified*) was represented six times. Parts A–D were presented in a single paragraph in Studies 1 and 2, and in segments in Study 3. Alterations in part D for *Unspecified* condition are indicated in brackets.

SCENARIO 1

- (A) Sasha is a manager at a large factory. She is in charge of scheduling shifts for all the managers to complete safety trainings.
- (B) Today Sasha has to assign shifts, and she knows afternoon shifts are always preferred to morning shifts.
- (C) *Reciprocity*: Sasha thinks about some managers who recently were a great help to her during the planning of the safety training curriculum.
Impartiality: Sasha thinks about which managers had the morning shifts last week, since she trades off shifts week to week.
Charity: Sasha thinks about a couple of managers who were struggling to adjust to having newborns at home.
Unspecified: Sasha thinks about the managers and the available shifts. She opens the scheduling document and selects some managers' names.
- (D) Sasha assigns those managers the better afternoon shifts.

SCENARIO 2

- (A) Dan referees street hockey games in a small city. He is responsible for making calls during the games.
- (B) In one game, Dan is unsure whether a player has just committed a penalty.
- (C) *Reciprocity*: Dan thinks about the player and how he had recently helped Dan clear out a large mess from his basement.
Impartiality: Dan thinks about the technical rules of the game and determines that the player's move was not illegal.
Charity: Dan thinks about the player and how he had just recently returned to hockey after recovering from a major car accident.
Unspecified: Dan thinks about the player's movements in relation to the other player, and the location of the hockey stick.
- (D) Dan decides to not call out this penalty.

SCENARIO 3

- (A) Carol is asked to judge a contest at her company's annual picnic. The winner will receive a large gift certificate to a home improvement store.
- (B) Carol watches her fellow employees participating in the contest.
- (C) *Reciprocity*: One coworker had recently put in a lot of overtime helping Carol complete a project at work.
Impartiality: One coworker met all the judging criteria Carol had been given.
Charity: One coworker was part of a department that had just been downsized and would soon be laid off.
Unspecified: One coworker from the shipping department got on stage and performed an interesting juggling act.
- (D) She decides to give this coworker the highest score.

SCENARIO 4

- (A) Katie is part of a work group at school with two other students. The other students each have a different idea about the best way to present the project.
- (B) They ask Katie to pick which idea she likes best. Katie gives it some thought.
- (C) *Reciprocity*: One of the other students had recently picked Katie's idea for another part of the project.
Impartiality: One of the other students had presented an idea that fit all the requirements laid out in the syllabus.
Charity: One of the other students had just received her first chemotherapy treatment after being diagnosed with lymphoma.
Unspecified: She gets up to take a short break. She buys a soda from the vending machine and returns to the group.
- (D) Katie decides to choose that student's idea [which student's idea to choose].

SCENARIO 5

- (A) Jon is the instructor of an engineering course at a small college. The students of the class have just finished designing an electric car for a competition.
- (B) Jon must decide which student will be chosen to drive the car for the judges at the competition.
- (C) *Reciprocity*: Jon thinks about the students. One student had recently helped Jon raise money for the project at a fundraiser outside of school.
Impartiality: Jon thinks about the rules of the competition. The driver of the car is required to have certain qualifications, and only one student qualifies.
Charity: Jon thinks about the students. One student had recently lost his home to a devastating tornado.
Unspecified: Jon thinks about the students and the dates of the competition. He pulls up the competition schedule on his computer.
- (D) Jon chooses this [a] student to drive the car.

SCENARIO 6

- (A) Craig is part of a team of catering staff who pool all the tips they receive in an evening. All the employees have slightly different jobs and some serve more tables than others.
- (B) Tonight, Craig must distribute the tips among them.
- (C) *Reciprocity*: Craig thinks about a couple team members who had made his night a lot easier by running an extra errand that usually took Craig a long time to do.
Impartiality: Craig thinks about the catering company's rule that tips should be allocated based on how many tables employees served. Some employees served a lot of tables that evening.
Charity: Craig thinks about a couple of employees whose spouses were recently laid off from their jobs and who were on the verge of home foreclosure.
Unspecified: The evening had gone smoothly, and the client had written a check to pay for the catering at the end of the night. Craig thinks about some employees who worked that evening.
- (D) He gives the largest portions of the tip money to them.

SCENARIO 7

- (A) Janice delivers a presentation at an important meeting at work. After she finishes, her coworkers praise the presentation and her achievements.

- (B) Janice thinks about the help she received on the presentation.
- (C) *Reciprocity*: One coworker had recently located an important document that Janice had lost and which was critical for the presentation.
Impartiality: One coworker had coordinated the meetings related to the presentation, and it was standard for the presenter to thank the coordinator.
Charity: One coworker who helped had recently found out that her husband, a pilot in the Air Force, was being deployed for the third time.
Unspecified: She notices a few notes she had added to the end of her presentation. She returns to the podium.
- (D) Janice acknowledges this [a] coworker aloud.

SCENARIO 8

- (A) Jim coaches a soccer team at a local elementary school in a small city. There are 20 children on the team and they have gathered on the field for practice.
- (B) Today, Jim is teaching the kids how to kick a goal.
- (C) *Reciprocity*: Jim thinks about the children on the team. One child had recently come over and tutored Jim's own son in math.
Impartiality: Jim thinks about the soccer league's guideline that all the kids should be able to kick a goal. One child is not able to do it.
Charity: Jim thinks about the children on the team. One child's parents were recently evicted and were now facing homelessness.
Unspecified: Jim demonstrates kicking to the children and they practice. One child is very enthusiastic and won't stop kicking goals, even during the breaks.
- (D) He lets this child have the most kicks.

SCENARIO 9

- (A) Brian is a landscaper who tends several lawns and gardens in the town of Fairfield. One day, he is short on time and can only make it to some of his clients.
- (B) He looks at his client list and tries to figure out where he will go.
- (C) *Reciprocity*: One of his clients had recently given him a valuable new referral for a large landscaping design job.
Impartiality: He checks the contracts for his clients and notices that the contract for one client specifically stipulates daily visits.
Charity: One of his clients recently had surgery and was unable to clear the leaves off his front stairs in order to safely use them.
Unspecified: He notices that several clients are about to receive large shipments of tulip bulbs. He checks the shipping status and looks at his map.
- (D) Brian makes a plan to go to that [a] client.

SCENARIO 10

- (A) Naomi has a popular blog that she uses to discuss issues on children's education. Many people in several districts in her city regularly access the blog for tips.
- (B) Naomi is writing a post on tutoring companies and wants to include links to tutors.
- (C) *Reciprocity*: Naomi thinks about one tutoring company with several different locations, which recently sent

Naomi a packet of information on math learning that she used to write a well-liked post.

Impartiality: Naomi has a rule to include links that are helpful for her readers all over the city. She figures out the locations of tutoring companies in different districts across the city.

Charity: Naomi thinks about her readers in one part of the city who recently lost funding for their in-school tutoring program. She figures out the locations of tutors that would be accessible and affordable for them.

Unspecified: Naomi spends the afternoon making phone calls, checking things online, and writing some notes in a document. She opens the blog program and begins to draft her post.

- (D) Naomi includes these [some] links on her blog.

SCENARIO 11

- (A) Jessica recently graduated from college and moved to a major metropolitan area. Several acquaintances from college are in the city for a conference.
- (B) Jessica has only enough room in her small apartment to provide one person with a place to stay.
- (C) *Reciprocity*: She thinks about her acquaintances. One of them had recently helped Jessica at a different conference by connecting her with a contact who had a job opening.
Impartiality: She thinks about her acquaintances. She emails them all and lets them know the first person to reply can stay at her apartment. She receives a reply that evening.
Charity: She thinks about her acquaintances. One of them had recently been overwhelmed by very expensive student loan payments and was barely able to pay them.
Unspecified: She thinks about her acquaintances. They had all moved far from the college they attended in the middle of the country. Jessica receives a phone call.
- (D) Jessica offers this [an] acquaintance a place to stay.

SCENARIO 12

- (A) Carla is a medical director at a dermatology clinic that employs several physicians. Recently, a new surgical technology had been demonstrated to the group.
- (B) Several of the physicians come to Carla and request time off to obtain training in this new technique.
- (C) *Reciprocity*: One of the physicians had recently covered several of Carla's duties when she took time off to attend a business development seminar in Germany.
Impartiality: It is a clinic policy that training leaves are granted based on the number of years physicians have been employed with the group. Only one had completed the required number of years.
Charity: One of the physicians had recently gone through a messy divorce after surviving a domestic assault, and her finances had been largely wiped out.
Unspecified: Carla consults the website of the company that trains physicians in the new surgical technique. She notes the dates and locations of the training sessions.
- (D) Carla lets this [a] physician take time off for the training.

SCENARIO 13

- (A) Rick manages the purchasing department for an online retailer. Around the holidays, many companies and manufacturers send Rick holiday gifts.

- (B) Today, Rick is deciding what to do with a basket full of chocolates.
- (C) *Reciprocity*: One employee had recently come in on his day off to organize Rick's messy filing system.
Impartiality: He has a rule that a different employee will receive whatever gift was sent to Rick that day. He checks his list for the name of the next employee in line.
Charity: One employee's husband was a police officer who had recently been seriously injured in the line of duty.
Unspecified: He puts the basket on his desk. He checks his calendar and realizes he has a meeting in five minutes.
- (D) Rick gives the chocolates to this [an] employee.

SCENARIO 14

- (A) Al teaches a creative writing course at a bookstore downtown. At the end of each course, he collects each student's favorite poem and prints it in his popular literary journal.
- (B) Today he is laying out the poems. One must be placed on the "feature" page at the beginning.
- (C) *Reciprocity*: One student had recently placed a new ad for Al's creative writing course at the coffee shop she owns.
Impartiality: Al had the students vote for their favorite out of all the students' poems. One poem received the majority of the votes.
Charity: One student had recently returned from Afghanistan, and was dealing with post-traumatic stress disorder.
Unspecified: Al opens his email program and downloads all the files that the students sent to him.
- (D) Al puts this [a] student's poem in the featured spot.

SCENARIO 15

- (A) Gary is the driver of a large tour bus that takes hundreds of passengers to a casino. The huge crowd tends to get impatient as they wait to board the bus.
- (B) Gary is about to begin the boarding process for the passengers.
- (C) *Reciprocity*: Gary recognizes several passengers from a previous trip. They had won big payouts at the blackjack table at the casino, and had given Gary a large tip.
Impartiality: Passengers assigned to the seats in the back are to board first. Gary consults his passenger list and identifies those seated in the back rows.
Charity: Gary recognizes several passengers from a previous trip. They were a group of breast cancer survivors who had undergone difficult treatments.
Unspecified: Gary locates the passenger list for the trip. He checks the date and time that it was last updated. He determines how many empty seats would be available on the bus.
- (D) Gary lets these passengers board the bus first [some passengers board the bus].

SCENARIO 16

- (A) Brenda works at a farm stand. There is a large selection of plants, fruits and vegetables for sale.
- (B) Today the farm stand is busy with many customers, and Brenda is ringing up sales at the counter.
- (C) *Reciprocity*: Brenda recognizes a couple buying fruit and several potted plants. They had recently given her a discount at the hardware store that they own.
Impartiality: A couple is purchasing some fruit and several potted plants. They hand Brenda a coupon they had cut out of the paper for a discount.

Charity: Brenda recognizes a couple buying fruit and several potted plants. They had recently lost their son, who was away at college, in a terrible car accident.

Unspecified: A couple places several potted plants and a few pounds of fruits and vegetables on the counter. Brenda weighs the fruits and vegetables.

- (D) Brenda gives them the three plants for the price of one.

SCENARIO 17

- (A) Jackson is a music teacher at an elementary school. The students are learning about percussion, and one instrument, the snare drum, is always the class favorite.
- (B) Today Jackson is preparing his lesson plan for music class and must assign the instruments.
- (C) *Reciprocity*: One student had stayed after class last week to help Jackson put away all the instruments in the storage closet, a task that usually takes him a long time to do.
Impartiality: Jackson trades off who gets the snare drum each class. He looks at the class roster and determines who was next in line for it.
Charity: One student was going through very difficult times at home and was barely able to stay engaged in the classroom.
Unspecified: Jackson pulls down the boxes of percussion instruments from the storage closets and makes sure all the pieces are there. He consults the class list.
- (D) Jackson assigns the snare drum to this [a] student in his lesson plan.

SCENARIO 18

- (A) Mario is a cardiologist who runs a clinic for patients who are at risk of having a heart attack. A famous nutritionist is visiting Mario's clinic and offers to give two patients personalized diet advice.
- (B) Mario must select which patients will receive the free consultations.
- (C) *Reciprocity*: Two patients had recently helped Mario find an excellent assisted living facility for his mother.
Impartiality: The diet requires patients to have certain cholesterol levels and two patients fit the requirements.
Charity: Two patients had recently moved to the area because they were displaced from their home after flooding.
Unspecified: Mario notes the dates that the nutritionist is available to give the consultations on his calendar. He opens the patient records program.
- (D) Mario invites these [some] patients to come in for the nutritionist's consultations.

SCENARIO 19

- (A) Max is a photographer who is expanding his business to include weddings. In order to build a wedding portfolio, he invites couples to submit their names to be considered for free wedding photography.
- (B) Today he needs to choose who will receive the offer.
- (C) *Reciprocity*: One couple had recently recommended Max to a new hotel who needed photography for their website.
Impartiality: Max enters all the couples' names into a computer program and has the program randomly select one couple.
Charity: The father of the bride-to-be in one couple had recently been killed in a sudden workplace accident.

Unspecified: Max begins to sort through the emails of the couples who contacted him. He notes their names and contact information.

- (D) Max offers this [a] couple the free wedding photography services.

SCENARIO 20

- (A) Michelle is an intern at the White House. Several students from the high school she attended in her hometown are part of mentorship program for young people interested in a career in politics.

- (B) Michelle is deciding which of these students to bring to an event at the White House.

- (C) *Reciprocity:* One of the students had volunteered for several years in a row at an annual fundraiser that Michelle coordinated.

Impartiality: The program has a policy that seniors should be given priority to attend White House events. One student in the program is a senior.

Charity: One of the students in the program had been raised in foster care since the age of 12 after losing both of his parents in a shooting.

Unspecified: Michelle locates the phone number of the director of the mentorship program and gives her a call.

- (D) Michelle decides to bring this student to the event [which student to bring to the event].

SCENARIO 21

- (A) Anne is the director of an emergency response team in a small town in New England. After a major storm, the town is in need of clean drinking water.

- (B) A major shipment of water arrives in town and Anne must decide how to distribute it.

- (C) *Reciprocity:* The company who sent the shipment has a large warehouse located in the center of Anne's town. A group of their employees was working to repair that warehouse.

Impartiality: Anne consults the list of emergency zones given to her by the response team. The team has organized the list in order to facilitate the fastest distribution. One zone is at the top of the list.

Charity: Some residents of the town had their homes completely destroyed in the storm. They were all displaced to a shelter without electricity or running water in one area of the town.

Unspecified: Anne helps the emergency response team unload the massive shipment into smaller delivery trucks and boats. She selects an emergency zone from her list.

- (D) Anne sends the first portion of the shipment out to them.

SCENARIO 22

- (A) Sara collects book donations for a nonprofit organization in a major city. Thousands of children throughout the city receive free books through the program.

- (B) Sara has received a huge donation of brand new books from a popular publisher, which she will send out today.

- (C) *Reciprocity:* Sara thinks of a school in one borough that often sends teachers to help out with the organization's fundraisers.

Impartiality: Sara opens the master list of the schools that receive book donations. She notes which school is next in line for a donation.

Charity: Sara thinks of a school in one borough that was very poorly funded and was recently labeled "underperforming" by the state.

Unspecified: Sara sorts the books by level of reading difficulty. She repackages the books, labels them, and prints the addresses of the recipients.

- (D) Sara ships the donation out to them.

SCENARIO 23

- (A) Tania is an executive assistant for a design company in Los Angeles. One afternoon, she is asked to make several deliveries in the city.

- (B) Tania is not sure she'll be able to get all the deliveries out in time. She begins to plan her route.

- (C) *Reciprocity:* One delivery was for a designer in the lighting department who recently gave Tania an enthusiastic recommendation for a raise.

Impartiality: Tania plots out what appears to be the most direct route to all the delivery locations on her map. One location is just next door.

Charity: One delivery was for a designer whose son was in the hospital and who was clearly very stressed out at work.

Unspecified: Tania picks up the outgoing delivery packages and makes sure she can fit them all in her car. She types in an address on her GPS.

- (D) Tania makes this delivery first.

SCENARIO 24

- (A) Bill is a fisherman who catches lobsters in the town of Rockport. When he gets back to shore one afternoon, there is a mixup with some orders.

- (B) Bill notices there is an extra lot of six lobsters.

- (C) *Reciprocity:* The owner of the business next door to the dock had recently given Bill his extra set of tickets to a baseball game.

Impartiality: Whenever there are extra lobsters, the fishermen give them to the manager who redistributes them to another order.

Charity: A family that lived close by was struggling through tough financial circumstances and Bill often saw the father returning from the food bank.

Unspecified: Bill looks over the order forms closely, checks the orders he already packed and speaks to his coworker.

- (D) Bill packs up the lobsters and brings them over to him.