Does Empathy Promote Helping by Activating Altruistic Motivation or Concern About Social Evaluation? A Direct Replication of Fultz et al. (1986)

Thomas G. McCauley

University of California, San Diego

William H. B. McAuliffe

Rethink Priorities

Michael E. McCullough

University of California, San Diego

Author note: Corresponding author: Michael McCullough, University of California, San Diego. E-mail: memccullough@ucsd.edu. Supporting materials and preregistration documents for this project can be accessed at https://osf.io/2egsd/.

We gratefully acknowledge the John Templeton Foundation for their support of this work through Grant #61539 (New Measures for the Experimental Study of Autonomous Prosocial Motivation) to Michael McCullough. We also appreciate Daniel Batson's helpful feedback on a previous version of this manuscript.

Draft version 1, 1/4/22. This paper has not been peer reviewed.

Word count: 7,722

Abstract

Empathy motivates people to help needy others. Does it do so by activating genuine concern, or by activating more self-interested goals that helping needy others might enable them to fulfill? The empathy-altruism hypothesis claims that empathic concern reflects a non-instrumental desire to improve the welfare of a person in need. To rule out the alternative hypothesis that empathy motivates prosocial behavior by first generating fear of appearing selfish, Fultz et al. (1986) manipulated empathy for a needy target using perspective-taking instructions; they also manipulated whether the subject's opportunity to help was subject to social evaluation. However, Fultz et al.'s (1986) experiments were underpowered. Here, we conducted a large-*N* pre-registered replication of Experiment 2 in Fultz et al. (1986). We also administered self-report measures of moral identity and endorsement of the principle of care to test whether these traits reflect altruistic desires or desires to avoid disapprobation. We found that volunteering did not differ between the high and low social evaluation conditions, and that volunteering was not significantly higher in the high-empathy condition. These results sit uneasily with Fultz et al. (1986)'s evidence in support of the empathy-altruism hypothesis. We also failed to find evidence that the principle of care or moral identity internalization reflect altruistic motivation. Consistent with the empathy-altruism hypothesis, however, we did find that self-reported empathic concern predicted helping.

Keywords: Morality, empathy, replication

Introduction

Like many social animals, humans frequently help relatives and cooperative partners. There are strong evolutionary arguments for why such help would in some situations be underwritten by altruistic motivation—that is, a genuine desire to improve the beneficiary's welfare (Trivers, 1971). However, empirically establishing the existence of altruistic motivation is difficult because, in any given case of helping, it is always possible to posit an egoistic motive—that is, a desire to help others as a means of improving one's own welfare (Batson, 2011). For example, a father may take good care of his children because he values their welfare, or instead because he fears that if he did not people would think that he is a terrible dad. Although cases of heroic self-sacrifice (e.g., Colby & Damon, 1992; Oliner & Oliner, 1988) do not appear to be explained by selfish motives, systematic evidence that ordinary people tend to help others under certain circumstances, even in the absence of selfish incentives to do so, is necessary to establish that altruistic motivation is a species-typical trait that explains a wide range of helping behavior. Here, we attempt to directly replicate an experiment by Fultz, Batson, Fortenbach, McCarthy, and Varney (1986) which represents a key piece of evidence for altruistic motivation.

Historical Background

In the early 1980s, Daniel Batson and his colleagues commenced a series of experiments in search of altruistic motivation. Based on earlier findings that empathy (defined in this literature as feelings of sympathetic concern for needy or distressed others, not as experiencing the same emotion as others) appears to precede acts of helping (e.g., Krebs, 1975), Batson proceeded to test the *empathy-altruism hypothesis*, which claims that empathy for a suffering person reliably elicits altruistic motivation (Batson et al., 1981; Batson, 2010). Batson's research on empathy has been highly influential within social psychology and allied fields. Nearly all prominent theories of the

origins of empathy and prosocial motivation are either premised on the truth of the empathy-altruism hypothesis or focused on accommodating it (e.g., De Waal, 2008; Goetz et al. 2010; Hoffman, 2000; Preston, 2013; Sznycer et al., 2019; Zaki, 2020).

Although other laboratories have contributed independent evidence in favor of the empathy-altruism hypothesis (e.g., FeldmanHall et al., 2015; McAuliffe et al., 2018), the wide consensus in its favor owes to Batson's research program. He and his colleagues conducted experiments that pitted the predictions of the empathy-altruism hypothesis to an alternative hypothesis that claimed either that there is no causal link between empathy and helping or that empathy generates an egoistic motive to help (for a review, see Batson, 2011). In a typical experiment, subjects first encounter a person in a state of need who could benefit from the subject's help (e.g., volunteering to spend time with a lonely classmate or to help raise money for a financially strapped university student). Next, subjects undergo an experimental manipulation of empathy—the most common of which involves reading perspective-taking instructions that ask participants to either (a) imagine what the needy person might be feeling (the high-empathy condition) or (b) remain objective as they consider the needy person's plight (the low-empathy condition). They simultaneously undergo an experimental manipulation of an opportunity for subjects to fulfill an egoistic motivation by helping the needy person (Batson, 2011). These egoistic incentives have included extrinsic rewards for helping (Batson et al., 1989) and exiting the helping situation altogether, thereby avoiding vicarious negative affect (Batson et al., 1981). If empathically aroused subjects persisted in helping even in the absence of an egoistic incentive, then one is licensed to infer that empathy could be activating an altruistic goal. On the other hand, if empathy selectively increases helping in the presence of a given incentive, then one can infer that empathy arouses egoistic motivation.

One of the first egoistic alternatives to the empathy-altruism hypothesis to be explicitly articulated was based on the recognition that a reputation for callous indifference can result in social exclusion (Archer et al., 1981). Thus, potential helpers have an incentive to help needy victims when others will know if they had such an opportunity, even if they place no intrinsic value on the welfare of victims. Archer et al. (1981) asserted that empathy is in fact sensitive to the threat of negative social evaluation. According to this "empathy-specific evaluation hypothesis," appearances of concern are either insincere (often called "crocodile tears") or the product of self-deception about one's own sentiments.

Fultz, Batson, Fortenbach, McCarthy, and Varney (1986) pitted the empathy-altruism hypothesis against the empathy-specific evaluation hypothesis by simultaneously manipulating empathy (using perspective-taking instructions) and the anonymity of the helping opportunity. Subjects in the high social evaluation condition were led to believe that the experimenter knew that the subject's partner was in need (i.e., was lonely and looking to make a friend), that the subject's partner knew the subject had the opportunity to help (by volunteering for a successive experiment that would have allowed the subject and the partner to get to know each other), and that the experimenter knew subjects' self-reported willingness to volunteer on behalf of the partner. In contrast, subjects in the low social evaluation condition were led to believe that the experimenter was unaware of the partner's need, that the partner did not know the subject was in a position to help, and that the experimenter was not privy to subjects' willingness to volunteer. Fultz et al. (1986) found that subjects who were assigned to the imagine-other condition volunteered to spend more time with the lonely stranger than did subjects who were assigned to the remain-objective condition, regardless of whether the subject believed that the other subject and the experimenter were in a position to learn about the subject's choices (and potentially disapprove of them). This

pattern of findings implied to Fultz and colleagues that empathy-induced helping is underwritten not by a motivation to avoid social pressure, but instead by a motivation to help the needy other.

The Present Study

Fultz et al (1986) has been cited more than 327 times on Google Scholar as of November 2021 even though its main experiment (Experiment 2) comprised only 32 subjects who were spread across four different cells in a 2 x 2 between-subjects design. Consequently, the experiment was underpowered for the range of effect sizes one ordinarily expects to find in laboratory social psychology, raising the possibility that its findings are spurious (Loken & Gelman, 2017). The low power is especially concerning because ruling out the empathy-specific evaluation hypothesis depends on *not* observing an interaction between empathy and social evaluation: Very large samples are required to detect interactions (Aguinis, 1995). There is therefore a risk of having sufficient power to detect main effects but not interactions, which would bias the results in favor of the empathy-altruism hypothesis.

In the present paper, we conducted a replication of Fultz et al.'s (1986) experiment 2 with a larger sample size. The empathy-altruism hypothesis leads to our primary prediction:

Prediction 1: Based on the empathy-altruism hypothesis, one would predict that subjects who receive imagine-other instructions will volunteer to spend more time with the needy subject than subjects who receive remain-objective instructions, regardless of whether they were assigned to the high or low social evaluation condition. This prediction stands in contrast to the prediction one would formulate based on the empathy-specific evaluation hypothesis: On this latter view, one would predict that subjects who receive imagine-other instructions will volunteer to spend more time with the needy subject only if they are under high social evaluation.

In addition to arousing empathic concern, the opportunity to help a needy other might also arouse self-oriented personal distress, so that subjects might help to relieve their own distress, rather than to improve the welfare of the needy person. For instance, past research has found that imagine-other instructions increase empathic concern but not personal distress (Batson, Fultz, and Schoenrade, 1987; McAuliffe et al., 2018). Additionally, Fultz et al. (1986) found that self-reported empathic concern, but not personal distress, predicted volunteering behavior regardless of whether subjects were in the high or low social evaluation condition. Based on the empathy-altruism hypothesis, we therefore make a second prediction:

Prediction 2: Self-reported empathic concern will predict subjects' decisions to volunteer regardless of whether they were assigned to the high or low social evaluation condition, even after controlling for self-reported personal distress.

Individual differences in prosocial motivation

Holding aside whether empathy promotes volunteering on its own, scholars have proposed traits that might influence whether people respond to the opportunity to help in situations that arouse empathic concern and social evaluation pressure. However, self-reports of prosocial motivation are often contaminated by social desirability bias, obscuring the true relationship between individual differences in prosocial motivation and helping behavior (Graziano & Tobin, 2002). For this reason, it is useful to make sure that self-report measures that are presumed to reflect an unselfish desire to help strangers do indeed predict whether people help strangers in the absence of potential social evaluation. Therefore, we tested three preregistered predictions

regarding three traits that have been hypothesized to promote different types of motivations for helping.

First, we examined the endorsement of the principle of care (the internalized moral value that one ought to help people in need; Wilhelm & Bekkers, 2010). Past research has revealed that people who endorse the principle of care (Bekkers & Ottoni-Wilhelm, 2016) donate more to charity, but it is unknown whether this effect is driven by principlism—i.e., an intrinsic desire to abide by an endorsed moral principle (Batson, 2011)—or by a desire to appear principled. On the assumption that principle of care endorsement is sincere, we made the following prediction:

Prediction 3: Endorsement of the principle of care will predict volunteering more strongly for subjects in the low social evaluation condition than those in the high social evaluation condition.

Next, we examined whether moral identity internalization (the centrality of moral traits to one's self-concept) and moral identity symbolization (the desire to have one's moral traits observed in public settings; Aquino and Reed, 2002), predict helping. Internalization is hypothesized to promote helping when social and material rewards for helping are absent, while symbolization is thought to promote helping to secure such rewards. Those who profess to have a strong internalized moral identity engage in more prosocial behavior (Hertz & Krettenhauer, 2016), but studies of this association have not manipulated the presence or absence of rewards for helping. To address these gaps in the literature, we included two additional predictions:

Prediction 4: Self-reported moral identity internalization will predict volunteering more strongly for subjects in the low social evaluation condition than in the high social evaluation condition.

Prediction 5: Self-reported moral identity symbolization will predict volunteering more strongly for subjects in the high social evaluation condition than in the low social evaluation condition.

Method

Pre-registration and Institutional Approval

Our hypotheses and analysis plan were preregistered, and all study materials, data, and analysis code can be found at https://osf.io/fqu36. We submitted an amendment to our original preregistration because, in the time since we submitted the original preregistration, we learned that there are more appropriate methods to analyze the data (viz., fitting data to models that can account for a zero-inflated outcome; controlling for experimenter effects and suspicion), which are reflected in the results reported here. We report the results of the original preregistration in the supplemental materials as well, but the results of the original and amended analyses were not qualitatively different. We report the results of non-preregistered Bayesian analyses, calculated using the BFpack package in R (Mulder et al., 2019), to evaluate the strength of the evidence for our hypotheses. All study materials and procedures were approved by the Institutional Review Boards at the University of Miami and University of California, San Diego.

Participants

We intended to collect data from either N = 400 total subjects, or when we had collected data from N = 75 non-suspicious participants in each cell, whichever came first. However, as stated in the amendment to our preregistration, we were forced to terminate data collection due to the Covid-19 pandemic. The final number of subjects included in analyses was N = 280 (M_{age} = 19.94, SD_{age} = 1.96; *females* = 197). Data were collected from N = 129 subjects at the University

of Miami, and $N = 151$ subjects at the University of California, San Diego. Three subjects failed to provide a coherent response to the volunteering opportunity (e.g., one participant endorsed volunteering for 2, 3, 4, 5, and 6 hours) and were removed from analyses, leaving a final sample of $N = 277$ subjects included in the data analysis.

Procedure

We made minor adaptations of the protocol to ensure that it would have experimental realism in our laboratory (e.g., see below about participants recycling their partner's note), but otherwise we duplicated Fultz et al.'s (1986) protocol as faithfully as possible. See supplemental materials for more details about differences between Fultz et al.'s (1986) protocol and the protocol used in our experiment.

Pre-experiment interaction. After arriving at the lab, subjects were greeted by a research assistant and told that they would be participating in an "impression formation" experiment to see how people form impressions of others based on verbal communication alone and in the absence of non-verbal or visual cues. The subjects were then led to believe (erroneously) that they had been partnered with another subject of the same sex who had recently completed the study (in reality, there was no other person). In the high social evaluation condition, the research assistant asked subjects if they knew their partner, who was named Mark Tycho for male subjects and Janet Tycho for female subjects (there was no one named Mark or Janet Tycho at the institutions where the study was conducted, and no subject indicated that they knew Mark or Janet). After subjects responded negatively, the research assistant told them that their partners had affirmed that they also did not know them. In the low social evaluation condition, the research assistant did not ask whether subjects knew their partner. Thus, anonymity was maintained in the low social evaluation condition, but not in the high social evaluation condition.

Next, subjects received written introductory instructions stating that either they or their partner would be assigned to be the "Communicator," who would then write as note to the other subject, designated the "Listener," who would then read the note. Subjects were always assigned to the Listener role, and were told that their interaction partner had already completed the tasks in his/her role as Communicator.

Perspective-taking manipulation. Subjects then received the Communicator's note, along with "mindset" instructions that they were asked to read before reading the note. Subjects in the high-empathy condition received imagine-other instructions, and were asked to focus on imagining the feelings of the interaction partner while reading the note. Subjects in the low-empathy condition, in contrast, received remain-objective instructions, and were asked to focus on the facts described in the note. See the supplemental materials for the text of both mindset instructions.

Need situation. Subjects then read the note. The note indicated that the Communicator and recently transferred to the University of Miami/UC San Diego from a university in their hometown. The note goes on to express how lonely the communicator has felt since transferring. See the supplemental materials for the text of the note.

Emotion Response Questionnaire (ERQ). After reading the note, subjects completed a questionnaire to indicate their emotional reactions to reading the note, including their endorsement of five adjectives that reflect empathic concern ("compassionate", "moved", "softhearted", "sympathetic", and "tender") and five adjectives that reflect personal distress ("alarmed", "distressed", "disturbed", "troubled", and "upset"), along with 20 other distractor adjectives unrelated to empathy or distress. (See the supplemental materials for all items from the ERQ.) For each adjective, subjects endorsed how much they felt each adjective described their emotional state on a Likert scale, with response options ranging from 1 (*not at all*) to 7 (*extremely*). We formed a

unit-weighted composite for both the empathic concern (M = 4.86, SD = 1.19, *McDonald's* ω = 0.82) and personal distress items (M = 2.59, SD = 1.16, *McDonald's* ω = 0.82).

Impression Questionnaire. Subjects also completed a questionnaire indicating their impressions of their interaction partner (See the supplemental materials for all items from the Impressions Questionnaire.). We did not analyze data from the Impression Questionnaire, as it was administered only to facilitate the social evaluation manipulation and enhance the plausibility of the ruse.

Volunteering opportunity. The experimenter delivered to the subject a letter from the professor in charge of the experiment. The letter contained an opportunity to volunteer for a second, ostensibly unrelated study about long-term relationships. Subjects were informed that if they agreed to volunteer, then they would be paired with the communicator for an initial meeting, and the experimenter would occasionally ask the subject to report on the development of the friendship. The subject was asked to indicate on a form whether they were willing to volunteer, and, if so, for how many hours.

In the high social evaluation condition, the letter indicated that both the subject and communicator would be made aware of the study opportunity, and that the subject and communicator would be recruited for the long-term relationship study only if both people agreed to participate. The subject was then led to believe that the research assistant had read the Communicator's note, and was comparing the subject's impressions of the Communicator against the content of the note. Thus, in this condition the subject had reason to fear that they could appear callous for failing to volunteer (given that they know about the partner's loneliness, and the partner and experimenter will know that the subject knows).

In the low social evaluation condition, the letter indicated that the communicator would be contacted about participating in the long-term relationships study only if the subject indicated a desire to participate, and the subject was told to place the partner's note back in its envelope and recycle it. Thus, participants believed that they, but not the experimenter, would be aware of the partner's loneliness. Consequently, in the low social evaluation condition the subject presumably realized that if they do not help at all, then the partner cannot blame them for failing to act on an opportunity to help. Similarly, the experimenter will not construe participation in the follow-up study as a means to relieving the partner's loneliness.

Social evaluation manipulation check. Subjects completed a six-item questionnaire that probed their experience of having been asked to volunteer for the long-term relationships study. Four items were distractors, but two items assayed whether the subject believed that the experimenter and needy person knew that the participant had had an opportunity to relieve the needy person's need: "My interaction partner would know if I chose to not volunteer" and "My response to the volunteer questionnaire was anonymous." Subjects indicated how much they agreed with each item on a self-report Likert scale, with response options ranging from 1 (*definitely not*) to 7 (*definitely yes*).

Principle of Care Scale. Subjects' self-reported endorsement of the principle of care was measured using the eight-item Principle of Care Scale (Bekkers & Ottoni-Wilhelm, 2016). We formed a unit-weighted composite from responses to the eight items (M = 3.38, SD = 0.33, *McDonald's* ω = 0.83).

Moral Identity Scale. Subjects' self-reported moral identity was measured using the Moral Identity Scale (Aquino & Reed, 2002), which consisted of ten items that map onto two subscales: moral identity internalization and moral identity symbolization. Subjects who participated in the

study at the University of Miami provided their responses during pre-testing data collection, while subjects at UCSD did so at the end of the study. We formed unit-weighted composites for both subscales, with five items mapping onto the internalization subscale (M = 3.26, SD = 0.34, *McDonald's* ω = 0.76), and five items mapping onto the symbolization subscale (M = 3.26, SD = 0.75, *McDonald's* ω = 0.84).

Suspicion probe. Finally, before being dismissed, we probed subjects for suspicion and debriefed them with a full explanation of how and why we had deceived them. For the suspicion probe, we used a funnel questionnaire, which is a three-question verbal interview in which subjects were asked increasingly specific questions about the study to determine whether they had become aware of either the deception or the experiment's true hypothesis (Aronson et al., 1990).

Analysis plan

Analyses were preregistered on the Open Science Framework (https://osf.io/fqu36).

Controlling for experimenter effects. Eleven undergraduate research assistants (RAs) proctored experimental sessions, with four research assistants proctoring sessions at the University of Miami, and seven proctoring session at the University of California, San Diego. To control for the effects of test sites and research assistants, we included 10 dummy variable codes (*session proctor* = 1, *did not proctor* = 0) for the 11 RAs, as coding for research assistants also effectively controls for the effect of test site.

Controlling for suspicion. We coded for suspicion based on subjects' responses to the suspicion probe. Subjects were categorized as suspicious if, during the funnel questionnaire, they indicated that the communicator was not real; that the communicator received different sets of instructions (or was told what to say or do); or that the experiment was designed to evaluate their willingness to volunteer or their empathic responses to the communicator's plight. We also

distinguished between whether subjects expressed conviction in their suspicion, which we term "hard" suspicion, or whether they merely expressed uncertainty or doubt (which we call "soft suspicion;" see supplemental materials for details about the coding procedure).

Model selection. Since the number of hours subjects volunteered to help the partner is a discrete variable and cannot take on negative values, fitting our data to an ordinary least squares model would be inappropriate. Instead, we compared four competing models to determine which model was most appropriate for our data: a negative binomial model, a zero-inflated negative binomial model, a zero-inflated Poisson model, and a multinomial ordered logit model[1]. Each model included the number of hours volunteered as the dependent variable, and the following predictor variables: 10 dummy-coded covariates to control for the research assistant who ran the session; 2 dummy-coded covariates to control for hard and soft suspicion; 1 predictor for the empathy manipulation; 1 predictor for the social evaluation manipulation; and 1 predictor for the interaction of the empathy manipulation with the social evaluation manipulation.

Results

We report the results from our modified pre-registration. Results from the original pre-registration are included in supplemental materials.

Manipulation check

The effect of the perspective-taking manipulation on empathic concern. Subjects in the imagine-other condition (M = 5.04, SD = 1.18) scored higher on the Emotion Response Questionnaire than subjects in the remain-objective condition (M = 4.68, SD = 1.17; $t(275) = 2.59$, $p = .01$, *Cohen's d* = 0.31, *95%* [0.07, 0.55]), indicating the perspective-taking instructions increased subjects' empathic concern for the needy target.

[1]Our modified pre-registration indicated that we would also fit a zero-truncated Poisson model, but we were unable to fit the model due to a surplus of zeroes in the data, so it was excluded from model testing.

Social evaluation manipulation. Subjects in the high social evaluation condition (M = 5.09, SD = 2.08) indicated greater agreement with the statement “My interaction partner would know if I chose to not volunteer” than did subjects in the low social evaluation condition (M = 2.14, SD = 1.88; $t(278) = 12.44$, $p < .001$, *Cohen's d* = 1.49, *95%* [1.20, 1.78]).). However, the subjects in the high (M = 4.04, SD = 2.66) and low (M = 4.18, SD = 2.67) social evaluation conditions did not differ in their agreement with the statement "My response to the volunteer questionnaire was anonymous” ($t(278) = -0.44$, p =.658, *Cohen's d* = -0.05, *95%* [-0.29, 0.18]). Thus, although the social evaluation manipulation was successful in leading subjects in the low social evaluation condition to believe their partner would not know if they volunteered, it was unsuccessful in leading them to believe that their decision to volunteer was anonymous, perhaps because subjects believed that the experimenter would view their responses at some point after the experiment.

Results

Model selection

To select the model that was most appropriate for modelling the underlying distribution of subjects’ volunteering hours, we calculated Akaike Information Criteria (AIC) fit indices for the negative binomial, zero-inflated negative binomial, zero-inflated Poisson, and multinomial ordered logit models; we then selected the model with the lowest AIC. Next, we compared the evidence for each model by calculating Akaike weights, which can be compared using the Kullback–Leibler information quantity $I(f, g)$ to measure the discrepancy between the true model, relative to an approximation of the true model (Wagenmakers & Farrell, 2004). Results of the model fit analyses are shown in Table S1. The negative binomial model had the best fit, and a comparison of the Akaike weights indicated that the negative binomial model is 1.05 times more likely to be the true

model, relative to the zero-inflated negative binomial model[2]. Therefore, we retained the negative binomial model for the remainder of confirmatory and exploratory analyses.

Results for the perspective-taking model

First, we tested whether the perspective-taking instructions promoted volunteering. Results are reported in Table 1, and descriptives are reported in Table 2. Neither the perspective-taking manipulation, nor the social evaluation manipulation, nor the interaction of those two manipulations influenced subjects' volunteering hours (*p*s > .699; see Table S2 for the coefficients for the suspicion and experimenter covariates). Thus, we found no evidence that the perspective-taking instructions or social evaluation manipulation affected subjects' decisions about how much time they offered to volunteer. We also did not find that perspective-taking influenced subjects' willingness to volunteer in the low social evaluation condition differently than in the high social evaluation condition.

In conducting a Bayesian analysis, one seeks to evaluate which of two hypotheses is better supported by a pattern of empirical results. The strength of support for a focal hypothesis, versus its alternative, is expressed as a Bayes Factor (Kass & Raftery, 1995). Bayes factors are centered around 1, and can range from zero to infinity. We organized our Bayesian analyses so that as values increase from 1, they indicate increasingly strong support for the experimental hypothesis over the null hypothesis; as values decrease from 1, they indicate increasingly strong support for the null hypothesis.

By Kass and Raftery's (1995, p. 777) rules of thumb, Bayes factors ranging from 1 to 3.2 indicate only anecdotal support for the focal hypothesis (as they put it, "not worth more than a bare mention"); values ranging from 3.2 to 10 indicate "substantial" support for the focal hypothesis;

[2]Although we were able to compute fit indices for the zero-inflated negative binomial model, the solution was computationally singular, so we were unable to compute standard errors or *p*-values for parameter estimates.

values ranging from 10 to 100 indicate “strong support” for the focal hypothesis; and values exceeding 100 indicate “decisive” support for the focal hypothesis. The Bayes factors for the three effects described here ranged from 0.876 to 1.861, which imply that the data provide no meaningful support for our hypotheses (Kass & Raftery, 1995). Bayes factors for these hypothesis tests (and all the others) are reported in Tables 1, 3, 4, 5, and 6.

Table 1

Coefficients for the Negative Model That Includes Predictors for the Perspective-taking Manipulation, the Social Evaluation Manipulation, and the Interaction Between Perspective-taking and the Social Evaluation Manipulation.

Predictors	*b* (SE)	*P*	*BF*
Perspective-taking manipulation	0.09 (0.23)	.699	1.861
Social evaluation manipulation	-0.00 (0.23)	.987	0.975
Empathy manipulation x Social evaluation manipulation	-0.03 (0.32)	.934	0.876

Note: The *b*s reported for the negative binomial model are logged coefficients. BF is the Bayes factor associated with each coefficient, compared to the hypothesis that the coefficient is either zero or negative.

Table 2

Descriptive Statistics of Volunteering Decisions by Condition.

	Low perspective-taking	High perspective-taking
Low social evaluation	Mean = 1.99, SD = 2.74 *N* = 70 Volunteered = 44.3%	Mean = 2.16, SD = 3.02 *N* = 69 Volunteered = 42%

High social evaluation	Mean = 2.19, SD = 2.32 *N* = 70 Volunteered = 27.1%	Mean = 2.25, SD = 2.71 *N* = 71 Volunteered = 36.8%

Notes. Volunteered = The percentage of participants in a given condition that volunteered for at least one hour.

Results for the individual differences models

Next, in four separate models, we tested our preregistered predictions about self-reported empathic concern and individual differences in prosocial motivations. For each model, we removed the dummy-coded predictor variable reflecting the perspective-taking manipulation with either (a) empathic concern (including personal distress as a covariate), (b) endorsement of the principle of care, (c) moral identity internalization, or (d) moral identity symbolization. Each model also featured a predictor term for the social evaluation manipulation and the interaction of the focal predictor and the social evaluation manipulation; dummy coded terms for hard and soft suspicion, respectively; and ten dummy coded terms for the experimenter.

Results for all models are shown in Tables 3-6 (see Tables S3-S6 for coefficients for the suspicion and experimenter covariates). For the model in which self-report empathic concern was the focal predictor, empathic concern positively predicted volunteering behavior (b = 0.31, SE = 0.10, *95%* [0.11, 0.51], p = .002) even after controlling for personal distress (p = .673). The Bayes factor for this association (BF = 810.18) indicates that our data provide "decisive" support for the hypothesis that empathic concern is associated with volunteering (Kass & Raftery, 1995, p. 777). For the model in which the moral identity symbolization composite was the focal predictor, the

social evaluation manipulation appeared to reduce volunteering (*b* = -1.54, *SE* = 0.74, *95%* [-3.01, -0.09], *p* = .036), and symbolization's association with volunteering behavior was dependent on the social evaluation manipulation (*b* = 0.09, *SE* = 0.04, *95%* [0.04, 0.91], *p* = .032). When we probed the interaction between moral identity symbolization and social evaluation using simple slopes analyses, we found (consistent with our hypothesis) that symbolization was positively associated with hours volunteered for subjects in the high social evaluation condition (*b* = 0.35, *SE* = 0.15, *95%* [0.05, 0.65], *p* = .02), but not for subjects in the low social evaluation condition (*b* = -0.12, *SE* = 0.16, *95%* [-0.44, 0.20], *p* = .46; see Figure 1). The Bayes factor for this coefficient (*BF* = 62.43) implied "strong support" for the hypothesis that moral identity symbolization is associated with volunteering among subjects in the high social evaluation condition (Kass & Raftery, 1995, p. 777).

Table 3

Coefficients for the Model That Predicts Volunteering Hours as a Function of Self-reported Empathic Concern, the Social Evaluation Manipulation, the Interaction of Empathic Concern and the Social Evaluation Manipulation, and Self-reported Personal Distress.

Predictors	*b* (SE)	*p*	*BF*
Empathic concern	0.31 (0.10)	.002**	810.18
Social evaluation manipulation	0.67 (0.73)	.360	4.55
Empathic concern x Social evaluation manipulation	-0.14 (0.14)	.344	0.21
Personal distress	-0.03 (0.07)	.673	0.51

Note: The *b*s for all models reported here are logged coefficients. BF is the Bayes factor associated with each coefficient, compared to the hypothesis that the coefficient is either zero or negative.

Table 4

Coefficients for the Model That Predicts Volunteering Hours as a Function of Self-reported Endorsement of the Principle of Care, the Social Evaluation Manipulation, and the Interaction of the Principle of Care and the Social Evaluation Manipulation.

Predictors	*b* (SE)	*p*	*BF*
Principle of care	0.04 (0.04)	.435	0.62
Social evaluation manipulation	0.51 (1.73)	.769	1.60
Principle of care x Social evaluation manipulation	-0.02 (0.06)	.769	1.60

Note: The *b*s for all models reported here are logged coefficients. For the principle of care and the social evaluation manipulation, BF is the Bayes factor associated with each coefficient, compared to the hypothesis that the coefficient is either zero or negative; for the interaction effect, the Bayes factor is compared to the hypothesis that the coefficient is either zero or positive.

Table 5

Coefficients for the Model That Predicts Volunteering Hours as a Function of Self-reported Endorsement of Moral Identity Internalization, the Social Evaluation Manipulation, and the Interaction of Moral Identity Internalization and the Social Evaluation Manipulation.

Predictors	*b* (SE)	*P*	*BF*
Moral identity internalization	0.05 (0.08)	.499	0.48
Social evaluation manipulation	0.75 (1.59)	.640	2.13
Moral identity internalization x Social evaluation manipulation	-0.04 (0.10)	.650	2.08

Note: The *b*s for all models reported here are logged coefficients. For moral identity internalization and the social evaluation manipulation, BF is the Bayes factor associated with each coefficient, compared to the hypothesis that the coefficient is either zero or negative; for the interaction effect, the Bayes factor is compared to the hypothesis that the coefficient is either zero or positive.

Table 6

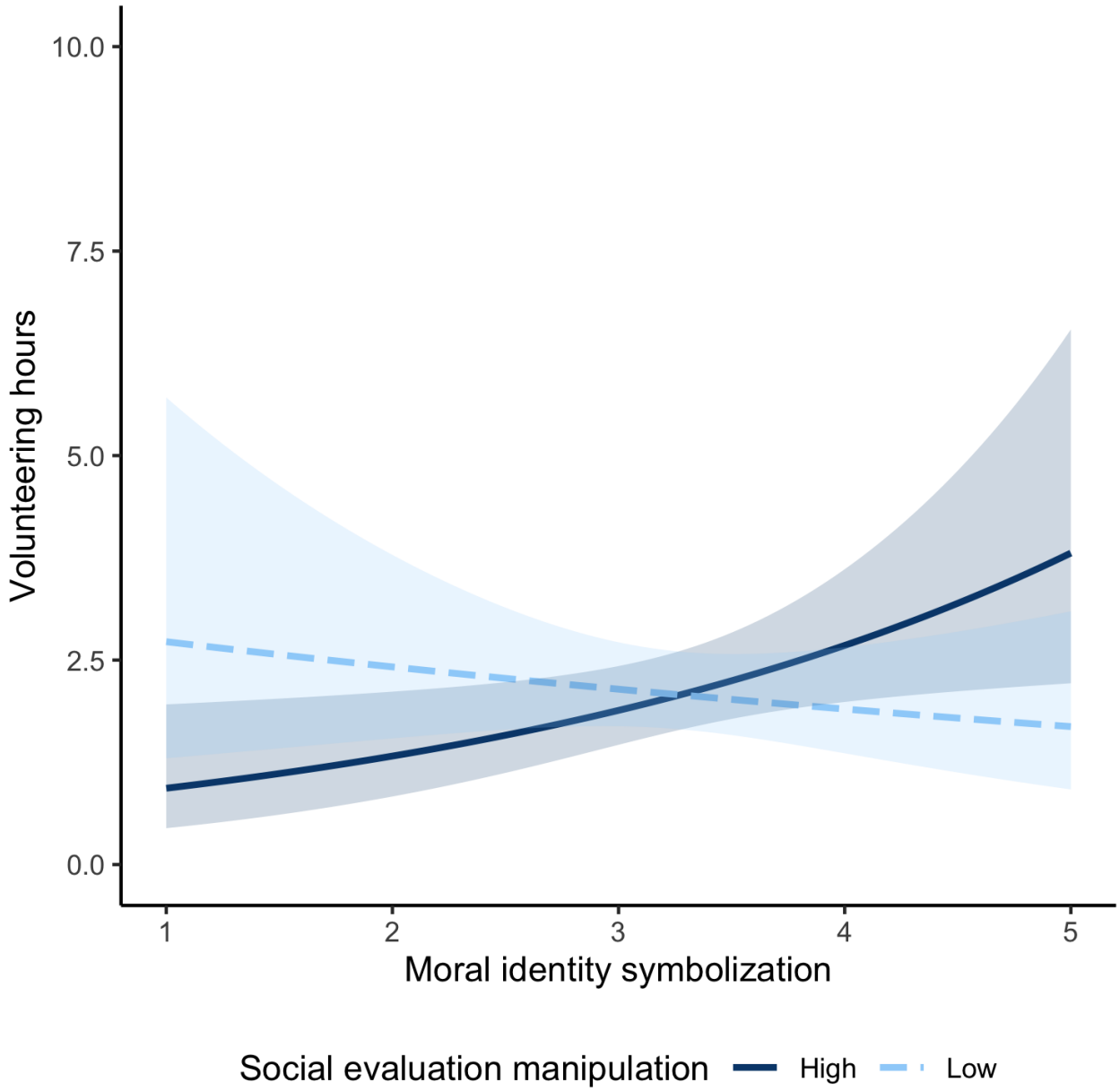
Coefficients for the Model That Predicts Volunteering Hours as a Function of Self-reported Endorsement of Moral Identity Symbolization, the Social Evaluation Manipulation, and the Interaction of Moral Identity Symbolization and the Social Evaluation Manipulation.

Predictors	*b* (SE)	*P*	*BF*
Moral identity symbolization	-0.02 (0.03)	.456	0.30
Social evaluation manipulation	-1.54 (0.74)	.036*	0.18
Moral identity symbolization x Social evaluation manipulation	0.09 (0.04)	.032*	62.43

Note: The *b*s for all models reported here are logged coefficients. BF is the Bayes factor associated with each coefficient, compared to the hypothesis that the coefficient is either zero or negative.

Figure 1

Two-way interaction of moral identity symbolization and social evaluation. The dashed orange line shows the association of volunteering and moral identity symbolization in the high social evaluation condition; the solid blue line shows the same association in the low social evaluation condition. The shaded area indicates the 95% confidence region for the effects of the high and low social evaluation conditions, respectively.



Discussion

Daniel Batson and his colleagues' research has persuaded many behavioral scientists that empathy is a source of altruistic motivation (Batson, 2011), and several experiments designed to test the empathy-altruism hypothesis have produced data inconsistent with the hypothesis that empathically motivated helping is driven by fear of social punishment (e.g., Batson, Dyck et al.,1988, Experiments 2-4; Batson & Weeks, 1996). Here, we conducted a preregistered replication of Fultz, Batson, Fortenbach, McCarthy, and Varney's (1986, Study 2) landmark experiment, as Fultz et al.'s (1986) study revealed that subjects who were instructed to imagine the plight of a needy person, as opposed to subjects who were instructed to remain objective, were more likely to help the needy person, even after another plausible egoistic motive for helping had been removed from the helping encounter. We also used this opportunity to advance a secondary goal of testing whether traits that have been theorized to measure principlism and egoism are more likely to promote helping in the absence or presence of selfish incentives, respectively.

Contrary to Fultz et al.'s (1986) findings, we failed to replicate the effect of the empathy manipulation and the social evaluation manipulation on the number of hours that subjects agreed to volunteer. Because the empathy-altruism hypothesis predicts a main effect of empathy on helping, our failure to produce that effect here is a cause for reflection. We also did not find any causal evidence in favor of the empathy-specific evaluation hypothesis, which entails the prediction that imagining the plight of the other would motivate subjects to help only when they believed that they would be negatively evaluated if they failed to help. Consistent with the empathy-altruism hypothesis, however, self-reported empathic concern did significantly predict the number of hours that subjects volunteered.

We also found that moral identity symbolization predicted volunteering, and that this association was qualified by an interaction with the social evaluation manipulation: Moral identity symbolization predicted volunteering in the high evaluation condition, but not the low evaluation condition. This was as predicted, since a desire to appear moral to others should promote helping only in the presence of potential social evaluation. In contrast, we predicted that moral identity internalization and endorsement of the principle of care would be especially predictive of volunteering to help in the low social evaluation condition. However, neither moral identity internalization nor the principle of care significantly predicted volunteering in either condition.

Ineffective Manipulations?

If perspective-taking instructions are effective in producing group differences in empathic concern, then why did they not promote volunteering here? Recent research revealed that subjects who read imagine-other instructions wrote more emotionally supportive notes to another anonymous subject who was experiencing a stressful life event (McAuliffe et al., 2018), and had stronger desires to punish transgressors on behalf of victims (Pfattheicher et al., 2019, Studies 2a-2b). However, these are relatively costless forms of prosociality, raising the possibility that perspective-taking instructions' impact on empathy is too weak to have a measurable effect on more costly behaviors, like offering one's valuable free time (Inzlicht & Hutcherson, 2017; Neuberg et al., 1997). Batson (1997) preferred perspective-taking instructions over other empathy manipulations because he believed they were less likely than other empathy manipulations to simultaneously activate egoistic motives, but the time may be ripe for researchers to develop stronger, but equally pure manipulations of empathy.

The prospect of being judged harshly for failing to help also did not affect helping. It is possible that subjects in the high social evaluation condition were unconcerned about being judged

for choosing to not volunteer, or did not believe that observers would view their volunteering decision as morally tinged. However, it could also be that subjects were indeed concerned with how their decisions could be judged, but we simply failed to manipulate assurance of anonymity. Indeed, although one of our manipulation checks revealed that subjects in the high social evaluation more strongly believed that their communication partner would know if they chose not to volunteer, our second manipulation check indicated that subjects in the high social evaluation condition were not more likely to believe that their volunteering decision was anonymous than subjects in the low social evaluation condition. The failure of the second manipulation check may be due to subjects' accurate recognition, even in the low social evaluation condition, that *someone* would indeed see their volunteering decision: After all, the researcher would need to see the subject's decision in order to arrange for the subject to participate in the follow-up study. If subjects were actually reflecting on this fact when answering the question about anonymity—but not reflecting on whether researchers would know whether they had failed to help a needy other—then the failure of the social-evaluation manipulation to affect subjects' scores on this question is neither surprising nor especially probative as a test of whether subjects behaved as they did out of a fear of negative social evaluation: Subjects could have acknowledged that their decision was not completely anonymous while still feeling assured that nobody would judge them for callousness.

Other alternative explanations are tenable as well. For instance, subjects in the low social evaluation condition might have failed to draw the inference that neither the experimenter nor the communicator were in a position to negatively evaluate the subject's decision. That is, subjects in the low social evaluation condition might have neglected to realize that no one person had all of the information necessary to condemn them for callousness. Additionally, it could be that subjects were more worried about appearing unwilling to aid the advancement of scientific understanding

than they were about appearing unconcerned about a fellow student's social isolation. If so, then the realization that someone would view their rejection of participation in the follow-up study would be sufficient to induce concern with social evaluation in all conditions.

Whether one wishes to accept these post hoc interpretations or not, it remains the case that our social-evaluation manipulation failed to change subjects' beliefs about whether their volunteering decision would be made anonymous. While it might be tempting to interpret this result as evidence that our experiment was based on a false premise—that we weren't in a position to pit the empathy-altruism hypothesis against the empathy-specific evaluation hypothesis in the first place because we didn't manipulate evaluation concerns perfectly—failures to replicate manipulation checks should cause every bit as much concern about the replicability of previously published findings as failures to replicate major conclusions should. (For an enlightening example of a scholarly interchange on this issue, see Cheung et al., 2016, and Finkel, 2016). Our failure to manipulate subjects' perceptions that their behavior would remain anonymous, therefore, should not alleviate concerns about our failure to replicate Fultz et al.'s (1986) main result: Instead, in light of the large effect size that Fultz and colleagues (1986) obtained on this manipulation check (d = 0.94; computed from Fultz et al.'s published results), and our large sample size—which jointly should have given us ample statistical power for detecting a true effect (*estimated power* = 100%)—it should heighten them.

Experimental Expertise?

Some authors seek to attribute direct replication attempts that yield non-significant results to failures to adapt the protocol to novel circumstances or the inexperience of the replicators (e.g., Cunningham & Baumeister, 2016; Fabrigar & Wegener, 2016; Schwarz & Clore, 2016), so perhaps we lacked the ability or knowledge to recreate the experimental protocol in ways that

maximized the impact of the experimental manipulations. Such a speculation would not be unfounded because even though we did find that our perspective-taking manipulation significantly influenced empathic concern (d = 0.31), the effect was smaller in magnitude than the effect that Fultz et al.'s (1986) obtained (d = 1.07; computed from Fultz et al.'s published results) and that McAuliffe et al. (2020) discovered in their comprehensive meta-analytic review of all known experiments (*Hedge's g* = 0.68, 95% CI: .61, .76). However, it is noteworthy that Fultz et al.'s (1986) effect size of d = .1.07 was .39 units above the mean effect size of d = .68 that McAuliffe et al. (2020) reported: Our effect size estimate of d = .31 was nearly as far away (.37 units), albeit in the opposite direction. Thus, while one might justifiably wonder why our perspective-taking manipulation was so weak, one would be equally justified in wondering why Fultz et al.'s manipulation was so strong.

Admittedly, we did not consult any of the authors from Fultz et al. (1986) when developing the protocol for our laboratory, who might have detected some problematic aspect of our protocol. However, including input from the original authors has not been found to systematically affect the results of replication attempts (Klein et al., 2020; Schlingloff et al., 2020). Moreover, we have experience in conducting experiments on the empathy-altruism hypothesis that have successfully manipulated both empathy and fear of social evaluation (e.g., McAuliffe, 2017; McAuliffe et al., 2018). Our background in the area does not rule out the possibility that we made important methodological errors, of course, but does at least suggest that the conditions necessary to produce the results observed by Fultz et al. (1986) are not easily understood, even by experimenters with some experience in the area.

Individual differences in prosocial motivation

Subjects with high scores on the moral symbolization scale were more likely to volunteer, but only if they had been assigned to the high social evaluation condition. This finding is consistent with research suggesting that moral symbolization reflects concerns about how one's moral character might be perceived by others. For example, other work has shown that self-reported moral symbolization predicts public, but not private, prosocial behavior (Mayer et al., 2013).

Why did the moral identity internalization and principle of care endorsement measures we examined here fail to predict volunteering? Perhaps these constructs do not actually measure factors that cause helping behavior. Most of the studies that have documented an association of endorsement of the principle of care with a measure of helping behavior have relied on self-reports of helping behavior from survey studies (de Wit & Bekkers, 2016, 2020; Wilhelm & Bekkers, 2010), or measures of one's prosocial intentions; only a few studies have included direct observations of helping behavior (Wilhelm & Bekkers, 2016). Likewise, the meta-analytic association of moral identity internalization with observable prosocial behavior is modest (Hertz & Krettenhauer, 2016), and no previous study to our knowledge has examined the association of moral identity symbolization with observable helping behavior. To be sure, some self-report measures of prosocial traits do predict prosocial behavior—and not just self-reports of prosocial behavior (Thielmann, Spadaro, & Balliet, 2020)—but moral identity internalization and endorsement of the principle of care may simply be weaker predictors of helping behavior than others are. Alternatively, it is possible that subjects viewed volunteering to befriend the communication partner as a nice thing to do, but not necessarily a moral thing to do, in which case people's internalized moral identities or espousals of the principle of care might have been

irrelevant to how they perceived and responded to their opportunity to interact with their communication partner.

Statistical Considerations

When we began this project, we believed the main improvement we would make over Fultz et al. (1986) was increasing sample size: Small samples have been identified as a central cause of low replicability in experimental social psychology (Fanelli, Costas, & Ioannidis, 2017). However, we also identified three potential shortcomings in Fultz et al.'s (1986) analysis strategy. First, they evaluated the effect of the experimental manipulations using analysis of variance, in effect treating volunteering as a continuous outcome. Treating counts as though they are continuous can substantially bias regression coefficients (O'Hara & Kotze, 2010), especially when the dependent variable is zero-inflated as was the case here. We used a negative binomial regression model, thereby accommodating the actual distribution of volunteering intentions. Second, Fultz et al. (1986) excluded from analysis participants who expressed suspicion during the funnel debrief. Insofar as suspicion is non-random (e.g., perhaps participants lower in dispositional trust are more likely to report suspicion), excluding suspicious participants reduces the generalizability of the results. Even more seriously, excluding suspicious participants could interfere with the tendency for randomization to balance conditions on unobserved covariates, if one condition is more likely to elicit suspicion than another (Anderson et al., 2021). We addressed selection bias by retaining all suspicious participants. This "intent-to-treat" approach risks underestimating the effect of the experimental manipulation by including participants who were not immersed in the study. To balance these concerns, we also controlled for suspicion, thereby partialing out indirect paths between condition assignment and volunteering that are mediated by suspicion. Finally, Fultz et al. (1986) used three experimenters but did not

statistically account for experimenter effects, potentially leading to artificially small standard errors. We used a fixed-effects approach to partial out the systematic effects of individual experimenters (McNeish & Stapleton, 2016). Fultz et al.'s (1986) analytic decisions reflected standard practice in experimental social psychology at the time and may not have biased their results. However, we believe that analytic issues deserve as much attention as small sample size when attempting to replicate past findings, given ever-evolving standards for data analysis.

Conclusion

Our results sit uneasily next to Fultz et al.'s (1986) finding that manipulated empathy increases helping behavior, irrespective of the presence of an incentive to help to avoid appearing callous. In the work we report here, neither the manipulation of empathy nor the manipulation of fear of social evaluation significantly impacted volunteering. Until further work reveals otherwise, we believe Fultz et al. (1986)'s major conclusion—that empathically aroused helping does not reflect a fear of social evaluation for behaving callously towards a person in need, but instead, reflects a genuine desire to raise the welfare of a needy other—should be regarded as an unsettled issue. However, our results do not contradict the empathy-altruism hypothesis entirely: Individual differences in empathy were positively associated with individual differences in volunteering, and there was *not* a significant interaction between the empathy and social evaluation manipulations, a key prediction made by the empathy-specific evaluation hypothesis. Fultz et al. (1986) is not unique in the empathy-altruism literature for featuring a relatively small sample and never having been subject to direct replication. Going forward, social psychologists might therefore consider attempting to directly replicate other widely cited experiments that rule out alternatives to psychology's main source of experimental evidence that empathy is a reliable source of altruistic motivation.

Context Statement

People who experience empathic emotion for a needy person often help them in response, but what are empathic helpers actually trying to accomplish? One possibility, stated formally as the *empathy-altruism hypothesis*, is that empathy promotes genuine concern for the needy person's welfare. Another possibility, however, stated formally as the *empathy-specific evaluation hypothesis*, is that empathy makes people worry about whether others will think less of them if they fail to help. As part of our ongoing research into the motivational foundations of altruism, we conducted this experiment in efforts to replicate key experimental findings that are often invoked in support of the empathy-altruism hypothesis. We found correlational evidence that people who experienced empathy for the needy person were more likely to help them, but we did not find that an experimental manipulation designed to increase empathy—or a manipulation designed to increase concern about social evaluation—affected how much people helped. Thus, these results fail to furnish strong causal support for either the empathy-altruism hypothesis or the empathy-specific evaluation hypothesis. Society needs more prosocial behavior, so we hope researchers will continue to develop new tests of the hypotheses we evaluated here, along with others.

References

Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management*, *21*(6), 1141-1158.

Anderson, K., Zamarro, G., Steele, J., & Miller, T. (2021). Comparing performance of methods to deal with differential attrition in randomized experimental evaluations. *Evaluation Review*, *45*(1-2), 70-104.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, *100*(401), 464-477.

Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423-1440.

Archer, R. (1984). The farmer and the cowman should be friends: An attempt at reconciliation with Batson, Coke, and Pych. *Journal of Personality and Social Psychology*, *46*(3), 709-711.

Archer, R. L., Diaz-Loving, R., Gollwitzer, P. M., Davis, M. H., & Foushee, H. C. (1981). The role of dispositional empathy and social evaluation in the empathic mediation of helping. *Journal of Personality and Social Psychology*, *40*(4), 786-796.

Aronson, E., Ellsworth, P. C., Carlsmith, J. M., & Gonzalez, M. H. (1990). *Methods of research in social psychology*. McGraw-Hill Humanities, Social Sciences & World Languages.

Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, *18*(2), 139-152.

Bardsley, N. (2008). Dictator game giving: altruism or artefact?. *Experimental Economics*, *11*(2), 122-133.

Batson, C. D. (1997). Self-other merging and the empathy-altruism hypothesis: Reply to Neuberg et al. (1997). *Journal of Personality and Social Psychology*, *73*(3), 517-522.

Batson, C.D. (2010) Empathy-induced altruistic motivation. In M. Mikulincer & P. R. Shaver (Eds.), in *Prosocial motives, emotions, and behavior: The better angels of our nature* (pp. 15–34). New York: American Psychological Association.

Batson, C.D. (2012). The empathy-altruism hypothesis: issues and implication. *Empathy — From Bench to Bedside*. MIT Press, Cambridge.

Batson, C. D., & Ahmad, N. (2001). Empathy-induced altruism in a prisoner's dilemma II: What if the target of empathy has defected?. *European Journal of Social Psychology*, *31*(1), 25-36.

Batson, C. D., Ahmad, N., Lishner, D. A., Tsang, J., Snyder, C. R., & Lopez, S. J. (2002). Empathy and altruism. *The Oxford handbook of hypo-egoic phenomena*, 161-174.

Batson C.D., Ahmad N. & Stocks E.L. (2011) Four forms of prosocial motivation: Egoism, altruism, collectivism and principlism. In *Social Motivation* (Dunning D., ed.), Psychology Press, New York, pp. 103–126.

Batson, C. D., Batson, J. G., Griffitt, C. A., Barrientos, S., Brandt, J. R., Sprengelmeyer, P., & Bayly, M. J. (1989). Negative-state relief and the empathy—altruism hypothesis. *Journal of Personality and Social Psychology*, *56*(6), 922.

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation?. *Journal of Personality and Social Psychology*, *40*(2), 290.

Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and Social Psychology Bulletin*, *23*(7), 751-758.

Batson, C. D., Lishner, D. A., Cook, J., & Sawyer, S. (2005). Similarity and nurturance: Two possible sources of empathy for strangers. *Basic and Applied Social Psychology*, *27*(1), 15-25.

Batson, C. D., & Shaw, L. L. (1991). Evidence for altruism: Toward a pluralism of prosocial motives. *Psychological Inquiry, 2*, 107–122.

Batson, C. D., Sympson, S. C., Hindman, J. L., Decruz, P., Todd, R. M., Weeks, J. L., ... & Burns, C. T. (1996). " I've been there, too": Effect on empathy of prior experience with a need. *Personality and Social Psychology Bulletin*, *22*(5), 474-482

Bekkers, R., & Ottoni-Wilhelm, M. (2016). Principle of care and giving to help people in need. *European Journal of Personality*, *30*(3), 240-257.

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., ... & Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication?. *Journal of Experimental Social Psychology*, *50*, 217-224.

Cheung, I., Campbell, L., LeBel, E.P., Ackerman, R.A., Aykutoğlu, B., Bahník, Š., Bowen, J.D., Bredow, C.A., Bromberg, C., Caprariello, P.A. and Carcedo, R.J. (2016). Registered replication report: Study 1 from Finkel, Rusbult, Kumashiro, & Hannon (2002). *Perspectives on Psychological Science*, *11*(5), 750-764.

Cunningham, M. R., & Baumeister, R. F. (2016). How to make nothing out of something: Analyses of the impact of study sampling and statistical interpretation in misleading meta-analytic conclusions. *Frontiers in Psychology, 7*, Article 1639.

De Waal, F. B. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual Review of Psychology*, *59*, 279-300.

De Wit, A., & Bekkers, R. (2016). Exploring gender differences in charitable giving: The Dutch case. *Nonprofit and Voluntary Sector Quarterly*, *45*(4), 741-761.

De Wit, A., & Bekkers, R. (2020). Can Charitable Donations Compensate for a Reduction in Government Funding? The Role of Information. *Public Administration Review*, *80*(2), 294-304.

Dovidio, J. F., Allen, J. L., & Schroeder, D. A. (1990). Specificity of empathy-induced helping: Evidence for altruistic motivation. *Journal of Personality and Social Psychology*, *59*(2), 249.

Eckel, C., Grossman, P. J., Johnson, C. A., de Oliveira, A. C., Rojas, C., & Wilson, R. (2011). Social norms of sharing in high school: teen giving in the dictator game. *Journal of Economic Behavior & Organization*, *80*(3), 603-612.

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin*, *101*(1), 91-119.

Fabrigar, L. R., & Wegener, D. T. (2016). Conceptualizing and evaluating the replication of research results. *Journal of Experimental Social Psychology*, *66*, 68-80.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785-791.

FeldmanHall, O., Dalgleish, T., Evans, D., & Mobbs, D. (2015). Empathic concern drives costly altruism. *Neuroimage*, *105*, 347-356.

Finkel, E. J. (2016). Reflections on the commitment–forgiveness registered replication report. *Perspectives on Psychological Science*, *11*(5), 765-767.

Fultz, J., Schaller, M., & Cialdini, R. B. (1988). Empathy, sadness, and distress three related but distinct vicarious affective responses to another's suffering. *Personality and Social Psychology Bulletin*, *14*(2), 312-325.

Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: an evolutionary analysis and empirical review. *Psychological bulletin*, *136*(3), 351-374.

Gotowiec, S., & van Mastrigt, S. (2019). Having versus doing: The roles of moral identity internalization and symbolization for prosocial behaviors. *The Journal of Social Psychology*, *159*(1), 75-91.

Graziano, W. G., & Tobin, R. M. (2002). Agreeableness: Dimension of personality or social desirability artifact? *Journal of Personality*, *70*, 695–727.

Hertz, S. G., & Krettenauer, T. (2016). Does moral identity effectively predict moral behavior?: A meta-analysis. *Review of General Psychology, 20*(2), 129-140.

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795.

Khadjavi, M., & Lange, A. (2015). Doing good or doing harm: experimental evidence on giving and taking in public good games. *Experimental Economics, 18*(3), 432-441.

Klein, R. A., Cook, C. L., Ebersole, C. R., Vitiello, C., Nosek, B. A., Chartier, C. R., ... & Ratliff, K. (2020). Many Labs 4: *Failure to replicate mortality salience effect with and without original author involvement*. PsyArXiv. 10.31234/osf.io/vef2c.

Krebs, D. (1975). Empathy and altruism. *Journal of Personality and Social psychology, 32*(6), 1134-1146.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 1-12.

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, *355*(6325), 584-585.

Mayer, D. M., Aquino, K., Greenbaum, R. L., & Kuenzi, M. (2012). Who displays ethical leadership, and why does it matter? An examination of antecedents and consequences of ethical leadership. *Academy of Management Journal, 55*(1), 151–171.

McAuliffe, W.H.B. (2017). *The psychology of common knowledge explains the appearance of altruistic and moral motivation* (Doctoral dissertation, University of Miami).

McAuliffe, W.H.B., Forster, D. E., Philippe, J., & McCullough, M. E. (2018). Digital altruists: Resolving key questions about the empathy–altruism hypothesis in an Internet sample. *Emotion*, *18*(4), 493-506.

McCullough, M. E., Worthington, E. L., Jr., & Rachal, K. C. (1997). Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology*, *73*(2), 321–336. https://doi.org/10.1037/0022-3514.73.2.321

McNeish, D., & Stapleton, L. M. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research*, *51*(4), 495-518.

Mulder, J., Gu, X., Olsson-Collentine, A., Tomarken, A., Böing-Messing, F., Hoijtink, H., Meijerink, M., et al. (2019). BFpack: Flexible bayes factor testing of scientific theories in r. *arXiv preprint arXiv:1911.07728*.

Neuberg, S. L., Cialdini, R. B., Brown, S. L., Luce, C., Sagarin, B. J., & Lewis, B. P. (1997). Does empathy lead to anything more than superficial helping? Comment on Batson et al.(1997). *Journal of Personality and Social Psychology*, *73*(3).

O'Hara, R. B., & Kotze, D. J. (2010). Do not log-transform count data. *Methods in Ecology and Evolution*, *1*(2), 118-122.

Pfattheicher, S., Sassenrath, C., & Keller, J. (2019). Compassion magnifies third-party punishment. *Journal of Personality and Social Psychology*, *117*(1), 124-141.

Preston, S. D. (2013). The origins of altruism in offspring care. *Psychological Bulletin*, *139*(6), 1305-1341.

Ryan, R. M., & Deci, E. L. (2000b). When rewards compete with nature: The undermining of intrinsic motivation and self-regulation. In C. Sansone & J. M. Harackiewicz (Eds.) *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 13–54). New York: Academic Press.

Schlingloff, L., Csibra, G., & Tatone, D. (2020). Do 15-month-old infants prefer helpers? A replication of Hamlin et al.(2007). *Royal Society Open Science*, *7*(4), 191795.

Schwarz, N., & Clore, G. L. (2016). Evaluating psychological research requires more than attention to the N: A comment on Simonsohn's (2015) "small telescopes". *Psychological Science*, *27*(10), 1407-1409.

Sznycer, D., Delton, A. W., Robertson, T. E., Cosmides, L., & Tooby, J. (2019). The ecological rationality of helping others: Potential helpers integrate cues of recipients' need and willingness to sacrifice. *Evolution and Human Behavior*, *40*(1), 34-45.

Thielmann, I., & Hilbig, B. E. (2018). Is it all about the money? A re-analysis of the link between Honesty-Humility and Dictator Game giving. *Journal of Research in Personality*, *76*, 1-5.

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, *146*(1), 30-90.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192-196.

Weinstein, N., & Ryan, R. M. (2010). When helping helps: autonomous motivation for prosocial behavior and its influence on well-being for the helper and recipient. *Journal of Personality and Social Psychology*, *98*(2), 222-244.

Wilhelm, M. O., & Bekkers, R. (2010). Helping behavior, dispositional empathic concern, and the principle of care. *Social Psychology Quarterly*, *73*(1), 11-32.

Zaki, J. (2020). Integrating empathy and interpersonal emotion regulation. *Annual Review of Psychology*, *71*, 517-540.

Zhao, K., Ferguson, E., & Smillie, L. D. (2017). Individual differences in good manners rather than compassion predict fair allocations of wealth in the dictator game. *Journal of Personality*, *85*(2), 244-256.