AMERICAN PSYCHOLOGICAL ASSOCIATION

Emotion

© 2024 The Author(s)
ISSN: 1528-3542

2024, Vol. 24, No. 8, 1868–1884
https://doi.org/10.1037/emo0001339

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

Thomas G. McCauley[1], William H. B. McAuliffe[2], and Michael E. McCullough[1]
[1] Department of Psychology, University of California, San Diego
[2] Rethink Priorities, San Francisco, California, United States

When people experience empathy for a needy stranger, efforts to help are often not far behind. But does empathy actually *cause* prosocial behavior? And if so, does it activate genuine concern or more self-interested motivations? To rule out the alternative hypothesis that empathy motivates prosocial behavior by generating fear of social disapproval for acting selfishly, Fultz et al. (1986) manipulated empathy for a lonely stranger using perspective-taking instructions; they also manipulated whether subjects believed their decision to help would remain anonymous. However, Fultz et al. conducted their experiment decades ago, with few subjects, and before some potentially important cultural changes in college students' values and social lives. Here, in a preregistered replication with 280 undergraduates, we tested Fultz et al.'s key assertions. The perspective-taking and social evaluation manipulations influenced scores on the manipulation check measures mostly in theory-consistent ways but did not significantly influence helping. Consistent with theory, empathy was positively associated with prosocial behavior. We also found evidence that endorsement of the principle of care reflects genuine concern for needy strangers and that moral identity symbolization reflects a desire to help in order to avoid social disapproval. We consider these results a partially successful replication of key tenets of the empathy–altruism hypothesis, though questions remain about the conditions under which perspective-taking promotes prosocial behavior and about the generalizability of our findings to populations beyond undergraduate women circa 1986. Our results also help illuminate the motivational underpinnings of two individual differences that predicted prosocial behavior in previous research.

Keywords: morality, altruism, empathy, replication

Supplemental materials: https://doi.org/10.1037/emo0001339.supp

Why do we help other people? Social psychologists have proposed four distinct classes of motivations that might underlie our willingness to help others (Batson, 2022; Batson & Powell, 2003). These include (a) altruism (the desire to boost the welfare of another person), (b) egoism (the desire to obtain material, relational, or emotional benefits for oneself), (c) communalism (the desire to benefit one's social group), and (d) principlism (the desire to behave consistently with one's ethical principles, such as the Golden Rule or the duty to rescue). Social psychologists have focused much of their attention on efforts

This article was published Online First July 8, 2024.
Sara Konrath served as action editor.
Thomas G. McCauley https://orcid.org/0000-0001-7307-4391
Michael E. McCullough https://orcid.org/0000-0001-9833-2713
This project/publication was made possible through the support of a grant from the John Templeton Foundation Grant 61539. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation. This article was made possible through the support of Grant W911NF2010259 from the Army Research Institute. The views, opinions, and/or findings contained in this report (article) are those of the authors and shall not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documents.

Supporting materials and preregistration documents for this project can be accessed at https://osf.io/2egsd/?view_only=51d07a565a86482ebc17c0ba12b34205.

Thomas G. McCauley played a lead role in data curation and formal analysis, a supporting role in methodology, and an equal role in investigation, writing–original draft, and writing–review and editing. William H. B. McAuliffe played a supporting role in formal analysis and an equal role in conceptualization, investigation, methodology, writing–original draft, and writing–review and editing. Michael E. McCullough played a lead role in funding acquisition and supervision and an equal role in conceptualization, methodology, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Michael E. McCullough, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, United States. Email: memccullough@ucsd.edu

to determine whether prosocial behavior in any given situation is due to altruistic motivation or, alternatively, one or more egoistic motivations. However, establishing that altruistic motivation is responsible for helping is difficult: In any given case of helping, it is possible to posit an egoistic motive as well as an altruistic one (Batson, 2011). For example, one might help a stranger out of a concern for the stranger's welfare, but one also might help out of a desire to make a new friend.

The social psychologist C. Daniel Batson developed methods for identifying which motivations are active in any particular help-giving situation. Drawing on early evidence that empathy (defined in this literature as feelings of concern and pity for needy or distressed others) promotes (or at least is correlated with) helping behavior (Krebs, 1975; Stotland, 1969), Batson proposed the empathy–altruism hypothesis, which specifies that empathy for a distressed person reliably elicits altruistic motivation (Batson, 1991). To test the empathy–altruism hypothesis, Batson and colleagues sought to design experiments that could pit the predictions of the empathy–altruism hypothesis against alternative hypotheses specifying that empathy promotes helping via a nonaltruistic motivation. This experimental design came to be called the *cross-cutting experiment* (see Batson, 2011, for a review).

In a typical cross-cutting 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 college student or to help raise money for a financially strapped one). Next, subjects undergo an experimental manipulation of empathy—the most common of which involves 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 and focus solely on the facts as they consider the person's plight (the low-empathy condition). Subjects simultaneously undergo an experimental manipulation of a situational variable that varies whether they can fulfill an egoistic motivation by helping the needy person (Batson, 2011). These egoistic incentives have included hedonic rewards for helping (Batson et al., 1988, 1989, 1991) and the opportunity to avoid vicarious negative affect by limiting one's exposure to the needy person altogether (Batson et al., 1981). If empathically aroused subjects persist in helping regardless of whether an egoistic incentive is on offer, as Batson and colleagues have reasoned, then their helping might plausibly be motivated by altruism. On the other hand, if empathy selectively increases helping in the presence of a given egoistic incentive, then one might reasonably infer that empathy arouses egoistic motivation instead.

One of the first egoistic alternatives to the empathy–altruism hypothesis that Batson articulated was based on the observation that acquiring a reputation for callous indifference can result in social exclusion (Archer et al., 1981): People do not like being ostracized and therefore seek to avoid it. According to the empathy-specific evaluation hypothesis, empathy for a needy other might cause helping not because it activates altruistic motivation but instead because it motivates people to take actions that will help them avoid acquiring a reputation for callousness.

Fultz et al. (1986) pitted the empathy–altruism hypothesis against the empathy-specific evaluation hypothesis by simultaneously manipulating empathy (using perspective-taking instructions) and the perceived anonymity of the subject's decision to help a lonely subject who was eager to make a new friend. Subjects in the high social evaluation condition were led to believe (a) that the experimenter knew that the other subject needed a friend, (b) that their responses would be seen by the experimenter, and (c) that the lonely person would know whether the subject declined to help (by volunteering for a successive experiment that would have allowed the subject and the lonely person to get to know each other). 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 experimenter would not see their responses, and that their partner would not know that they were in a position to help. 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 subjects believed that their decisions to help would be anonymous. Additionally, Fultz et al. (1986) found that self-reported empathic concern, but not self-reported personal distress, predicted volunteering regardless of whether subjects were in the high or low social evaluation condition. This pattern of findings implied to Fultz and colleagues that empathy-induced helping is underwritten not by a motivation to avoid social pressure (or personal distress), but instead by altruistic motivation.

Fultz et al.'s (1986) claims about the empathy–altruism hypothesis actually rest on a network of hypotheses regarding the relationships among empathy, fear of negative social evaluation, and helping behavior, most of which Fultz et al. examined directly. First, they examined whether their perspective-taking manipulation increased (a) empathic emotion toward the lonely stranger and (b) personal distress. Next, they examined whether their manipulation of social evaluation influenced subjects' scores on two manipulation check items that measured subjects' perceptions that their decisions to help would be anonymous. Additionally, they examined whether the perspective-taking manipulation, the social evaluation manipulation, and their interaction influenced subjects' willingness to help the needy stranger. As expected, they found that the perspective-taking manipulation increased empathy (but not personal distress), that the social evaluation manipulation increased people's scores on two manipulation check measures, and that the perspective-taking manipulation increased subjects' willingness to help the stranger. They also found that self-reported empathic concern for the stranger was positively correlated with helping behavior. Perhaps most importantly—from their vantage point of wanting to pit the empathy–altruism hypothesis against the empathy-specific evaluation hypothesis—they found that the perspective-taking manipulation increased subjects' willingness to help regardless of whether their opportunity to help was under high or low social evaluation: They did not find a statistical interaction of the perspective-taking manipulation and the social evaluation manipulation.

On the basis of this pattern of results, Fultz and colleagues claimed victory for the empathy–altruism hypothesis, writing

> This pattern of results is not consistent with what would be expected if the empathically aroused individual is motivated to avoid looking bad in the eyes of others, but it is quite consistent with what would be expected if empatny [sic] evokes altruistic motivation to reduce the victim's need. (p. 767)

Batson's research has been influential. Most prominent theories of the origins of empathy and prosocial motivation either rely on the truth of the empathy–altruism hypothesis or seek to accommodate it (de Waal, 2008; Goetz et al., 2010; Hoffman, 2000; Preston, 2013; Sznycer et al., 2019; Zaki, 2020), and every graduate and

undergraduate social psychology textbook prescribed by the Society for Personality and Social Psychology includes a detailed description of the empathy–altruism hypothesis. Indeed, even though other laboratories have contributed independent evidence in favor of the empathy–altruism hypothesis (e.g., FeldmanHall et al., 2015; McAuliffe et al., 2018), the broad consensus in its favor owes to Batson's own research program. Fultz et al.'s (1986) work has been particularly generative: It has been cited hundreds of times, it features prominently in a number of undergraduate textbooks (e.g., Gilovich et al., 2018; Myers & Twenge, 2022) as an illustration of the empathy–altruism hypothesis, and it is the sole piece of evidence that Batson has marshalled in support of the idea that empathy promotes helping even in the absence of social incentives (Batson, 2011).

Despite its impact, however, there are reasons to wonder whether Fultz et al.'s (1986) results were reliable in the first place. First, it should be noted that Fultz et al. studied a small number of subjects—32 undergraduate women from the University of Texas—who were assigned to four cells in a 2×2 design, which probably underpowered their ability to detect an interaction effect, if one indeed exists. Many scientific fields are beset by so-called publication bias—a sociological phenomenon resulting from reviewers' and editors' special enthusiasm for publishing results that are statistically significant. The spectre of publication bias looms large in fields in which experiments with small sample sizes are the norm because the vagaries of sampling error will cause many of those experiments to spuriously produce statistically significant results even when the null hypothesis is true. Given how much the empathy–altruism literature relies on the veracity of an experiment with such a small sample size, we believe it warrants a direct replication.

As the action editor and reviewers have pointed out to us, there are also substantive reasons why a replication conducted decades later might not produce the same results, even if the original experiment's results were reliable at the time. Most obviously, Fultz et al.'s (1986) exclusive focus on college women creates uncertainty about whether their conclusions generalize to college men. More broadly, however, undergraduates have changed more psychologically over the past several decades than many people might realize. Analyses of cohort data over time have revealed that high school seniors and college undergraduates today differ in many ways from their 1980s counterparts—some of which could be relevant to how they would respond in Fultz et al.'s experiment today. For example, they are less involved in religion than in previous decades (Twenge et al., 2015). Because compassion has a weaker relationship with prosocial behavior for religious people than for nonreligious people (Saslow et al., 2013), it is possible that the effects of experimentally induced empathy on prosocial behavior are stronger today than they would have been when Fultz et al. conducted their research merely because today's young people are, as a group, less religious than they used to be.

Today's young people are also lonelier than in decades past (Twenge et al., 2019). Empathy is evidently related more strongly to moral/prosocial traits among lonely people than among nonlonely people (Jiao & Wang, 2018), so the high prevalence of loneliness among young people in the 21st century could contribute to making the effect of experimentally induced empathy on prosocial behavior stronger than it might have been in earlier decades.

Finally, today's young people have more experience with volunteering and expect to do more volunteering during their years in college than in decades past (Twenge et al., 2012). Empathy is more strongly related to prosocial behavior among people who do not volunteer in charitable organizations than among those who do—in fact, empathy appears to be negatively related to prosocial behavior among young people involved in volunteering (Nowakowska, 2022)—so the high rates of volunteering among young people today could contribute to reducing or even reversing the effect of empathy on prosocial behavior.

For these reasons—the small number of subjects in Fultz et al.'s (1986) original experiment, their exclusive focus on college women, and (as we came to recognize after we finished conducting this project) psychological changes in young people since the 1980s—we believe their influential and theoretically important experiment warrants a direct replication. In the experiment reported here, we directly replicated Fultz et al.'s (1986) Study 1 in order to evaluate their network of hypotheses—along with three novel ones—by testing seven preregistered predictions.

Predictions About Manipulation Checks

The first two of our predictions are lynchpins of Fultz et al.'s (1986) argument, and the validity of Fultz et al.'s experiment requires those predictions to be correct: They involve tests of whether Fultz et al.'s experimental manipulations of empathy and social evaluation actually do the jobs they were designed to do. First, we tested whether the perspective-taking manipulation increased self-reported empathic concern. Second, we tested whether the social evaluation manipulation convinced subjects that their decision to help the lonely subject would remain anonymous.

Predictions About Empathy and Prosocial Behavior

We also tested two preregistered predictions at the heart of the empathy–altruism hypothesis itself: First, we tested whether subjects who received imagine-other instructions volunteered to spend more time with the lonely subject than subjects who received remain-objective instructions, regardless of whether they were assigned to the high or low social evaluation condition. Second, we tested whether empathic concern was associated with volunteering regardless of whether subjects were assigned to the high or low social evaluation condition and even when simultaneously controlling for personal distress.

Predictions About Individual Differences in Prosocial Motivation

We also used this experiment as an opportunity to test three predictions about individual-difference variables that might have different associations with prosocial behavior when egoistic incentives are present than when such incentives are absent. (In this sense, our experiment is not only a direct replication of Fultz et al., 1986, but also an extension of it.) Holding aside whether empathic concern for any specific needy person promotes prosocial behavior on that person's behalf—and, if so, whether it does so by activating altruistic motivation (vs. egoistic motivations such as the desire to avoid social disapproval)—scholars have examined several individual differences also believed to measure the motivational substrates of prosocial behavior, including the willingness to engage

in prosocial behavior even in the absence of egoistic incentives (such as the fear of negative social evaluation we study here).

First, we examined whether the endorsement of the principle of care (the internalized moral value that one has a moral duty to help people in need; Wilhelm & Bekkers, 2010) predicts helping—even in the absence of social evaluation pressure. Past research has revealed that people who endorse the principle of care donate more to charity (Bekkers & Ottoni-Wilhelm, 2016), and it may motivate persistence in helping others, even as the egoistic incentives for doing so fluctuate. However, we currently do not know whether the association of endorsement of the principle of care with prosocial behavior association is driven by altruism or principlism—that is, by an intrinsic desire to abide by an endorsed moral principle (Batson, 2011)—or merely by a desire to *appear* altruistic or principled. In keeping with the developers' theory of the measure, we predicted that endorsement of the principle of care would predict prosocial behavior more strongly among subjects in the low social evaluation condition than among those in the high social evaluation condition.

Second and third, 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 & Reed, 2002) would predict helping. Internalization is hypothesized to promote helping even when social and material rewards for helping are absent, whereas symbolization is thought to promote helping precisely in order to secure such rewards. Those who profess to have a strong internalized moral identity engage in more prosocial behavior (Hertz & Krettenauer, 2016), and they also experience more empathy for others (Lefebvre & Krettenauer, 2019). As far as we are aware, however, studies of the associations of moral identity internalization with prosocial behavior have not involved experimental manipulation of whether social rewards for helping were on offer. Based on the developers' theory of these measures, we predicted that self-reported moral identity internalization would be associated with prosocial behavior more strongly for subjects in the low social evaluation condition than for those in the high social evaluation condition, and that self-reported moral identity symbolization would be associated with prosocial behavior more strongly for subjects in the high social evaluation condition than for those in the low social evaluation condition.

Method

Transparency and Openness

We report how we determined our sample size. Complete statements of our seven preregistered hypotheses, along with our preregistered study design and analysis plan, as well as study materials, data, and analysis code, can be found at https://osf.io/2egsd/?view_only=51d07a565a86482ebc17c0ba12b34205. We submitted an amendment to our original preregistration because we discovered more appropriate statistical methods for analyzing the data (viz., fitting data to models that can account for zero-inflated outcomes; controlling for experimenter effects and suspicion) in the months following our preregistration. The results we report here resulted from the new analyses. In Supplemental Materials, we report results based on the analyses described in the original preregistration, but the results of the original and amended analyses were not qualitatively different. We also 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. We created interaction plots using the interact_plot() function from the Interactions package in R. All materials and procedures were approved by the institutional review boards at the University of Miami and the University of California, San Diego. We report all manipulations, measures, and exclusions in these studies.

Subjects

We preregistered our intention to collect data until we had reached 400 total subjects, or until we had collected data from 75 nonsuspicious participants in each cell (300 total subjects), whichever came first. However, as stated in the amendment to our preregistration, we stopped collecting data early due to the COVID-19 pandemic. We ended up collecting data from 129 subjects at the University of Miami and 151 subjects at the University of California, San Diego, from 2017 until 2020. 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 hr) and were therefore removed from analyses (excluded subjects were evenly distributed across the experimental conditions), leaving 277 subjects (M_{age} = 19.93, SD_{age} = 1.97; *males* = 81, *females* = 195;[1] *Caucasian* = 93, *Asian* = 93, *Hispanic* = 49, *Black* = 16 *other* = 29; we did not collect data regarding nativity or immigration history, socioeconomic status, or clinical diagnoses). After the fact, we determined that a sample size of as few as n = 221 subjects was sufficient to detect main effects as small as Cohen's d = 1.04, which is the effect size for the perspective-taking instructions on volunteering (which we computed based on statistics in Fultz et al., 1986), with nearly 100% power. Thus, even with a smaller final sample than we had hoped to obtain, our experiment was well-powered with 277 subjects.

Procedure

We made minor changes to Fultz et al.'s (1986) protocol to enhance its experimental realism in our laboratory (e.g., see in the following section about participants recycling their partner's note), but otherwise we duplicated Fultz et al.'s (1986) protocol as faithfully as possible. In Supplemental Materials, we provide more details on the differences between Fultz et al.'s (1986) protocol and ours.

Preexperiment 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 improve scientists' understanding of how people form impressions of others based on verbal communication alone (i.e., in the absence of nonverbal or visual cues). Subjects were then led to believe that they had been partnered with another subject of the same sex. In reality, there was no other person. 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 Tycho or Janet Tycho at the institutions where the experiment was conducted, and subjects denied familiarity with a Mark Tycho or a Janet Tycho). After subjects responded negatively, the research assistant told them that their partner also denied knowing them.

[1] One subject did not provide information about sex.

Next, subjects received written introductory instructions stating that either they or their partner would be assigned to be the "communicator," who would then write a 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 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 we asked them to read before reading the communicator's note. Subjects in the high-empathy condition received imagine-other instructions and were asked to focus on imagining the feelings of their 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 Supplemental Materials for the text of both mindset instructions.

Need Situation. Subjects then read the communicator's note. Participants in our experiment received just a single note from the communicator, whereas participants in Fultz et al.'s (1986) experiment received two notes, one after the other; there was no difference in the content of the notes, only that the notes were sent via one, rather than two, sheets of article. The note indicated that the communicator had recently transferred from a university in his or her hometown. The note goes on to express how lonely the communicator has felt since transferring. See Supplemental Materials for the text of the note.

Emotion Response Questionnaire. 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 distractors unrelated to empathy or distress. See Supplemental Materials for all items from the Emotion Response Questionnaire. For each adjective, subjects endorsed how much each adjective described their emotional state on a Likert scale, with response options ranging from 1 (*not at all*) to 7 (*extremely*). We formed unit-weighted composites for the empathic concern items ($M = 4.86$, $SD = 1.19$, McDonald's $\omega = 0.82$) and the personal distress items ($M = 2.59$, $SD = 1.16$, McDonald's $\omega = 0.82$).

Impression Questionnaire. Subjects also completed a questionnaire indicating their impressions of their interaction partner (i.e., the communicator). See Supplemental Materials for all items from the Impression 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. Next, the experimenter delivered to the subjects a letter from the professor in charge of the experiment. The letter described an opportunity to volunteer for a second study, ostensibly unrelated to the first, about long-term relationships. Subjects were informed that if they agreed to volunteer, they would be paired with Janet/Mark for an initial meeting, and the experimenter would occasionally ask subjects to report on the development of the friendship. Subjects were asked to indicate on a form whether they were willing to volunteer and, if so, for how many hours.

Social Evaluation Manipulation. In the high social evaluation condition, the letter indicated that both the subject and the partner would be made aware of the study opportunity and that the subject and partner would be recruited for the relationship study only if both people agreed to participate. The subject was then led to believe that the research assistant had read Mark/Janet's note and was comparing the subject's impressions of Mark/Janet (gleaned from the Impression Questionnaire) with the text from the partner's note to determine how accurate the subject's impression was, ensuring that the research assistant would learn about the contents 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 knew about their partner's loneliness, and the partner and experimenter knew that the subject knew).

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

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 subjects believed that the experimenter and the partner knew that the participant had had an opportunity to relieve the partner's distress: "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. We measured subjects' endorsement of the principle of care with the eight-item Principle of Care Scale (Bekkers & Ottoni-Wilhelm, 2016), which includes items such as "When people are less fortunate, it is important to help them even if they are very different from us." Subjects indicated how much they agreed with each item on a self-report Likert scale, with response options ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). We formed a unit-weighted composite from the eight items ($M = 4.11$, $SD = 0.52$, McDonald's $\omega = 0.83$).

Moral Identity Scale. We measured subjects' self-reported moral identity with the Moral Identity Scale (Aquino & Reed, 2002), which describes a moral person and asks subjects to rate their agreement with items such as "Being someone who has these characteristics is an important part of who I am" from the five-item internalization subscale and "I often wear clothes that identify me as having these characteristics" from the five-item symbolization subscale. Subjects indicated how much they agreed with each item on a self-report Likert scale, with response options ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). We formed unit-weighted composites for both subscales, with five items for the internalization subscale ($M = 4.52$, $SD = 0.51$, McDonald's $\omega = 0.79$) and five for the moral symbolization subscale ($M = 3.26$, $SD = 0.75$, McDonald's $\omega = 0.84$).

Subjects who participated in the study at the University of Miami completed the three individual-difference measures during

pretesting data collection; subjects at the University of California, San Diego, did so at the end of the study.

Suspicion Probe. Finally, before ending the experimental session, we attempted to determine whether subjects were suspicious about the experiment, and we debriefed them with a full explanation of how and why we had deceived them. For the suspicion probe, we used a funnel debriefing, which is a three-question 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). See Supplemental Materials for the text of the questions we asked subjects.

The plans for analyzing the preregistered hypotheses can be found on the Open Science Framework (https://osf.io/2egsd/?view_only=51d07a565a86482ebc17c0ba12b34205).

Results

Here, we report the results from our modified preregistration. Results from the original preregistration are included in Supplemental Materials.

Testing Predictions About Manipulation Checks

We began by testing our two predictions about the manipulation checks, which involved examining whether the manipulations for inducing empathy and concern about social evaluation had their intended effects.

Prediction 1: Did the Perspective-Taking Manipulation Increase Empathy?

Yes. Consistent with Prediction 1, subjects in the imagine-other condition scored higher on the empathic concern composite ($M = 5.04$, $SD = 1.18$) than did subjects in the remain-objective condition, $M = 4.68$, $SD = 1.17$; $t(275) = 2.59$, $p = .01$, Cohen's $d = 0.31$, 95% CI [0.07, 0.55].

To provide a quantitative estimate of how strongly this evidence favors the focal hypothesis over the null hypothesis, we conducted a Bayesian analysis. In performing 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 (BF; Kass & Raftery, 1995). Bayes factors are centered on 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 and approach 0, they indicate increasingly strong support for the null hypothesis.

By Kass and Raftery's (1995) 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, values ranging from 10 to 100 indicate "strong support" for the focal hypothesis, and values exceeding 100 indicate "decisive" support for the focal hypothesis. We specified equal prior probabilities for the focal hypothesis that the effect would be greater than zero versus equal to or less than zero. We found that the Bayes factor associated with the test of whether the perspective-taking manipulation increased empathic concern was 3.10, implying that the focal hypothesis—that imagine-other instructions elicit more empathy than remain-objective instructions—is approximately three times more likely than the null hypothesis that the perspective-taking instructions do not influence empathy or reduce it. That said, the 3.10 value for the Bayes factor lies beneath the minimum value that Kass and Raftery identified as indicative of "substantial support" (3.2). So, even though the effect for the perspective-taking manipulation was statistically significant, the evidence was not overwhelmingly strong.

Prediction 2: Did the Social Evaluation Manipulation Convince Subjects That Their Responses Would Remain Anonymous?

Yes and No. Consistent with Prediction 2, subjects in the high social evaluation condition indicated greater agreement ($M = 5.09$, $SD = 2.08$) 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% CI [1.20, 1.78], BF = 1.64×10^{25}. However, contradicting Prediction 2, 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% CI [−0.29, 0.18], with a Bayes factor that indicates strong support for the null hypothesis (BF = 0.14). Thus, the social evaluation manipulation was successful in convincing subjects in the low social evaluation condition that their partner would not know if they volunteered, but it failed to convince them that their decision to volunteer was fully anonymous—plausibly because subjects understood that the experimenter would have access to their responses at the end of the experiment.

Testing Predictions About Main Effects

The number of hours subjects volunteered to help the partner is a count variable and cannot take on negative values, so fitting our data to an ordinary least squares model would have been inappropriate (Atkins & Gallop, 2007). Therefore, before testing Predictions 3–7, we compared four statistical models that are suitable for count data 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.[2]

For each of the regression models we used to test predictions 3–7, we included predictors that controlled for the effects of test sites, research assistants, and subjects' suspicion about the deceptive elements of the procedure (e.g., subjects' beliefs about whether the note from their partner was a sham). To control for the effects of test sites and research assistants, we included 10 dummy variable codes (1 = *administered the session*, 0 = *did not administer the session*) for the 11 undergraduate research assistants (RAs) who administered experimental sessions, with four RAs at one university and seven at the other. Coding for research assistants in this manner also controls for any effects of test site. To control for the effects of suspicion, we

[2] Our modified preregistration 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.

coded for suspicion based on subjects' responses to the suspicion probe. Subjects were categorized as suspicious if, during the funnel debriefing, they indicated that they believed (a) that the communicator was not real, (b) that the communicator received different mindset instructions than they received (or the communicator was told what to say or do), or (c) that the experiment was designed to evaluate subjects' 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 (1 = *hard suspicion*, 0 = *no hard suspicion*), or whether they merely expressed uncertainty or doubt, which we call "soft" suspicion (1 = *soft suspicion*, 0 = *no soft suspicion*). See Supplemental Materials for details about how we coded for hard and soft suspicion.

In sum, each candidate model described hereafter includes the number of hours volunteered as the dependent variable and the following predictor variables: one focal variable (e.g., the perspective-taking manipulation), one predictor for the social evaluation manipulation, one predictor for the interaction of the focal variable and the social evaluation manipulation, 10 dummy-coded covariates to control for the research assistant who ran the session, and two dummy-coded covariates to control for hard and soft suspicion.

To select the model that was most appropriate for modeling the underlying distribution of subjects' volunteering hours, we calculated Akaike information criterion fit indices for the negative binomial, zero-inflated negative binomial, zero-inflated Poisson, and multinomial ordered logit models. Consistent with our updated preregistration, we then selected the model with the lowest Akaike information criterion. 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).

We selected the best statistical model for each prediction separately (e.g., for Prediction 5, we compared four models, all of which included the covariates controlling for the research assistant who ran the session and suspicion, endorsement of the principle of care as the focal predictor, the social evaluation manipulation, and the interaction of endorsement of the principle of care and the social evaluation manipulation; for Prediction 6, we compared four models that included the same set of predictors, except that they featured moral identity internalization as the focal predictor).[3] A comparison of the Akaike weights indicated that the negative binomial model provided the best fit in each case ($1.05 < \textit{Akaike Weights} < 278.56$), indicating that the negative binomial model is more likely to be the true model, relative to a zero-inflated negative binomial model, a zero-inflated Poisson model, and a multinomial ordered logit model (see Supplemental Table S1 for results of the model selection analyses for all predictions). Therefore, we retained the negative binomial model for the remainder of all analyses.

Finally, we note that, for the models that include personal distress as a covariate, some additional subjects were excluded from analyses for failing to provide a response to an item from the Emotion Response Questionnaire used to form the personal distress composite ($n = 3$) or the empathic concern composite ($n = 3$) or for failing to indicate how many hours they would volunteer ($n = 3$).

Table 1
Descriptive Statistics of Volunteering Decisions for Subjects in Each of the Four Combinations of the Perspective-Taking and Social Evaluation Manipulations

Social evaluation condition	Low perspective-taking	High perspective-taking
	Our experiment	
Low social evaluation	$M = 1.99$, $SD = 2.74$ $N = 70$ Volunteered = 55.7%	$M = 2.16$, $SD = 3.02$ $N = 69$ Volunteered = 58%
High social evaluation	$M = 2.19$, $SD = 2.32$ $N = 70$ Volunteered = 72.9%	$M = 2.25$, $SD = 2.71$ $N = 68$ Volunteered = 63.2%
	Fultz et al.'s (1986) experiment	
Low social evaluation	$M = 1.29$ $N = 7$	$M = 2.44$ $N = 9$
High social evaluation	$M = 0.67$ $N = 9$	$M = 1.71$ $N = 7$

Note. Volunteered = the percentage of participants in a given condition who volunteered for at least 1 hr. Fultz et al. (1986) did not include information about standard deviations for volunteering nor the number of participants who volunteered in each cell.

Prediction 3: Was the Effect of Perspective-Taking on Volunteering Moderated by Social Evaluation?

No, but There Was No Main Effect of Perspective-Taking Either. First, we tested whether the perspective-taking instructions promoted volunteering. Descriptive and inferential statistics appear in Tables 1 and 2. Refuting Prediction 3, the perspective-taking manipulation did not influence the number of hours subjects volunteered. This finding contradicts Fultz et al. (1986)—and the mainstay of Batson's research, actually (Batson, 2011). Consistent with preregistered Prediction 3, the interaction between perspective-taking and social evaluation was nonsignificant, but of course this is unsurprising in light of the overall ineffectiveness of the perspective-taking manipulation. We also did not find a main effect for the social evaluation manipulation (all ps $> .699$; see Supplemental Table S2 for the statistics related to the suspicion and experimenter covariates and Supplemental Table S7 for analyses that do not include experimenter effects and the effects of suspicion). The Bayes factors for the three effects modeled to test Prediction 3 ranged from 0.88 to 1.86, which imply no meaningful support for any of the hypotheses (Kass & Raftery, 1995). Bayes factors for these hypothesis tests (and all others) appear in Tables 2–6. We have also included the descriptive statistics for volunteering reported in Fultz et al. (1986) for comparison.

In response to editorial comments, we also examined whether these conclusions generalized over sex and ethnicity. To do so, we reran the model separately for (a) men versus women and (b) Caucasian versus Asian versus non-Caucasian–non-Asian subjects (this tripartite distinction among ethnicities was the only one we could create that would give us reasonable numbers of subjects in each category). We found that the results for the subgroup analyses were qualitatively identical to results that included all subjects in

[3] We selected the same model for Predictions 3 and 4, since they are derived from the same model.

Table 2

Coefficients for the Negative Binomial Regression Model That Includes Predictors for the Perspective-Taking Manipulation, the Social Evaluation Manipulation, and the Interaction of the Perspective-Taking and Social Evaluation Manipulations

Predictor	*b* (*SE*)	*p*	Bayes factor
Perspective-taking manipulation	0.09 (0.23)	.699	1.86
Social evaluation manipulation	−0.00 (0.23)	.987	0.98
Empathy Manipulation × Social Evaluation Manipulation	−0.03 (0.32)	.934	0.88

Note. The *b*s reported for the negative binomial model are logged coefficients. For the terms associated with empathic concern, the social evaluation manipulation, and personal distress, the Bayes factor reflects the strength of the evidence for the hypothesis that the effect is positive versus the evidence for the hypothesis that it is zero or negative. For the interaction of empathic concern and the social evaluation manipulation, the Bayes factor reflects the strength of the evidence for the hypothesis that the coefficient is negative versus the evidence for the hypothesis that the coefficient is either zero or positive. *SE* = standard error.

analyses. Results of the statistical tests for these analyses are reported in Supplemental Tables S8–S12.

We ran two linear regression models that included the empathy manipulation, social evaluation manipulation, and their interaction. The first model included self-reported personal distress as the outcome, and the second model included self-reported empathic concern as the outcome. We found no significant predictors in either model (*p*s > .216). See Supplemental Tables S13–S14 for the coefficients from both models and Supplemental Tables S15–S16 for descriptive statistics for the effect of the empathy and social evaluation manipulations upon empathic concern and personal distress.

Prediction 4: Did Self-Reported Empathic Concern Predict Volunteering (Regardless of Whether Subjects Were Assigned to the High or Low Social Evaluation Condition), Even After Controlling for Self-Reported Personal Distress?

Yes. To test this prediction, we calculated a negative binomial model in which we regressed the number of hours subjects volunteered on their scores on self-reported empathic concern while controlling for (a) which level of the social evaluation manipulation they were assigned to, (b) the interaction of empathic concern with the social evaluation manipulation, and (c) self-reported personal distress. Even though self-reported empathic concern had a significant bivariate correlation with self-reported personal distress, $r(273) = 0.29$, 95% CI [0.18, 0.39], $p < .001$, Table 3 shows that empathic concern was positively associated with volunteering ($b = 0.31$, $SE = 0.10$, 95% CI [0.11, 0.51], $p = .002$) even after controlling for its association with personal distress ($b = -0.03$, $SE = 0.07$, 95% CI [−0.17, 0.11], $p = .673$). The Bayes factor for this association (BF = 810.18) indicates that our data provide "decisive" support for the hypothesis that self-reported empathic concern is associated with volunteering (Kass & Raftery, 1995, p. 777). Thus, in keeping with a recent meta-analysis (Yin & Wang, 2023), we found strong support for preregistered Hypothesis 4 (see Table 3).

In response to an editorial request, we also ran an exploratory analysis testing whether there was an indirect effect of perspective-taking instructions on helping that was mediated by self-reported empathic concern. We tested the indirect effect by fitting a structural equation model in which we regressed volunteering hours on the perspective-taking manipulations (0 = *remain-objective*, 1 = *imagine-other*), and then, in a separate equation, regressed volunteering on the perspective-taking manipulation and state empathic concern. The direct effect of the perspective-taking manipulation on the number of hours volunteered was nonsignificant ($c = -0.07$, $SE = 0.32$, $p = .826$). With respect to the indirect effect, the perspective-taking manipulation had a significant positive effect on state empathic concern ($a = 0.36$, $SE = 0.14$, $p = .012$), state empathic concern had a significant positive association with volunteering ($b = 0.48$, $SE = 0.14$, $p < .001$), and the indirect effect of the perspective-taking instructions on volunteering was significant (*indirect effect* = 0.17, $SE = 0.08$, $p = .041$). The total effect was nonsignificant ($p = .762$). These results suggest that perspective-taking instructions affect volunteering, potentially through their intermediate effect on empathic concern. These results are consistent with Batson et al.'s (1986; Batson, 2018) claim that perspective-taking instructions exert their

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

Predictor	*b* (*SE*)	*p*	Bayes factor
Empathic concern	0.31 (0.10)	.002**	810.18
Social evaluation manipulation	0.67 (0.73)	.360	4.55
Empathic Concern × Social Evaluation Manipulation	−0.14 (0.14)	.344	0.21
Personal distress	−0.03 (0.07)	.673	0.51

Note. The *b*s for the models reported here are logged coefficients. For the terms associated with empathic concern, the social evaluation manipulation, and personal distress, the Bayes factor reflects the strength of the evidence for the hypothesis that the effect is positive versus the evidence for the hypothesis that it is zero or negative. For the interaction of empathic concern and the social evaluation manipulation, the Bayes factor reflects the strength of the evidence for the hypothesis that the coefficient is negative versus the evidence for the hypothesis that the coefficient is either zero or positive. *SE* = standard error.
** $p < .01$.

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

Predictor	*b* (*SE*)	*p*	Bayes factor
Principle of care	0.96 (0.25)	<.001***	20,020.12
Social evaluation manipulation	3.63 (1.34)	.007**	299.66
Principle of Care × Social Evaluation Manipulation	−0.86 (0.32)	.007**	287.404

Note. The *b*s for all models reported here are logged coefficients. For the terms associated with the principle of care and the social evaluation manipulation, the Bayes factor reflects the strength of the evidence for the hypothesis that the effect is positive versus the evidence for the hypothesis that it is zero or negative. For the interaction of the principle of care with the social evaluation manipulation, the Bayes factor reflects the strength of the evidence for the hypothesis that the coefficient is negative versus the evidence for the hypothesis that the coefficient is either zero or positive. *SE* = standard error. ** $p < .01$. *** $p < .001$.

effect on helping via their intermediate effects on empathic concern. See Supplemental Table S17 for the coefficients from the indirect effects model.

Finally, in response to an editorial request, we ran an additional exploratory analysis to test whether the perspective-taking and social evaluation manipulations influenced the decision to volunteer rather than the amount of time spent volunteering. We ran a logistic regression that regressed the decision to volunteer (0 = *did not volunteer*, 1 = *volunteered nonzero number of hours*) on the perspective-taking and social evaluation manipulations. There was a significant main effect for the social evaluation manipulation: Participants in the high evaluation condition were more likely to volunteer than those in the low evaluation condition ($b = 0.76$, $SE = 0.36$, $p = .036$). The effects of the perspective-taking manipulation and its interaction with the social evaluation manipulation were nonsignificant (*p*s > .286). See Supplemental Table S18 for coefficients from the logistic regression.

Testing Predictions About Individual Differences in Prosocial Motivation

In three separate models, we tested preregistered predictions regarding the associations of three individual difference variables (endorsement of the principle of care, moral identity internalization, and moral identity symbolization) with volunteering under conditions in which social evaluation pressure is high versus in which social evaluation pressure is low. Each of the individual differences models had the same form as the models used to test the main effect hypotheses: one focal variable (either the principle of care, moral identity internalization, or moral identity symbolization), one predictor for the social evaluation manipulation, one predictor for the interaction of the focal variable and the social evaluation manipulation, 10 dummy-coded covariates to control for the research assistant who ran the session, and two dummy-coded covariates to control for hard and soft suspicion. Results for all models are shown in Tables 4–6 (see Supplemental Tables S4–S6 for statistics related to the suspicion and experimenter covariates).

Prediction 5: Did Endorsement of the Principle of Care Predict Volunteering More Strongly for Subjects in the Low Social Evaluation Condition Than for Subjects in the High Social Evaluation Condition?

Yes. For the model in which the principle of care was the focal predictor, both the principle of care ($b = 0.96$, $SE = 0.25$, 95% CI [0.47, 1.46], $p < .001$, BF = 20,020.12) and the social evaluation manipulation ($b = 3.63$, $SE = 1.34$, 95% CI [1.03, 6.28], $p = .007$, BF = 299.66) were positively associated with volunteering,[4] and, consistent with our hypothesis, the association between the principle of care and volunteering was modified by the social evaluation manipulation ($b = -0.86$, $SE = 0.32$, 95% CI [−1.49, −0.24], $p = .007$, BF = 287.40). The Bayes factors for the main effect of the principle of care (BF = 20,020.12) and the interaction of the principle of care and social evaluation manipulation (BF = 287.40) implied "decisive support" for the hypothesis that the principle of care positively predicts volunteering and that this relationship is particularly strong among subjects in the low social evaluation condition. When we probed the interaction between the principle of care and social evaluation using simple slopes analyses, we found that the principle of care was positively and significantly associated with hours volunteered for subjects in the low social evaluation condition ($b = 0.96$, $SE = 0.25$, 95% CI [0.48, 1.44], $p < .001$), whereas the positive effect for subjects in the high social evaluation condition was smaller and nonsignificant ($b = 0.10$, $SE = 0.21$, 95% CI [−0.31, 0.50], $p = .631$; see Figure 1).

Finally, in a follow-up simple effects analysis, we subset the data by the high and low social evaluation conditions and reran the model, omitting the social evaluation condition and interaction predictor terms. We found that there was "decisive support" for the main effect of the principle of care in the low social evaluation condition (BF = 3,090.35) but only "substantial support" for the main effect for participants in the high social evaluation condition (BF = 3.83).

Prediction 6: Did Internalized Moral Identity Predict Volunteering More Strongly for Subjects in the Low Social Evaluation Condition Than Those in the High Social Evaluation Condition?

No. For the model in which moral identity internalization was the focal predictor, neither moral identity internalization, the social evaluation manipulation, nor their interaction predicted volunteering (*p*s > .161; BFs < .33).

[4] Although the coefficient for the social evaluation manipulation was significant in this model, we cannot know what causes this suppressor effect (Rohrer, 2018), since the effect was neither predicted nor significant in other models that we tested, and thus do not attempt to further interpret this finding.

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

Predictor	*b* (*SE*)	*p*	Bayes factor
Moral identity internalization	0.37 (0.26)	.161	0.09
Social evaluation manipulation	1.08 (1.54)	.482	0.32
Moral Identity Internalization × Social Evaluation Manipulation	−0.23 (0.34)	.489	0.32

Note. The *b*s for all models reported here are logged coefficients. For the terms associated with moral identity internalization and the social evaluation manipulation, the Bayes factor reflects the strength of the evidence for the hypothesis that the effect is positive versus the evidence for the hypothesis that it is zero or negative. For the interaction of moral identity internalization and the social evaluation manipulation, the Bayes factor reflects the strength of the evidence for the hypothesis that the coefficient is negative versus the evidence for the hypothesis that the coefficient is either zero or positive. *SE* = standard error.

Prediction 7: Did Moral Identity Symbolization Predict Volunteering More Strongly for Subjects in the High Social Evaluation Condition Than Those in the Low Social Evaluation Condition?

Yes. For the model in which moral identity symbolization was the focal predictor, the social evaluation manipulation appeared to reduce volunteering ($b = -1.54$, $SE = 0.74$, 95% CI [−3.01, −0.09], $p = .036$, BF = 54.26), and the association of moral identity symbolization with volunteering behavior was modified by the social evaluation manipulation ($b = 0.47$, $SE = 0.22$, 95% CI [0.04, 0.91], $p = .032$, BF = 62.43). The main effect of moral identity symbolization was not significant ($p = .456$, BF = 0.30). 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% CI [0.05, 0.65], $p = .02$) but not for subjects in the low social evaluation condition ($b = -0.12$, $SE = 0.16$, 95% CI [−0.44, 0.20], $p = .46$; see Figure 2). In a follow-up simple effects analysis, we subsetted the data by the high and low social evaluation conditions and reran the model for each subset, omitting the social evaluation condition and interaction predictor terms. We found that, in the high social evaluation condition, there was decisive support for an association of volunteering with symbolization (BF = 213.1), but in the low social evaluation condition, there was no support for the association (BF = 0.15).

Discussion

Batson and colleagues' classic studies on empathy and prosocial behavior have convinced many social scientists that empathy is causally involved in the production of altruistic motivation. One of the most influential studies is Fultz et al.'s (1986) landmark experiment, which revealed that subjects instructed to imagine the plight of a needy person were more likely to help that person, even after one possible egoistic motive for helping—the desire to avoid negative social evaluation—had been removed from the helping encounter. We replicated their experiment here because of its considerable influence on the field and its small sample size. Moreover, given the limitations to generalizability associated with Fultz et al.'s (1986) exclusive focus on women, along with generational changes in the personalities and social lives of young people over the past 40 years (some of which, plausibly, might have changed how young people respond to opportunities to help other people), one might not necessarily expect the results to perfectly replicate even in a well-powered experiment.

By seeking to replicate Fultz et al.'s (1986) experiment with as much fidelity as we could muster, we were able to test several key assumptions and substantive predictions of the empathy–altruism hypothesis, including (a) the assumption that Fultz et al.'s perspective-taking manipulation actually promotes empathy; (b) the assumption that their social evaluation manipulation actually increases people's concerns about being evaluated by others; (c) the prediction that perspective-taking increases helping—even when social evaluation pressure is low; and (d) the prediction that empathy

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

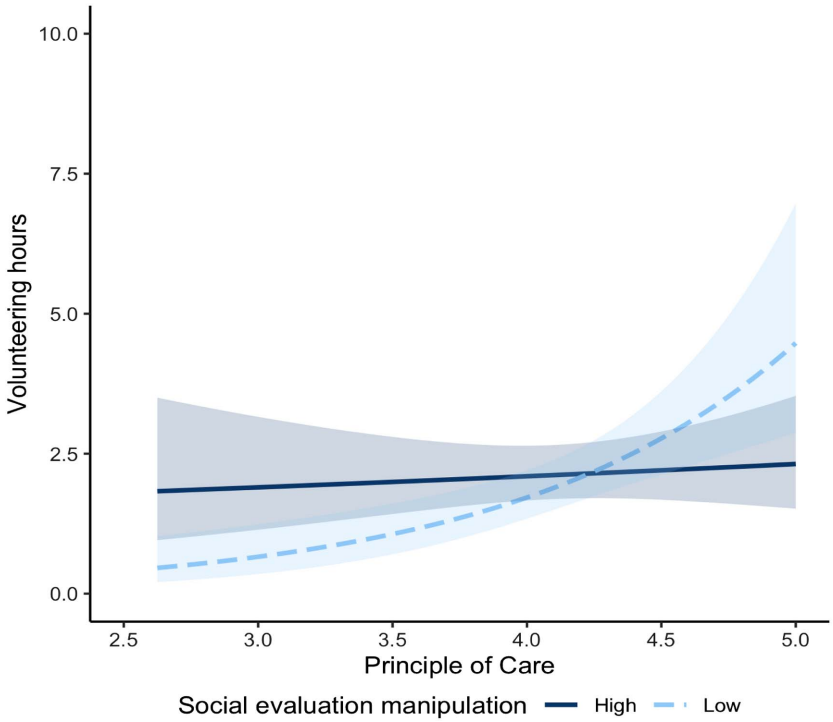
Predictor	*b* (*SE*)	*p*	Bayes factor
Moral identity symbolization	−0.11 (0.16)	.456	0.30
Social evaluation manipulation	−1.54 (0.74)	.036*	54.26
Moral Identity Symbolization × Social Evaluation Manipulation	0.47 (0.22)	.032*	62.43

Note. The *b*s for all models reported here are logged coefficients. For the terms associated with moral identity symbolization and the social evaluation manipulation, the Bayes factor reflects the strength of the evidence for the hypothesis that the effect is positive versus the evidence for the hypothesis that it is zero or negative. For the interaction of moral identity symbolization and the social evaluation manipulation, the Bayes factor reflects the strength of the evidence for the hypothesis that the coefficient is negative versus the evidence for the hypothesis that the coefficient is either zero or positive. *SE* = standard error.

* $p < .05$.

Figure 1

Two-Way Interaction of the Principle of Care and Social Evaluation



Note. The solid blue line shows the association of volunteering and the principle of care for subjects in the high social evaluation condition; the dashed blue line shows the same association for subjects 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. Lines are curved because they are plotted on the original scale of the volunteering responses (i.e., as the predicted probabilities of volunteering) rather than on a link scale (i.e., the log odds of volunteering). See the online article for the color version of this figure.

is associated with helping, even when controlling for the association of empathy and personal distress. We also used this opportunity to test whether three individual-difference variables that have been theorized to measure principlistic and egoistic prosocial motivations predict helping differently when subjects need to be concerned about social evaluation than when they do not. Finally, in response to an editorial suggestion, we looked for (and found) evidence that perspective-taking instructions have an indirect effect on helping via their intermediate effect on empathic concern (even though the direct and total effects of perspective-taking on helping were nonsignificant). In response to another editorial request, we looked for (and found) evidence that manipulated social evaluation pressure motivated subjects to volunteer but not evidence that perspective-taking instructions or the interaction of perspective-taking instructions and social evaluation pressure motivated subjects to volunteer.

Testing the Empathy–Altruism Hypothesis

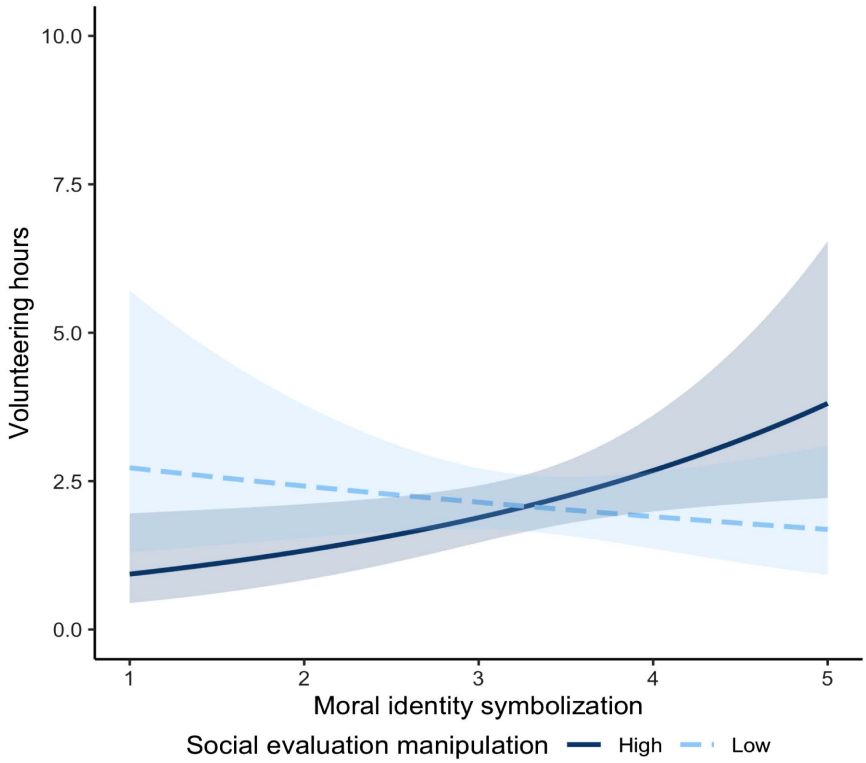
Considered together, we would call our efforts to replicate Fultz et al.'s (1986) findings a mixed success—and a partial victory for the empathy–altruism hypothesis. First, replicating Fultz et al. (1986), we did find that subjects we instructed to take the perspective of the other subject (who reported being lonely and desiring a friend) reported higher empathy toward the other subject than did subjects we instructed to remain objective (and that this effect led to a significant indirect effect of perspective-taking instructions on helping via empathy). Although the support we found for this prediction is unsurprising in light of recent meta-analytic evidence that the effect of perspective-taking instructions upon helping is driven primarily by the virtue of remain-objective instructions dampening feelings of empathy (McAuliffe et al., 2020), the effect was small in magnitude. Additionally, our Bayesian analysis indicated that the evidence that the effect is larger than zero was hardly overwhelming.

Second, our effort to replicate Fultz et al.'s (1986) evidence that their manipulation of social evaluation actually manipulated subjects' concerns about social evaluation was only a half-success: The manipulation convinced subjects that the other subject would not know whether they decided to volunteer, as Fultz et al. found. However, it did not convince subjects that their responses would be anonymous, presumably because they knew that the experimenter would eventually have access to their responses.

Although we did replicate Fultz et al.'s (1986) finding that self-reported empathy predicted the number of hours subjects volunteered to spend with the lonely subject—a finding entailed by the empathy–altruism hypothesis—we failed to replicate their finding that the perspective-taking manipulation increased the number of hours that

Figure 2

Two-Way Interaction of Moral Identity Symbolization and Social Evaluation



Note. The solid blue line shows the association of volunteering and moral identity symbolization for subjects in the high social evaluation condition; the dashed blue line shows the same association for subjects 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. Lines are curved because they are plotted on the original scale of the volunteering responses (i.e., as the predicted probabilities of volunteering) rather than being plotted on a link scale (i.e., the log odds of volunteering). See the online article for the color version of this figure.

subjects volunteered. Granted, we did find some evidence in an exploratory analysis that the perspective-taking manipulation had an indirect effect on helping via its intermediate effect on empathic concern, but the empathy–altruism hypothesis predicts a causal effect of empathy on helping, and our inability to demonstrate this effect by experimentally manipulating empathic concern with perspective-taking instructions is quite inconvenient for the hypothesis. To be sure, it is also the case that we 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 do so. Nevertheless, our failure to replicate one of the two findings most central to the empathy–altruism hypothesis is a cause for reflection. Other researchers have recently questioned about the robustness of this relationship more generally (Sassenrath et al., 2022).

Predictions About Individual Differences in Prosocial Motivation

We also took advantage of the fact that we experimentally manipulated social evaluation pressure to test predictions about three individual-difference variables hypothesized to predict prosocial behavior differently when social evaluation pressure is high than when it is low. First, we examined whether endorsement of the principle of care—the internalized moral value that one has a moral duty to help people in need (Wilhelm & Bekkers, 2010)—predicts helping even in the absence of social evaluation. Consistent with our preregistered prediction, we found that subjects with high scores on the Principle of Care Scale were more likely to volunteer, but only if they had been assigned to the low social evaluation condition. This pattern of findings suggests that endorsement of the principle of care may actually reflect the extent to which subjects hold moral principles about care: Those who scored high on this measure were more likely to help when social evaluation pressure was low. These results comport well with the theory that endorsement of the principle of care reflects a genuine, intrinsic motivation to improve the welfare of a needy person, even in the absence of social incentives and disincentives. On the other hand, the fact that the social evaluation manipulation did not produce a main effect on volunteering or unambiguously pass its manipulation checks makes it more difficult to interpret its interaction with the principle of care.

We also examined whether a measure of moral identity internalization (the centrality of moral traits to one's self-concept)

and a measure of moral identity symbolization (the desire to have one's moral traits observed in public settings; Aquino & Reed, 2002) predict helping differently when social evaluation pressure is low than when it is high. Internalization is hypothesized to promote helping even when social and material rewards for helping are absent, whereas symbolization is thought to promote helping precisely in order to secure such rewards. The evidence failed to support our preregistered prediction that moral identity internalization would be associated with prosocial behavior more strongly for subjects in the low social evaluation condition than for those in the high social evaluation condition. One possible explanation is that moral identity internalization does not actually measure factors that cause helping behavior. The meta-analytic association of moral identity internalization with observable prosocial behavior is modest (Hertz & Krettenauer, 2016), and though some self-report measures of prosocial traits do predict prosocial behavior, moral identity internalization may simply be a weaker predictor of helping behavior than other candidate measures of prosocial disposition. Of course, the fact that our prediction here was preregistered means that perhaps it merits more attention than some of the other published findings in the literature.

On the other hand, the evidence did support our prediction that subjects with high scores on moral identity *symbolization* would volunteer more, but only when assigned to the high social evaluation condition. Moral identity symbolization is hypothesized to promote helping behavior in public contexts, tracking the extrinsic rewards that follow from having one's moral character recognized by others (Winterich et al., 2013). Other work has shown that self-reported moral symbolization predicts public, but not private, prosocial behavior (Mayer et al., 2012), which suggests that the measure might indeed reflect people's desire to advertise themselves as good people.

Possible Explanations for Our Failures to Replicate

Why did we not perfectly replicate Fultz and colleagues' findings? Several explanations come to mind. In addition to the possibility that our failure to replicate was due to Fultz et al.'s (1986) small sample size, which necessarily limited the precision of their effect size estimates, and the psychological changes that young people might have experienced over the past 4 decades (some of which might have altered how they responded to the experimental conditions), it is also conceivable that our experimental manipulations were ineffective, that our lack of expertise with Fultz et al.'s methods limited our experiment's validity, and that the statistics we chose to use produced different results than Fultz and colleagues' statistical approach would have. We address these latter three possibilities in turn.

Ineffective Manipulations?

Our perspective-taking manipulation successfully increased empathy (relative to a remain-objective condition), so why did it not increase volunteering? A recent experiment 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). However, sending a supportive note in the course of an ongoing experiment is a relatively costless form of prosocial behavior, raising the possibility that perspective-taking instructions might be too weak to have a measurable effect on more costly helping behaviors, such as volunteering one's free time to help another person (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 manipulations of empathy.

Why did the social evaluation manipulation also 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 relevant. However, it could also be that subjects were indeed concerned with how their decisions could be judged and that we simply failed to manipulate assurance of anonymity. Indeed, although one of our manipulation checks revealed that subjects in the high social evaluation condition were more convinced that their communication partner would know if they chose not to volunteer, our second manipulation check indicated that subjects in the low social evaluation condition were not more convinced that their volunteering decision would remain anonymous. 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.

Other explanations are plausible as well. For instance, subjects in the low social evaluation condition might have failed to infer that the experimenter and the communicator were not in a position to negatively evaluate their decisions. That is, subjects in the low social evaluation condition might not have realized that neither the experimenter nor the other subject possessed all of the information necessary to condemn them for callousness. Alternatively, subjects might have been more worried about appearing unwilling to aid the advancement of scientific understanding than they were about appearing unconcerned about a fellow student's loneliness (Orne, 1962). If so, then the realization that someone would know about their decision not to participate in the follow-up study might have been sufficient to induce concern with social evaluation in both conditions.

Of course, these concerns are militated by the fact that two of the three individual-difference variables we examined—endorsement of the principle of care and moral symbolization—predicted volunteering in interaction with the social evaluation manipulation in ways that make sense only if the manipulation *did* successfully manipulate social evaluation pressure: Subjects endorsing the principle of care volunteered more of their time, but only if they were in the low social evaluation, and subjects scoring high on the moral symbolization scale volunteered more of their time—but only if they were in the high social evaluation condition. In short, the social evaluation manipulation might not have been as effective as we would have liked, but it might have been effective enough to create some of the psychological effects it was designed to create (unless the interactions were false positives).

Whether one accepts these post hoc interpretations, it remains the case that our social evaluation manipulation failed to change subjects' beliefs about whether their volunteering decision would be made anonymous. It is tempting to interpret this result as evidence that our experiment was based on a false premise—that we were not in a position to pit the empathy–altruism hypothesis against the empathy-specific evaluation hypothesis in the first place because we did not manipulate evaluation concerns clearly—but failures to

replicate manipulation checks should cause as much concern about the replicability of previously published findings as failures to replicate the substantive results. (For an enlightening example of a scholarly interchange on this issue, see Cheung et al., 2016; Finkel, 2016). Our failure to manipulate subjects' perceptions that their behavior would remain anonymous, therefore, should not alleviate worries about our failure to replicate Fultz et al.'s (1986) main result. Instead, in light of the large effect size that Fultz et al. (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%)—perhaps it should heighten them.

Experimental Expertise?

Some authors attribute unsuccessful replications to the replicators' inexperience (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 crucial ways. 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. (1986) obtained ($d = 1.07$; computed from Fultz et al.'s published results) and that McAuliffe et al. (2020) discovered in their 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 standard deviation units above the mean effect size of $d = .68$ that McAuliffe et al. (2020) reported: Our effect size here of $d = .31$ was nearly as far away (.37 standard deviation units), albeit in the opposite direction. Thus, while one might wonder why our perspective-taking manipulation here was relatively weak, one might also wonder why Fultz et al.'s manipulation was relatively strong.

Admittedly, we did not consult any of the authors from Fultz et al. (1986) when we were developing the protocol for our laboratory, and perhaps they would 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., 2022; Schlingloff et al., 2020). Moreover, we do have experience in conducting experiments on the empathy–altruism hypothesis in which we have successfully manipulated both empathy and fear of social evaluation (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 it 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.

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 et al., 2017). However, we also identified three potential shortcomings in Fultz et al.'s (1986) analysis strategy. First, they evaluated the effect of their experimental manipulations using analysis of variance, in effect treating volunteering as a continuous outcome. Treating count variables (e.g., the number of hours volunteered) as though they are continuous can substantially bias regression coefficients (O'Hara & Kotze, 2010), especially when the dependent variable is zero-inflated (e.g., if more subjects volunteered 0 hr than would be expected under distributional assumptions), 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 their analysis participants who expressed suspicion during the funnel debriefing. Insofar as suspicion is nonrandom (e.g., perhaps participants lower in trust are more likely to report suspicion), excluding suspicious participants reduces the generalizability of the results. Even more worryingly, excluding suspicious participants could interfere with the tendency for randomization to balance conditions on unobserved covariates—for example, if one condition were more likely to elicit suspicion than the other (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 statistically, 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 analytic issues deserve as much attention as small sample size when attempting to replicate past findings, given ever-evolving standards for data analysis. Notwithstanding the changes we made to improve the strength of our statistical analyses and causal inferences, we arrive at exactly the same conclusions as when we analyzed the data in the same way Fultz et al. did (see Supplemental Methods for details regarding the analysis). Thus, we cannot really attribute our failure to faithfully replicate all of their results to the data-analytic changes we made, either.

Constraints on Generality

Compared to Fultz et al. (1986), which included only female, nonsuspicious participants of unknown ethnicity, our study advanced the generality of inferences about empathy and altruism by sampling men, suspicious participants, and a heterogeneous sample of Black, White, Asian, and Hispanic students from U.S. universities on both sides of the country. Still, our sample comprised undergraduate students exclusively, most of whom were young adults, potentially limiting the generalizability of our findings. Although some research has examined how empathic concern varies across diverse cultures (e.g., Chopik et al., 2017), few attempts have been made to replicate the core predictions from the empathy–altruism hypothesis outside of Western, educated, industrialized, rich, and democratic cultures—and even fewer attempts have been made to test how concern about social evaluation might modulate empathic concern outside of the United States and Europe. Given the widespread cultural variance in reputational concern (Romano et al., 2021) and prosocial behavior directed toward strangers (Gächter et al., 2010), the association between empathic concern, social evaluation, and helping behaviors

ought to be more closely scrutinized in samples beyond the United States and Europe.

Finally, our experiment focused on how concern about social evaluation influences helping in the presence of empathy (which we manipulated by instructing half of our subjects to take the perspective of the other person while considering that person's plight and the other half of our subjects to remain objective while doing so), but future researchers interested in testing these two hypotheses jointly might also examine how concern about social evaluation affects helping even when subjects are not cued to adopt any particular psychological perspective. Although perspective-taking prompts increase empathy in a potential helper, they also might alert the helper that other people—including the experimenter—are aware of their opportunity to help (McAuliffe, 2017). In a no-instructions control condition, in contrast, the absence of a prompt ought to reduce participants' concern that others expect them to empathize with the target while spontaneously evincing similar amounts of empathy as imagine-other instructions (McAuliffe et al., 2020). Including a no-instructions condition might therefore provide an even more stringent test than the one to which we submitted it in the direct replication of Fultz et al. (1986) reported herein.

Conclusion

Our results sit uneasily next to Fultz et al.'s (1986) finding, grounded in Batson's empathy–altruism hypothesis, that empathy increases helping behavior, irrespective of the presence of an incentive to help in order to avoid appearing callous. In the work we report here, neither the manipulation of empathy nor the manipulation of fear of social evaluation significantly affected volunteering. Until further work reveals otherwise, we believe Fultz et al.'s (1986) major conclusion—that empathically aroused helping does not reflect a fear of negative social evaluation 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 lend some support to the hypothesis: Perspective-taking instructions promoted empathy, the social evaluation manipulation did induce the perception that the lonely person would know about subjects' choices, 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, all of which important are assumptions or substantive predictions that are entailed by the empathy–altruism hypothesis and conventional methods for testing it. Additionally, our findings that endorsement of the principle of care and concerns about advertising one's moral virtue predict helping in theoretically specified ways (i.e., when social evaluation pressure is low or high, respectively) should give some encouragement to researchers who seek to identify the individual differences that underwrite prosocial behavior. Fultz et al. (1986) are not unique in the empathy–altruism literature for featuring a relatively small sample and never having been subject to direct replication. Going forward, we hope researchers will consider our findings here as they continue to explore the motivational foundations of prosocial behavior.

References

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. https://doi.org/10.1177/0193841X211034363

Aquino, K., & Reed, A., II. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423

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. https://doi.org/10.1037/0022-3514.40.4.786

Aronson, E., Ellsworth, P. C., Carlsmith, J. M., & Gonzales, M. H. (1990). *Methods of research in social psychology*. McGraw-Hill.

Atkins, D. C., & Gallop, R. J. (2007). Rethinking how family researchers model infrequent outcomes: A tutorial on count regression and zero-inflated models. *Journal of Family Psychology*, *21*(4), 726–735. https://doi.org/10.1037/0893-3200.21.4.726

Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer*. Erlbaum.

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. https://doi.org/10.1037/0022-3514.73.3.517

Batson, C. D. (2011). *Altruism in humans*. Oxford University Press.

Batson, C. D. (2018). *A scientific search for altruism: Do we only care about ourselves?* Oxford University Press.

Batson, C. D. (2022). Prosocial motivation: A Lewinian approach. *Motivation Science*, *8*(1), 1–10. https://doi.org/10.1037/mot0000217

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), 922–933. https://doi.org/10.1037/0022-3514.56.6.922

Batson, C. D., Batson, J. G., Slingsby, J. K., Harrell, K. L., Peekna, H. M., & Todd, R. M. (1991). Empathic joy and the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, *61*(3), 413–426. https://doi.org/10.1037/0022-3514.61.3.413

Batson, C. D., Bolen, M. H., Cross, J. A., & Neuringer-Benefiel, H. E. (1986). Where is the altruism in the altruistic personality? *Journal of Personality and Social Psychology*, *50*(1), 212–220. https://doi.org/10.1037/0022-3514.50.1.212

Batson, C. D., Duncan, B., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation? *Journal of Personality and Social Psychology*, *40*(2), 290–302. https://doi.org/10.1037/0022-3514.40.2.290

Batson, C. D., Dyck, J. L., Brandt, J. R., Batson, J. G., Powell, A. L., McMaster, M. R., & Griffitt, C. (1988). Five studies testing two new egoistic alternatives to the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, *55*(1), 52–77. https://doi.org/10.1037/0022-3514.55.1.52

Batson, C. D., & Powell, A. A. (2003). Altruism and prosocial behavior. In T. Millon & M. J. Lerner (Eds.), *Handbook of psychology, volume 5: Personality and social psychology* (pp. 463–484). Wiley. https://doi.org/10.1002/0471264385.wei0519

Bekkers, R., & Ottoni-Wilhelm, M. (2016). Principle of care and giving to help people in need. *European Journal of Personality*, *30*(3), 240–257. https://doi.org/10.1002/per.2057

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., Carcedo, R. J., Carson, K. J., Cobb, R. J., Collins, N. L., Corretti, C. A., DiDonato, T. E., Ellithorpe, C., Fernández-Rouco, N., Fuglestad, P. T., … Yong, J. C. (2016). Registered replication report: Study 1 from Finkel, Rusbult, Kumashiro, & Hannon (2002). *Perspectives on Psychological Science*, *11*(5), 750–764. https://doi.org/10.1177/1745691616664694

Chopik, W. J., O'Brien, E., & Konrath, S. H. (2017). Differences in empathic concern and perspective taking across 63 countries. *Journal of Cross-Cultural Psychology*, *48*(1), 23–38. https://doi.org/10.1177/0022022116673910

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. https://doi.org/10.3389/fpsyg.2016.01639

de Waal, F. B. M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, *59*(1), 279–300. https://doi.org/10.1146/annurev.psych.59.103006.093625

Fabrigar, L. R., & Wegener, D. T. (2016). Conceptualizing and evaluating the replication of research results. *Journal of Experimental Social Psychology*, *66*, 68–80. https://doi.org/10.1016/j.jesp.2015.07.009

Fanelli, D., Costas, R., & Ioannidis, J. P. (2017). Meta-assessment of bias in science. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(14), 3714–3719. https://doi.org/10.1073/pnas.1618569114

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

Finkel, E. J. (2016). Reflections on the commitment–forgiveness registered replication report. *Perspectives on Psychological Science*, *11*(5), 765–767. https://doi.org/10.1177/1745691616664695

Fultz, J., Batson, C. D., Fortenbach, V. A., McCarthy, P. M., & Varney, L. L. (1986). Social evaluation and the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, *50*(4), 761–769. https://doi.org/10.1037/0022-3514.50.4.761

Gächter, S., Herrmann, B., & Thöni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1553), 2651–2661. https://doi.org/10.1098/rstb.2010.0135

Gilovich, T., Keltner, D., Chen, S., & Nisbett, R. E. (2018). *Social psychology* (5th ed.). W.W. Norton.

Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: An evolutionary analysis and empirical review. *Psychological Bulletin*, *136*(3), 351–374. https://doi.org/10.1037/a0018807

Hertz, S. G., & Krettenauer, T. (2016). Does moral identity effectively predict moral behavior? A meta-analysis. *Review of General Psychology*, *20*(2), 129–140. https://doi.org/10.1037/gpr0000062

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press. https://doi.org/10.1017/CBO9780511805851

Inzlicht, M., & Hutcherson, C. A. (2017). Psychology: People work less hard for others. *Nature Human Behaviour*, *1*(7), Article 0148. https://doi.org/10.1038/s41562-017-0148

Jiao, J., & Wang, J. (2018). Can lonely people behave morally? The joint influence of loneliness and empathy on moral identity. *Journal of Consumer Psychology*, *28*(4), 597–611. https://doi.org/10.1002/jcpy.1040

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Klein, R. A., Cook, C. L., Ebersole, C. R., Vitiello, C., Nosek, B. A., Hilgard, J., Ahn, P. H., Brady, A. J., Chartier, C. R., Christopherson, C. D., Clay, S., Collisson, B., Crawford, J. T., Cromar, R., Gardiner, G., Gosnell, C. L., Grahe, J., Hall, C., Howard, I., … Ratliff, K. A. (2022). Many Labs 4: Failure to replicate mortality salience effect with and without original author involvement. *Collabra: Psychology*, *8*(1), Article 35271. https://doi.org/10.1525/collabra.35271

Krebs, D. (1975). Empathy and altruism. *Journal of Personality and Social Psychology*, *32*(6), 1134–1146. https://doi.org/10.1037/0022-3514.32.6.1134

Lefebvre, J. P., & Krettenauer, T. (2019). Linking moral identity with moral emotions: A meta-analysis. *Review of General Psychology*, *23*(4), 444–457. https://doi.org/10.1177/1089268019880887

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. https://doi.org/10.5465/amj.2008.0276

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., Carter, E. C., Berhane, J., Snihur, A. C., & McCullough, M. E. (2020). Is empathy the default response to suffering? A meta-analytic evaluation of perspective-taking instructions' effects on empathic concern. *Personality and Social Psychology Review*, *24*(2), 141–162. https://doi.org/10.1177/1088868319887599

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. https://doi.org/10.1037/emo0000375

McNeish, D., & Stapleton, L. M. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research*, *51*(4), 495–518. https://doi.org/10.1080/00273171.2016.1167008

Mulder, J., Gu, X., Olsson-Collentine, A., Tomarken, A., Böing-Messing, F., Hoijtink, H., Meijerink, M., Williams, D. R., Menke, J., Fox, J.-P., Rosseel, Y., Wagenmakers, E.-J., & van Lissa, C. (2019). *BFpack: Flexible Bayes factor testing of scientific theories in R*. ArXiv. https://doi.org/10.48550/arXiv.1911.07728

Myers, D. G., & Twenge, J. M. (2022). *Social psychology* (14th ed.). McGraw Hill Education.

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), 510–516. https://doi.org/10.1037/0022-3514.73.3.510

Nowakowska, I. (2022). Volunteerism in the last year as a moderator between empathy and altruistic social value orientation: An exploratory study. *Current Directions in Personality Psychology*, *10*(1), 10–20. https://doi.org/10.5114/cipp.2021.108258

O'Hara, R., & Kotze, J. (2010). Do not log-transform count data. *Nature Precedings*, *1*, 118–122. https://doi.org/10.1038/npre.2010.4136.1

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776–783. https://doi.org/10.1037/h0043424

Preston, S. D. (2013). The origins of altruism in offspring care. *Psychological Bulletin*, *139*(6), 1305–1341. https://doi.org/10.1037/a0031755

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42.

Romano, A., Giardini, F., Columbus, S., de Kwaadsteniet, E. W., Kisfalusi, D., Triki, Z., Snijders, C., & Hagel, K. (2021). Reputation and socio-ecology in humans. *Philosophical Transactions of the Royal Society B*, *376*(1838), Article 20200295. https://doi.org/10.1098/rstb.2020.0295

Saslow, L. R., Willer, R., Feinberg, M., Piff, P. K., Clark, K., Keltner, D., & Saturn, S. R. (2013). My brother's keeper? Compassion predicts generosity more among less religious individuals. *Social Psychological and Personality Science*, *4*(1), 31–38. https://doi.org/10.1177/1948550612444137

Sassenrath, C., Vorauer, J. D., & Hodges, S. D. (2022). The link between perspective-taking and prosociality—Not as universal as you might think. *Current Opinion in Psychology*, *44*, 94–99. https://doi.org/10.1016/j.copsyc.2021.08.036

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), Article 191795. https://doi.org/10.1098/rsos.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. https://doi.org/10.1177/0956797616653102

Stotland, E. (1969). Exploratory investigations of empathy. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4, pp. 271–314). Academic Press.

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. https://doi.org/10.1016/j.evolhumbehav.2018.07.005

Twenge, J. M., Campbell, W. K., & Freeman, E. C. (2012). Generational differences in young adults' life goals, concern for others, and civic orientation, 1966–2009. *Journal of Personality and Social Psychology*, *102*(5), 1045–1062. https://www.apa.org/pubs/journals/releases/psp-102-5-1045.pdf

Twenge, J. M., Exline, J. J., Grubbs, J. B., Sastry, R., & Campbell, W. K. (2015). Generational and time period differences in American adolescents' religious orientation, 1966–2014. *PLOS ONE*, *10*(5), Article e0121454. https://doi.org/10.1371/journal.pone.0121454

Twenge, J. M., Spitzberg, B. H., & Campbell, W. K. (2019). Less in-person social interaction with peers among US adolescents in the 21st century and links to loneliness. *Journal of Social and Personal Relationships*, *36*(6), 1892–1913. https://doi.org/10.1177/0265407519836170

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

Wilhelm, M. O., & Bekkers, R. (2010). Helping behavior, dispositional empathic concern, and the principle of care. *Social Psychology Quarterly*, *73*(1), 11–32. https://doi.org/10.1177/0190272510361435

Winterich, K. P., Aquino, K., Mittal, V., & Swartz, R. (2013). When moral identity symbolization motivates prosocial behavior: The role of recognition and moral identity internalization. *Journal of Applied Psychology*, *98*(5), 759–770. https://doi.org/10.1037/a0033177

Yin, Y., & Wang, Y. (2023). Is empathy associated with more prosocial behaviour? A meta-analysis. *Asian Journal of Social Psychology*, *26*(1), 3–22. https://doi.org/10.1111/ajsp.12537

Zaki, J. (2020). Integrating empathy and interpersonal emotion regulation. *Annual Review of Psychology*, *71*(1), 517–540. https://doi.org/10.1146/annurev-psych-010419-050830

Received February 21, 2023
Revision received October 27, 2023
Accepted January 11, 2024 ■