UC San Diego
UC San Diego Previously Published Works

Title

Is Empathy the Default Response to Suffering? A Meta-Analytic Evaluation of Perspective Taking's Effect on Empathic Concern

Permalink

https://escholarship.org/uc/item/5j86s6m6

Journal

Personality and Social Psychology Review, 24(2)

ISSN

1088-8683

Authors

McAuliffe, William HB
Carter, Evan C
Berhane, Juliana
et al.

Publication Date

2020-05-01

DOI

10.1177/1088868319887599

Peer reviewed

Is Empathy the Default Response to Suffering? A Meta-Analytic Evaluation of Perspective Taking's Effect on Empathic Concern

William H. B. McAuliffe[1,2]* [ID], Evan C. Carter[3]*,
Juliana Berhane[1,4], Alexander C. Snihur[1,5],
and Michael E. McCullough[1,6]

Abstract

We conducted a series of meta-analytic tests on experiments in which participants read perspective-taking instructions—that is, written instructions to imagine a distressed persons' point of view ("imagine-self" and "imagine-other" instructions), or to inhibit such actions ("remain-objective" instructions)—and afterwards reported how much empathic concern they experienced upon learning about the distressed person. If people spontaneously empathize with others, then participants who receive remain-objective instructions should report less empathic concern than do participants in a "no-instructions" control condition; if people can deliberately increase how much empathic concern they experience, then imagine-self and imagine-other instructions should increase empathic concern relative to not receiving any instructions. Random-effects models revealed that remain-objective instructions reduced empathic concern, but "imagine" instructions did not significantly increase it. The results were robust to most corrections for bias. Our conclusions were not qualified by the study characteristics we examined, but most relevant moderators have not yet been thoroughly studied.

Keywords

Introduction

Perspective taking, the act of imagining the thoughts and feelings of others, is a common precursor to prosocial behavior (Batson, 2011). Researchers have also found that perspective taking causes empathic concern (an emotion that is congruent with and elicited by perceived suffering), which reflects altruistic motivation (i.e., a noninstrumental desire to improve the welfare of another person). But do people as a matter of course experience empathic concern for needy others they observe?

Here, we present a series of meta-analyses designed to address three questions: (a) Do people spontaneously empathize with those in distress? If so, then (b) could they experience even more empathic concern if they deliberately engaged in perspective taking? Finally, (c) do moderators—such as the identity of the victim or the medium by which participants learn about the victim's need—affect the extent to which people spontaneously empathize or successfully increase empathic concern via deliberate effort?[1] To assess to what extent our answers to these questions depend on specific assumptions about how publication bias affects the primary

literature, we compared the results from nine different estimators, each of which depend on either a different model of how publication bias works or how to best correct for it.

Extant theorizing is divided on whether people will experience empathic concern in response to distressed others in the normal course of experience. On one hand, much research suggests that people avoid empathic concern by default, at least when helping requires a considerable sacrifice (Cameron & Payne, 2011; Zaki, 2014). This tendency might explain why numerous tragedies—especially those involving large

[1]University of Miami, Coral Gables, FL, USA
[2]Harvard Medical School, Boston, MA, USA
[3]U.S. Army Research Laboratory, Albuquerque, NM, USA
[4]Bucknell University, Lewisburg, PA, USA
[5]Florida International University, Miami, USA
[6]University of California, San Diego, USA

*These authors contributed equally to this article.

Corresponding Author:
William H. B. McAuliffe, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115, USA.
Email: williamhbmcauliffe@gmail.com

numbers of people occurring far away—routinely fail to sustain bystanders' emotional attention (Loewenstein & Small, 2007; Slovic, Västfjäll, Erlandsson, & Gregory, 2017). On the other hand, empathizing may come more naturally in situations where the perceived costs of helping do not overwhelm how much observers value victims. For example, the experiments that established the relationship between empathic concern and altruistic motivation typically had participants learn about just one victim who is a fellow in-group member (Batson, 2011). Another set of experiments found that participants reported the same other-oriented thoughts and feelings upon deliberately trying to take the perspective of a single victim as when they just responded naturally. Thoughts that distracted from focusing on a victim's needs, such as thinking about her appearance rather than her plight, did not occur naturally. Rather, they were common only when participants deliberately attempted to *not* consider how the victim felt about her plight (Davis et al., 2004). Notably, participants who tried to emotionally distance themselves from the victim still reported strong other-oriented emotions, suggesting that they found it difficult to respond with apathy.

Even if people sometimes do spontaneously empathize with victims, it is nevertheless possible that they have untapped potential for how much empathic concern they could experience. For instance, multiple research groups have found that compassion training—which involves deliberately cultivating concern for others—increases helping of distressed groups relative to control trainings (Leiberg, Klimecki, & Singer, 2011; Weng et al., 2013). Given that neither research group intentionally recruited participants who were particularly low in trait empathic concern, the efficacy of compassion training implies that normally empathic people could, with effort, experience more empathic concern than they do by default.

The Present Study

Although our interests are in the causal antecedents of empathic concern, we drew on experiments that focused on the causal consequences of perspective taking. Specifically, beginning with Stotland (1969) and popularized by Toi and Batson (1982), researchers have frequently used so-called perspective-taking instructions to manipulate empathic concern toward a person in need or distress. Most often, participants assigned to a "high empathy" condition of a perspective-taking manipulation receive "imagine-other" instructions that ask them to imagine the thoughts and feelings of the distressed person. Although the exact instructions vary, the text from Batson, Early, and Salvarani (1997) is typical:

> While you are listening to the broadcast, try to *imagine how the person being interviewed feels about what has happened and how it was affected his or her life*. Try not to concern yourself with attending to all of the information presented. Just concentrate

on trying to imagine how the person interviewed in the broadcast feels. (p. 753; emphasis in original)

In some experiments, participants in the high-empathy condition instead receive "imagine-self" instructions that ask them to imagine how they *themselves* would feel were they experiencing the situation of the distressed person:

> While you are listening to the broadcast, try to *imagine how you yourself would feel if you were experiencing what has happened to the person being interviewed and how this experience would affect your life*. Try not to concern yourself with attending to all of the information presented. Just concentrate on trying to imagine how the person interviewed in the broadcast feels. (Batson, Early, & Salvarani, 1997, p. 753; emphasis in original)

Researchers then measure empathic concern (typically via self-report) to check that the manipulation had the intended effect. Batson, Early, and Salvarani (1997) found that both imagine-self and imagine-other instructions increase empathic concern relative to "remain-objective" instructions, which ask participants to actively refrain from considering the distress person's feelings:

> While you are listening to this broadcast, try to *be as objective as possible about what has happened to the person interviewed and how it has affected his or her life.* To remain objective, do not let yourself get caught up in imagining what this person has been through and how he or she feels as a result. Just try to remain objective and detached. (Batson, Early, & Salvarani, 1997, p. 753; emphasis in original)

Experiments that compare the effects of imagine instructions to remain-objective instructions speak to the reliability of perspective-taking instructions as a manipulation of empathic concern. However, the comparison between imagine-other or imagine-self instructions and remain-objective instructions is of less theoretical import because all three instructions all ask participants to *deliberately* engage in or inhibit perspective taking. Consequently, comparisons among these types of instructions do not reveal whether (a) imagine-other and imagine-self instructions upregulate default levels of empathic concern, (b) remain-objective instructions downregulate default levels of empathic concern, or (c) both. Assessing whether imagine-self and imagine-other instructions increase default levels of empathic concern is necessary for addressing whether people can deliberately upregulate empathic concern. Similarly, investigating whether remain-objective instructions decrease default levels of empathic concern would indicate whether people spontaneously empathize with others or remain detached (Batson, Eklund, Chermok, Hoyt, & Ortiz, 2007).

Fortunately, some experiments have included a control in which some participants receive either no instructions, instructions merely to behave as they normally would, or instructions that are irrelevant to perspective taking.[2] A recent preregistered experiment, for example, found that

participants who received no instructions ($n = 166$) reported experiencing high levels of empathy ($M = 5.67$, $SD = 1.07$, on a scale of 1 which meant "not at all" to 7 which meant "extremely"), which provides some evidence that they were spontaneously empathizing with the victim they were observing (McAuliffe, Forster, Philippe, & McCullough, 2018). Indeed, even participants in the remain-objective condition in this experiment were on average above the mid-point of the scale ($M = 5.07$, $SD = 1.28$). However, Blanton and Jaccard (2006) warned that positive scores on measures with arbitrary metrics do not necessarily reflect the presence of the construct in question. To clarify whether positive levels of empathic concern reported in control conditions are meaningful, we focused on whether participants who read remain-objective instructions reported less empathic concern. Whether participants in the remain-objective condition still experience some empathic concern is beyond the scope of the present study.

Moderators

Although many experiments have found that perspective-taking instructions create significant condition differences in empathic concern, assessing the reliability of perspective-taking instructions' effects is important in light of replicability issues in social psychology (Open Science Collaboration, 2015). The present literature is certainly not above suspicion, as many experiments that have used perspective-taking instructions have had modest sample sizes (e.g., see the experiments cited in Batson, 2011). Studies with small samples are prone to false positives (Loken & Gelman, 2017) and can create the illusion of robust effects if researchers and journal editors disfavor the publication of null effects (Ioannidis, 2005). Thus, we attempted to uncover as many relevant unpublished experiments as possible and included publication status as a possible moderator of perspective-taking instructions' effects on empathic concern.

A second moderator of interest is the in-group status of the perspective-taking target. Experiments have found that people experience more empathic concern on behalf of in-group members than out-group members (Stürmer, Snyder, Kropp, & Siem, 2006). In fact, the plight of an out-group member is more likely to elicit schadenfreude than empathic concern, if the out-group is perceived as in competition with the observer's in-group (Cikara, Bruneau, & Saxe, 2011). Thus, it is possible that imagine-self and imagine-other instructions are more effective when applied to out-group targets, while remain-objective instructions are more effective when applied to in-group targets.

The third moderator we focus on here is the medium by which participants learned about the perspective-taking target's plight. The socioemotional capacities underlying empathy evolved in an ancestral environment that was characterized by face-to-face interactions and long predated the invention of writing and recording devices (Goetz, Keltner, & Simon-Thomas, 2010). Thus, it is no surprise that empathic responses are sensitive to features of the victim that are most apparent in person, such as facial expressions that indicate fear and vocalizations that reflect distress (Hoffman, 2000; Marsh, 2019). It is plausible, then, that natural empathic responses are muted when merely reading about another person's plight, because facial and auditory cues of distress are absent. Hearing an audio recording of a victim provides auditory cues to which the cognitive systems underlying empathic concern may be sensitive, but lack the visual features that may heighten emotional impact. Based on this reasoning, one may hypothesize imagine-other and instruction-self instructions are particularly effective in boosting a muted empathic response to reading about a victim, where remain-objective instructions are particularly effective in dampening a strong empathic response to seeing or hearing a person in distress.

Method

Inclusion Criteria

We included results from all experiments that used two or more types of instructions to manipulate empathic concern, measured empathic concern toward the perspective-taking target, and used a perspective-taking target that was presented as real (for example, we included reactions to documentaries, but not to films that presented fictitious events) and in distress or need. We excluded three experiments in which the perspective-taking instructions pertained to the environment, as we were interested in empathic reactions to human victims. We also excluded experiments that used hypothetic vignettes (e.g., Takaku, 2001), as people often cannot accurately predict how they feel after experiencing an event that has not yet occurred (Wilson & Gilbert, 2005), especially when the event involves observing a victim's plight (Karmali, Kawakami, & Page-Gould, 2017; Pedersen, McAuliffe, & McCullough, 2018). We examined only studies that used imagine-other, imagine-self, remain-objective, or no-instructions (or instructions to behave normally) control conditions. We acknowledge that explicit instructions are not the only ways to modulate perspective taking (for instance, one might have participants write a story about a person in a photograph from either a first-person or third-person perspective; Galinsky & Moskowitz, 2000), but here we are interested in how attempts to upregulate or downregulate perspective taking relative to baseline influence empathic concern. Finally, many of the experiments included independent variables other than perspective-taking instructions. However, we recorded only the effect sizes of the main effects of the perspective-taking instructions.

For three reasons, we also decided to focus exclusively on papers that assessed empathic concern via self-report. First, self-report was by far the most popular method for assessing

empathic concern among potentially relevant articles we examined. Physiological, functional magnetic resonance imaging (fMRI), and facial measures were much less common. Second, self-report measures of currently felt emotions measures often have a strong link to behavior, whereas other measurement methods have less consistent associations with behavior (Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005). Third, self-reports can more easily measure specific emotions, whereas other measurement methods typically tap only global aspects of affect, such as valence and arousal (Mauss & Robinson, 2009). Precise differentiation is crucial here because emotions that appear similar to empathic concern have different motivational consequences—for instance, empathic *distress* generates a selfish desire to avoid further negative affect, which can be accomplished by either avoiding the distressed target or relieving her distress (Batson, Fultz, & Schoenrade, 1987; FeldmanHall, Dalgleish, Evans, & Mobbs, 2015). Because we are interested in whether people naturally take an interest in the welfare of victims, whether people spontaneously experience empathic distress is beyond the scope of the current study.

Almost all of the self-report measures in the present meta-analysis involved participants reporting to what extent they felt *compassionate, tender, sympathetic*, and other, similar emotions on rating scales (e.g., Batson, Early, & Salvarani, 1997). These emotion adjectives were typically mixed in among distracter adjectives and were averaged or factored to create an overall empathy score for each participant. Multiple research groups have found that such empathy scores are highly reliable and are psychometrically distinct from indices of distress and sadness (Batson et al., 1987; Fultz, Schaller, & Cialdini, 1988). The exception was Cameron and Payne (2011), who had participants continuously update how upset they felt on behalf of a suffering target for 60 seconds. These authors provided us with participants' average rating in each perspective-taking condition.

Data Collection

The search for articles began on May 24, 2015, and concluded on September 19, 2017. We completed an update of our dataset on February 6, 2019. We used several methods to collect all relevant published and unpublished experiments. First, we searched PsycINFO, Google Scholar, and ProQuest Dissertations & Theses: Social Sciences databases using the keywords "empath*," "perspective*," "observ*" (perspective-taking instructions are sometimes called "observational sets"), "sympath*," "compassion*," and "altruis*" (many of the relevant articles are about altruism). There were no restrictions on when or where the study was conducted, save that we were able to evaluate only those abstracts written in English (although in one case a researcher reached out to us with the results from an experiment he had published in Italian). There were also no restrictions on the population studied, other than that we only examined studies that used

adult samples. However, it turned out that all of the studies used subclinical populations (usually college students), and most studies were conducted in North America or Western Europe. Each abstract was reviewed for potential relevance, and the paper was read if there was no definitive information in the abstract about whether the article reported any experiments that fit the inclusion criteria.

We also searched through the online versions of every conference program of the Society for Personality and Social Psychology (SPSP) from 2003 (the first one available online) to 2019. We identified potentially relevant abstracts by sequentially entering each of the keywords above (minus the asterisk) into the search function. The authors of the abstract were then contacted to confirm that the study was relevant and to request the necessary data to code the effect sizes.

Last, we made requests for unpublished experiments using two methods. We posted on the SPSP listserv three times, describing the inclusion criteria for the meta-analysis and providing our contact information for those who had conducted relevant studies. We also e-mailed several authors directly that we knew had used perspective-taking instructions before. We asked them if they had any relevant unpublished experiments, and also asked them if they had any colleagues who may have relevant unpublished studies. Named colleagues were contacted with the same request.

Overall, we identified and coded 177 effect sizes from 85 papers that met the inclusion criteria (see the full reference list at https://osf.io/5htyg/files/). We initially deemed an additional 39 effect sizes as relevant, but determined upon closer inspection that they did not meet all criteria. There were also two relevant projects that we were aware of but could not access (Balliet & Gold, 2005; Veerbeek, 2005). Similarly, we were not able to obtain the information to code three experiments to which we did have access (Rumble, Van Lange, & Parks, 2010; Stotland, 1969; Sun, Li, Lou, & Lv, 2011). Last, because we found only three studies that compared imagine-self instructions to a no-instructions control condition, we did not conduct a meta-analysis for this comparison. See Table 1 for the magnitude and variance of each effect size, including the imagine-self versus no-instructions control comparison.

Effect Size Coding

We coded each effect size as Hedge's *g*, which is the bias-corrected standardized mean difference between two groups. Effect sizes from experiments that involved deception were all, to our knowledge, based on results from participants who were not suspicious of the ruse. We calculated *g* using the sample size, means, and standard deviations of each group when available, and used test statistics or *p* values when the means and standard deviations were not available (Borenstein, Hedges, Higgins, & Rothstein, 2009). We contacted authors of manuscripts that did not provide sufficient information to code the effect size, which usually happened in cases where

Table 1. Characteristics of Experiments Included in Meta-Analysis.

Comparison	Author(s)	Exp.	Year	Pub.	g	v	n1	n2
Imagine-Other vs. Remain-Objective								
	ToiB	1	1982	2	0.55	0.05	42	41
	BatsonA	1	2001	2	1.16	0.11	20	20
	BatsonA	1	1999	2	1.07	0.07	30	30
	BatsonA	2	1999	2	1.21	0.06	40	40
	BatsonB	3	1989	2	1.12	0.08	30	30
	BatsonB	1	1991	2	1.51	0.07	36	36
	BatsonB	2	1991	2	1.04	0.06	36	36
	BatsonB	3	1991	1	1.27	0.04	54	54
	BatsonB	1	1995	2	1.04	0.06	40	40
	BatsonC	1	2002	2	1.56	0.14	18	18
	FinlayS	0	2000	2	0.00	0.04	53	54
	BatsonD	5	1988	2	0.60	0.08	24	24
	BatsonE	1	1997	2	0.55	0.10	20	20
	BatsonEW	1	2007	2	0.86	0.05	40	40
	BatsonK	2	2007	2	1.48	0.20	12	12
	BatsonK	1	1995	2	1.06	0.11	20	20
	BatsonK	2	1995	2	0.86	0.07	30	30
	BatsonM	1	1999	2	2.11	0.15	20	20
	BatsonP	1	1997	2	1.04	0.05	48	48
	BatsonP	2	1997	2	1.23	0.10	24	22
	BatsonP	3	1997	2	0.98	0.07	30	30
	BatsonS	1	1997	2	0.91	0.11	20	20
	BatsonS	2	1997	2	1.26	0.08	30	30
	BatsonW	1	1996	2	1.38	0.08	30	30
	BatsonW	2	1996	2	0.74	0.14	15	15
	FultzEW	2	1986	2	1.05	0.14	16	16
	BatsonT	3	1995	2	1.06	0.11	20	20
	DovidioA	0	1990	2	0.53	0.02	96	96
	DovidioT	1	2004	2	0.00	0.09	22	22
	GrazianoH	3	2007	2	0.21	0.02	123	122
	ManerL	1	2002	2	0.96	0.03	73	72
	MashuriH	1	2012	2	0.31	0.02	92	85
	MyersL	1	2014	2	0.69	0.11	19	19
	MyersL	2	2014	2	0.69	0.05	39	38
	Oceja	1	2008	2	1.37	0.08	30	30
	Oceja	2	2008	2	1.36	0.05	46	46
	OcejaJ	1	2007	2	0.51	0.16	12	12
	Pedersen	1	2012	0	0.36	0.04	53	53
	SchallerC	1	1988	2	0.63	0.05	45	45
	SchroederD	1	1988	2	0.65	0.03	60	60
	SibickyS	1	1995	2	0.69	0.05	42	42
	StocksL	1	2009	2	1.07	0.09	24	24
	StocksL	2	2009	2	0.65	0.09	24	24
	StocksL	2	2011	2	0.99	0.13	16	16
	Van Lange	1	2008	2	−0.33	0.07	27	27
	VescioS	1	2003	2	0.50	0.06	32	32
	VorauerS	1	2009	2	0.83	0.05	47	46
	OcejaH	1	2014	2	0.99	0.07	32	32
	OcejaH	2	2014	2	1.54	0.07	36	34
	OcejaH	3	2014	2	0.79	0.11	20	20
	López-PérezC	2	2014	2	1.92	0.14	20	20
	Buswell	2	2005	0	0.65	0.02	104	103

(Continued)

Table I. (Continued)

Comparison	Author(s)	Exp.	Year	Pub.	*g*	*v*	*n*I	*n*2
	Harrell	I	2006	2	0.64	0.02	92	91
	Jacobs	I	2011	0	0.29	0.02	249	183
	Mann	I	2010	0	0.63	0.03	62	62
	Poulin	2	2015	I	0.32	0.02	102	103
	Poulin	3	2015	I	0.46	0.04	57	56
	Poulin	4	2015	I	0.64	0.05	44	44
	OcejaS	2	2017	2	0.51	0.02	106	107
	AmbronaO	I	2017	2	0.54	0.05	42	41
	YagiO	0	2015	I	0.17	0.06	31	31
	LishnerS	0	2015	0	0.13	0.06	36	31
	LishnerR	0	2015	0	0.51	0.19	10	10
	LishnerF	I	2015	I	1.16	0.11	18	36
	LishnerW	0	2015	0	0.65	0.09	36	18
	LishnerB	0	2015	0	3.58	0.91	7	4
	LishnerF	2	2015	0	1.42	0.50	4	4
	LishnerM	0	2015	0	0.52	0.03	81	56
	LishnerD	0	2015	0	0.33	0.06	21	83
	LishnerC	0	2015	0	0.61	0.11	13	27
	LishnerR	0	2015	0	0.55	0.07	26	28
	LishnerW	0	2015	0	0.84	0.14	15	15
	Lishner	0	2015	0	0.85	0.11	42	12
	LishnerM	0	2015	0	0.33	0.06	30	32
	Betancourt	I	1990	2	0.97	0.03	75	75
	Betancourt	2	1990	2	0.96	0.08	28	28
	Davis	I	1983	2	0.50	0.03	84	74
	CialdiniS	I	1987	2	0.43	0.06	59	22
	CialdiniS	2	1987	2	0.92	0.12	17	17
	MironM	I	2017	2	0.87	0.06	35	37
	MironM	2	2015	I	0.35	0.03	63	61
	Habashi	2	2008	2	0.19	0.01	170	170
	OcejaA	2	2010	2	1.11	0.07	32	32
	Pagotto	I	2010	0	0.61	0.04	53	53
	Pagotto	2	2010	0	0.86	0.04	61	56
	Pagotto	3	2010	0	0.44	0.04	59	55
	McAuliffeF	0	2018	2	0.69	0.01	153	169
	VociP	0	2009	2	0.64	0.10	20	20
	HabashiG	2	2016	2	0.14	0.03	77	74
	McAuliffe	0	2017	I	0.15	0.02	101	108
	Harmon-JonesV	0	2004	2	0.47	0.06	37	36
	BekkersO	I	2015	0	0.13	0.02	125	105
	Faulkner	0	2018	2	0.27	0.02	120	120
	BuffoneP	0	2017	2	0.21	0.03	64	71
	López-PérezH	0	2017	2	0.62	0.03	70	70
	SmithK	0	1989	2	0.36	0.06	32	32
	SassenrathP	I	2017	2	0.73	0.07	30	29
	SassenrathP	2	2017	2	0.49	0.04	59	57
	CornishG	0	2018	2	0.46	0.02	102	102
	WondraM	I	2018	I	0.51	0.03	67	65
	WondraM	2	2018	I	1.04	0.03	65	65
	WondraM	3	2018	I	0.35	0.03	71	70
	PfattheicherS	2a	2019	2	0.59	0.04	53	52
	PfattheicherS	2b	2019	2	0.49	0.02	85	86
	PfattheicherS	3a	2019	2	0.37	0.03	66	67

Table 1. (continued)

Comparison	Author(s)	Exp.	Year	Pub.	*g*	*v*	*n*1	*n*2
	PfattheicherS	3b	2019	2	0.49	0.03	64	63
	PfattheicherS	4	2019	2	0.66	0.03	64	63
Imagine-Other vs. Imagine-Self								
	LishnerR	0	2015	0	0.27	0.08	26	26
	LishnerW	0	2015	0	0.10	0.13	15	15
	LishnerF	2	2015	0	0.35	0.35	6	4
	LishnerF	1	2015	1	0.49	0.09	21	21
	BatsonE	0	1997	2	−0.36	0.10	20	20
	López-PérezC	2	2014	2	0.84	0.10	20	20
	StocksL	2	2011	2	0.33	0.12	16	16
	MyersL	1	2014	2	−0.19	0.10	19	19
	MyersL	2	2014	2	−0.06	0.05	39	38
	FinlayS	1	2000	2	0.00	0.06	36	35
	LammB	1	2007	2	0.72	0.24	8	8
	Pagotto	1	2010	0	0.00	0.04	53	60
	Pagotto	2	2010	0	0.32	0.03	61	62
	Pagotto	3	2010	0	−0.11	0.03	59	56
	McAuliffeF	0	2018	2	0.12	0.01	153	169
	VociP	0	2009	2	0.66	0.10	20	20
	BuffoneP	0	2017	2	0.15	0.03	64	67
	BeussinkH	0	2017	2	0.03	0.02	97	103
	MinisteroP	1	2018	2	0.05	0.02	82	79
Imagine-Other vs. No Instructions								
	MironM	1	2017	2	0.15	0.06	35	34
	Mann	1	2010	0	0.15	0.03	62	61
	DovidioT	1	2004	2	0.00	0.09	22	22
	HepperH	2	2014	2	0.10	0.04	49	46
	Pedersen	1	2012	0	−0.12	0.04	53	53
	LishnerR	0	2015	0	0.30	0.19	10	9
	LishnerB	0	2015	0	0.72	0.24	7	9
	LishnerA	0	2015	1	−0.12	0.02	94	95
	McAuliffeF	0	2018	2	0.19	0.01	153	166
	BeussinkH	0	2017	2	0.06	0.02	96	97
	LeongC	0	2015	2	0.40	0.03	63	63
	DrweckiM	2	2011	2	0.26	0.07	31	29
	DrweckiM	3	2011	2	−0.08	0.1	19	21
	WondraM	1	2018	1	0.17	0.03	67	65
	WondraM	2	2018	1	0.30	0.03	65	65
	WondraM	3	2018	1	−0.03	0.03	71	70
	MinisteroP	1	2018	2	−0.17	0.03	82	75
	MinisteroP	2	2018	2	−0.27	0.03	54	101
No Instructions vs. Remain-Objective								
	LishnerB	0	2015	0	0.29	0.03	73	71
	Pedersen	1	2012	0	0.52	0.04	53	53
	DovidioT	1	2004	2	0.00	0.09	22	22
	LishnerR	0	2015	0	0.12	0.19	9	10
	Mann	1	2010	0	0.52	0.03	61	62
	LishnerS	0	2015	0	0.38	0.07	32	29
	LishnerB	0	2015	0	3.11	0.68	9	4
	MironM	1	2017	2	0.66	0.06	34	37
	CameronP	3	2011	2	0.18	0.04	58	54
	McAuliffeF	0	2018	2	0.51	0.01	166	169
	WondraM	1	2018	1	0.33	0.03	71	65

(continued)

Table 1. (continued)

Comparison	Author(s)	Exp.	Year	Pub.	*g*	*v*	*n*1	*n*2
	WondraM	2	2018	1	0.71	0.03	69	65
	WondraM	3	2018	1	0.40	0.03	73	70
Imagine-Self vs. Remain-Objective								
	LishnerF	0	2015	1	0.55	0.10	21	21
	BatsonE	0	1997	2	0.84	0.10	20	20
	MyersL	1	2014	2	1.65	0.14	19	19
	MyersL	2	2014	2	0.76	0.06	38	38
	BladerR	1	2014	2	0.81	0.04	49	49
	FinlayS	1	2000	2	0.00	0.05	35	35
	StocksL	2	2011	2	0.65	0.13	16	16
	López-PérezC	2	2014	2	1.44	0.12	20	20
	LishnerF	2	2015	0	0.74	0.37	4	6
	LishnerW	0	2015	0	0.69	0.13	15	15
	LishnerR	0	2015	0	0.27	0.07	26	28
	Pagotto	1	2010	0	0.56	0.04	60	53
	Pagotto	2	2010	0	0.51	0.03	62	56
	Pagotto	3	2010	2	0.51	0.04	56	55
	McAuliffeF	0	2018	2	0.55	0.01	169	169
	VociP	0	2009	2	−0.17	0.10	20	20
	BuffoneP	0	2017	2	0.06	0.03	67	71
Imagine-Self vs. No Instructions								
	McAuliffeF	0	2018	2	0.06	0.01	169	166
	BeussinkH	0	2017	2	0.08	0.02	96	103
	MinisteroP	1	2018	2	−0.21	0.03	75	79

authors focused on simple effects. When authors did not report the sample sizes of each group, we assumed equal numbers of participants in each group for even sample sizes, and placed the remainder in the experimental group for odd sample sizes. Some experiments had more than two perspective-taking conditions; we coded an effect size for each pair of conditions. For instance, an experiment that contained imagine-other, imagine-self, and remain-objective instructions (but not a no-instructions control condition) would have yielded three effect sizes, each of which would appear in a different meta-analytic sample.

There were two papers that reported nonsignificant effects, but the information to calculate effect sizes was not reported and the authors no longer had access to the original data. We coded these effect sizes as 0 to avoid exacerbating overestimation due to the censoring of nonsignificant effects. Although imputing zeroes risks underestimation insofar as nonsignificant effect sizes may represent positive effects that the studies simply did not have power to detect, it is an unbiased compromise because the nonsignificant effects may have also represented *negative* effect sizes that could not be statistically distinguished from zero. To check whether our decision to impute zeroes affected any of our conclusions, we re-ran all analyses excluding nonsignificant effect sizes that we did not have enough information to compute. Including the imputed zeroes did not qualitatively affect any results except where noted.

Coding Experiment Attributes

Initially, we coded the following properties of each study: (a) publication status, (b) the type of perspective-taking target, and (c) the medium by which the participant learned about the perspective-taking target. Publication status was coded as an ordinal variable (0 = unpublished and not currently being prepared for submission, $k = 42$; 1 = in preparation for submission, $k = 19$; 2 = published or in press, $k = 116$). We collapsed published and in-preparation studies together for moderation analyses. The type of perspective-taking target was coded as an ordinal variable (0 = Not salient out-group member, $k = 128$; 1 = Salient out-group member, $k = 41$; information not provided or the effect was coded collapsing across in-group and out-group targets, $k = 8$). Out-group membership was coded only when authors made it explicit that they intended for participants to perceive the perspective-taking target as an out-group member. Because people regard in-group members as heterogeneous (Mullen & Hu, 1989), we did not code for out-group membership in cases where the target incidentally had a characteristic that differed from the majority of participants. We do not know whether participants on average encoded incidentally different targets as in-group members or instead did not encode their group membership at all. Therefore, for the primary analyses we regarded the group membership moderator as primarily indicating the absence or presence of a salient out-group

member. As a robustness check, we re-ran the mixed-effect models such that incidentally different victims were also coded as out-group members. We defined "incidentally different victims" as perspective-taking targets who had observable characteristics (or were described as having characteristics) that (a) contrasted with most or all members of the population from which the study sampled, but (b) did not have those characteristics to affect participants' beliefs about whether the target was an in-group or out-group member. Examples of studies with incidentally different targets include undergraduates learning about a student from a different, nonrival university and online participants learning about an elderly man. How we coded the type of perspective-taking target did not qualitatively affect results except where noted. The medium by which participants learned about the target was coded as a nominal variable ($0 =$ written report or description, $k = 102$; $1 =$ audio recording, $k = 50$; $2 =$ film or photographs, $k = 20$; information not provided [coded as missing], $k = 5$).

While revising this article, we considered suggestions from referees to code for dispositional factors that theoretically should impact the efficacy of perspective-taking instructions. For instance, there is some evidence that "imagine" instructions increase empathic concern more among participants who are low in traits that promote prosocial behavior (e.g., agreeableness, trait empathic concern; Graziano, Habashi, Sheese, & Tobin, 2007; McAuliffe, Forster, et al., 2018; but see Bekkers, Ottoni-Wilhelm, & Verkaik, 2015). We did not code for personality traits in this meta-analysis because the vast majority of studies that met our inclusion criteria did not explore how the effect of perspective-taking instructions are moderated by dispositional factors.

We did, however, act on a suggestion to code the gender composition in each study, as this information was available in almost all cases. A secondary analysis of some of the studies included in the present meta-analysis suggest women on average experience more empathic concern than do men in response to needy others, and that this difference is not due to violations of measurement invariance (McAuliffe, Pedersen, & McCullough, 2018). Thus, one might expect that imagine-other and imagine-self instructions would increase empathic concern more among men than women, and the remain-objective instructions would have a bigger impact on women than men. Although at least two investigations found that gender did not impact the effect of perspective-taking instructions on empathic concern (Batson et al., 1999; Smith, Keating, & Stotland, 1989), they likely did not have enough statistical power to detect interaction effects (Ns < 100).

Reliability

We used the compute.es package to code effect sizes (Del Re, 2010). The third and fourth authors coded effect sizes collected through 2017 and categorized them according to the pair of instructions used. Reliabilities were excellent for

effect sizes and sample sizes ($ICC_g = 0.98$, $ICC_{N1} = 0.96$, $ICC_{N2} = 0.97$). However, reliability was poor for effect size variances ($ICC_v = 0.26$) due to a few large discrepancies (raw exact agreement was 82%, and several disagreements were of a trivially small magnitude). The first author resolved discrepancies after the reliabilities were computed. He also recomputed several variances to ensure that there was no systematic coding issue that accounted for the low interrater reliability. There was one case in which the coders agreed on the effect size and variance, but the values were implausibly large. The first author corrected these values after determining that the authors of the study had reported the standard errors of empathic concern in each condition as standard deviations. The third author wrote a brief description of the perspective-taking target, and the medium by which the participant heard about the perspective-taking target. The first author coded these descriptions for in-group status and communication medium, respectively. The first author coded for publication status and performed all coding duties in the 2019 update.

Analyses

We conducted all analyses using R 3.1.2 (R Core Team, 2013). In addition to the base package, we used several other packages (see below). The data and syntax to perform the analyses can both be found at https://osf.io/5htyg/files/.

We fit nine meta-analytic models for each of the five comparisons of pairs of perspective-taking instructions: Imagine-other versus imagine-self, imagine-other versus no-instructions, imagine-other versus remain-objective, imagine-self versus remain-objective, and no-instructions versus remain-objective. First, we fit the random-effects meta-analysis model in the metafor package (Viechtbauer, 2010). This method assumes that the individual studies estimate "true" effects drawn from a distribution of "true" effects. As such, the random-effects model estimates the mean (μ) of the distribution of "true" effects as well as the variance (τ^2; also known as between-study heterogeneity). A test statistic, Q, represents whether the variability in a sample of effect sizes is greater than what can be expected given within-study sampling error, and can therefore be used to calculate a p value for whether between-study heterogeneity is greater than zero. One can explicitly model between-study heterogeneity as a function of study-level moderators (i.e., a mixed-effects model). In this case, the interpretation would be that the true effect measured by a given study is determined by some study-level characteristic such as the methodology or the location where data were collected. For the mixed-effects model, the test statistic Q_e can be used to test whether residual heterogeneity exists after including moderators. We focused on the random-effects estimate and tested the effects of the moderators regardless of whether Q was statistically significant (Borenstein et al., 2009; Higgins & Thompson, 2002). We used restricted maximum-likelihood estimation to calculate between-study variance.

We also used several other meta-analytic techniques that methodologists have designed to be robust to the influence of artifacts such as publication bias to estimate the true underlying effect. There is no consensus regarding which of these methods is best in all situations. To avoid making an arbitrary decision about which estimator to rely on, we followed Carter, Schönbrodt, Gervais, and Hilgard (2019) by applying each estimator and examining the degree to which our findings changed based on which estimator was used. We implemented these methods as described in Carter et al. (2019).

First, we used the trim-and-fill technique (Duval & Tweedie, 2000) as implemented in the metafor package. The trim-and-fill technique models publication bias as "missing studies" to be imputed. It determines which studies are missing by assuming that a funnel plot—a plot of effect sizes against sample sizes (or some proxy thereof)—will be symmetric. The procedure first "trims" the studies from the more populous side of the plot to find the "true" center of the plot. Then the trimmed effects are put back and "missing" effect sizes are imputed to the less populous side of the funnel plot so that it mirrors the more populous side. A new effect size is then estimated from this modified dataset.

Next, we used the Weighted Average of Adequately Powered Studies (WAAP; Ioannidis, Stanley, & Doucouliagos, 2017; Stanley, Doucouliagos, & Ioannidis, 2017) method. This method involves first fitting an intercept-only weighted least squares (WLS) regression—where weights are the inverse variances of the effect size estimates—to get a meta-analytic estimate of the true underlying effect. A second intercept-only WLS regression is then fit to only those studies that show post hoc power of 80% or more to detect the estimate of the true underlying effect from the previous step. The logic behind this method is that meta-analyzing only data from studies that are adequately powered removes some of the potential influence from processes like publication bias because those processes tend to disproportionately affect low-powered studies.

The third method we applied was the Precision Effect Test (PET; Stanley & Doucouliagos, 2014). PET, like WAAP, is a WLS regression model with inverse-variance weights; however, for PET, the effect size estimates are regressed on their standard errors. The intercept of PET is a meta-analytic estimate of the true underlying effect for which any association between effect size and standard error has been statistically controlled. The logic behind this method is that it controls for small-study effects—that is, processes that result in smaller samples tending to show larger effects—the most worrisome of which is publication bias.

Next, we applied the Precision Effect Estimate with Standard Error (PEESE; Stanley & Doucouliagos, 2014). This method is the same as PET, but rather than regressing effect sizes on standard errors, it regresses them on variances (i.e., standard error squared). Essentially, this change allows for a nonlinear version of small-study effects. The fifth method we applied is the conditional combination of PET

and PEESE: PET-PEESE (Stanley & Doucouliagos, 2014). Through a simulation study, Stanley and Doucouliagos (2014) showed that PET should be preferred to PEESE when the true underlying effect is zero. However, when the true underlying effect is not zero, PEESE is the preferred method. By using the statistical significance of the PET estimate (with a one-tailed test) as a best guess for whether the true underlying effect is zero, we can choose whether to trust the PET estimate or the PEESE estimate (using a two-tailed test in either case) more. Thus, the PET-PEESE estimate is the same as that of PET when PET is statistically nonsignificant; otherwise, it is the same as that of PEESE.

Next, we applied the *p*-curve method for meta-analytic estimation (Simonsohn, Nelson, & Simmons, 2014). A *p*-curve is a distribution of the statistically significant *p* values in a dataset. Because the *p* values are solely a function of their effect sizes and sample sizes, it can be used to estimate the effect size most likely to have produced it, even without access to unpublished results. Thus, given the *p* values and sample sizes, the overall effect size can be inferred by minimizing the discrepancy between the expected *p*-curve given the sample sizes and effect sizes and the observed distribution of *p* values. Critically, when a *p*-curve is used to estimate the "true" effect size in this manner, it is designed to produce a corrected estimate of the average effect size of the studies submitted to it. Thus, it is not estimating the average of the distribution of true effects (μ) in the same way as all of the other methods we apply. Furthermore, the implementation of *p*-curve that we use here does not provide a confidence interval.

The seventh method we applied is the *p*-uniform technique (van Assen, van Aert, & Wicherts, 2015). This method is similar to *p*-curve save that *p*-uniform uses the Irwin-Hall estimation method and sets the estimate to zero if the average *p* value of the dataset is greater than .025. We used the puniform package (van Aert, 2018) to implement *p*-uniform.

The final method we applied was the three-parameter selection model first proposed by Iyengar and Greenhouse (1988) and recently discussed by McShane, Böckenholt, and Hansen (2016). Like the random-effects model described above, this model includes a parameter for both the average and the variance of the distribution of true underlying effects. The third parameter represents the probability that a nonsignificant finding enters the meta-analytic dataset. We fit this model using the default functions in the weightr package (Coburn & Vevea, 2017).

Results

In each of our five datasets, we treated the instruction type that we hypothesized would evince more empathic concern as the experimental group. The instruction type that we hypothesized would evince less empathic concern was treated as the control group. (We hypothesized that imagine-other instructions might yield more empathic concern than

Table 2. Parameter Estimates for Random-Effects Models.

Comparison	k	g	Q (p)	T	I^2
Imagine-Other vs. Imagine-Self	19	0.11 [0.02, 0.21]	20.21 (.321)	0.00 [0.00, 0.34]	0.00 [0.00, 72.51]
Imagine-Other vs. No Instructions	18	0.08 [−0.02, 0.17]	9.52 (.574)	0.09 [0.00, 0.23]	18.02 [0.00, 61.35]
Imagine-Other vs. Remain-Objective	107	0.68 [0.61, 0.76]	335.17 (<.001)	0.31 [0.27, 0.42]	68.41 [61.89, 80.27]
Imagine-Self vs. Remain-Objective	17	0.56 [0.37, 0.75]	39.00 (.001)	0.30 [0.16, 0.62]	64.28 [32.38, 88.19]
No Instructions vs. Remain-Objective	13	0.45 [0.35, 0.55]	17.53 (.063)	0.00 [0.00, 1.04]	0.01 [0.00, 96.65]

imagine-self instructions, as participants who imagine themselves suffering may focus more on their own distress than on the perspective-taking target; Batson, Early, & Salvarani, 1997.) We began our analyses by generating a random-effects estimate for each sample, examining the distributions of study-level sample sizes and statistical power, and then observing the effect of moderators (publication status and in-group status) on the estimates of the true underlying effect. Next, we generated several meta-analytic estimates based on the trim-and-fill, WAAP-WLS, PET, PEESE, PET-PEESE, *p*-curve, *p*-uniform, and the three-parameter selection models.

Random-Effects Meta-Analysis

The results from the random-effects models appear in Table 2. Imagine-other instructions yielded statistically significantly more empathic concern than imagine-self instructions (*g* = 0.12, 95% CI = [0.02, 0.21], *p* = .013). Although the effect size was small, this result is consistent with our speculation that imagine-other instructions manipulate empathic concern in a more straightforward manner than do imagine-self instructions. Similarly, comparing imagine-other instructions to remain-objective instructions yielded a slightly larger mean difference (*g* = 0.68, 95% CI = [0.61, 0.76], *p* < .001) than did the comparison between imagine-self and remain-objective instructions (*g* = 0.56, 95% CI = [0.37, 0.75], *p* < .001). However, the remain-objective instructions appear to be primarily responsible for these effects, as they reduced empathic concern relative to a no-instructions control by almost half of a standard deviation (*g* = 0.45, 95% CI = [0.35, 0.55], *p* <.001). In contrast, the uptick in empathic concern among participants who received imagine-other instructions relative to participants who received no instructions was nonsignificant (*g* = .08, 95% CI = [−0.02, 0.17], *p* = .117). Note that all of these results are from a completely uncorrected estimator, and therefore are overestimates to the extent that bias is present in the datasets.

The imagine-other versus imagine-self, imagine-other versus no-instructions, and no-instructions versus remain-objective instructions comparisons did not have statistically significant heterogeneity. Although these findings could mean that the effects are not moderated by any study-level factor, it is more likely that the test for heterogeneity was underpowered in these samples because they contained a small number of effect size estimates. The fact that the I^2 statistic, which quantifies the percentage of variation in effect sizes that is attributable to substantive differences between studies, had extremely large confidence intervals in these samples reflects this uncertainty in the true amount of between-study heterogeneity. It is possible, of course, that heterogeneity is driven by one or two effect sizes (see Figure 1) having a large influence on relatively small meta-analytic datasets.

The imagine-self versus remain-objective comparison did have statistically significant heterogeneity. The point estimate of between-study variance here were "substantial" (i.e., 50% < I^2 < 80%; Higgins & Thompson, 2002), but the estimates here too were so uncertain that the confidence intervals ranged from no or mild heterogeneity (e.g., ≤ 42.61%) to extreme heterogeneity (e.g., > 80%). Only the imagine-other versus remain-objective comparison, which was based on a much larger number of studies than any of the other comparisons, had substantial, statistically significant heterogeneity accompanied by a reasonable confidence interval (68.42%, 95% CI: [62.87%, 80.27%]). We view the uncertain but potentially large amount of heterogeneity in our samples as ample justification for examining the effects of moderators on the random-effects estimates.

Mixed-Effects Meta-Analysis

We added publication status (0 = unpublished, 1 = in preparation or published), in-group status (0 = not salient out-group target, 1 = salient out-group target), and information medium (two dummy codes: 0 = written, 1 = audio; 0 = written, 1 = visual) as covariates to each of our random-effects models (see Table 3). Only two moderators had a statistically significant effect on the magnitude of effect size. First, out-group status for the imagine-other versus no-instructions comparison changed the sign of the effect (*b* = −0.36, 95% CI = [−0.68, –0.05], *p* = .024). This finding yields the intriguing implication that deliberate perspective taking could *reduce* empathic concern for out-group members. However, this effect was not replicated in the imagine-other versus remain-objective comparison, which had more power. Moreover, this effect was nonsignificant when regarding all potential out-group targets as perceived out-group targets (*b* = −0.20, 95% CI = [−0.47, 0.06], *p* = .135). Thus, we are inclined to view the negative effect of out-group status on
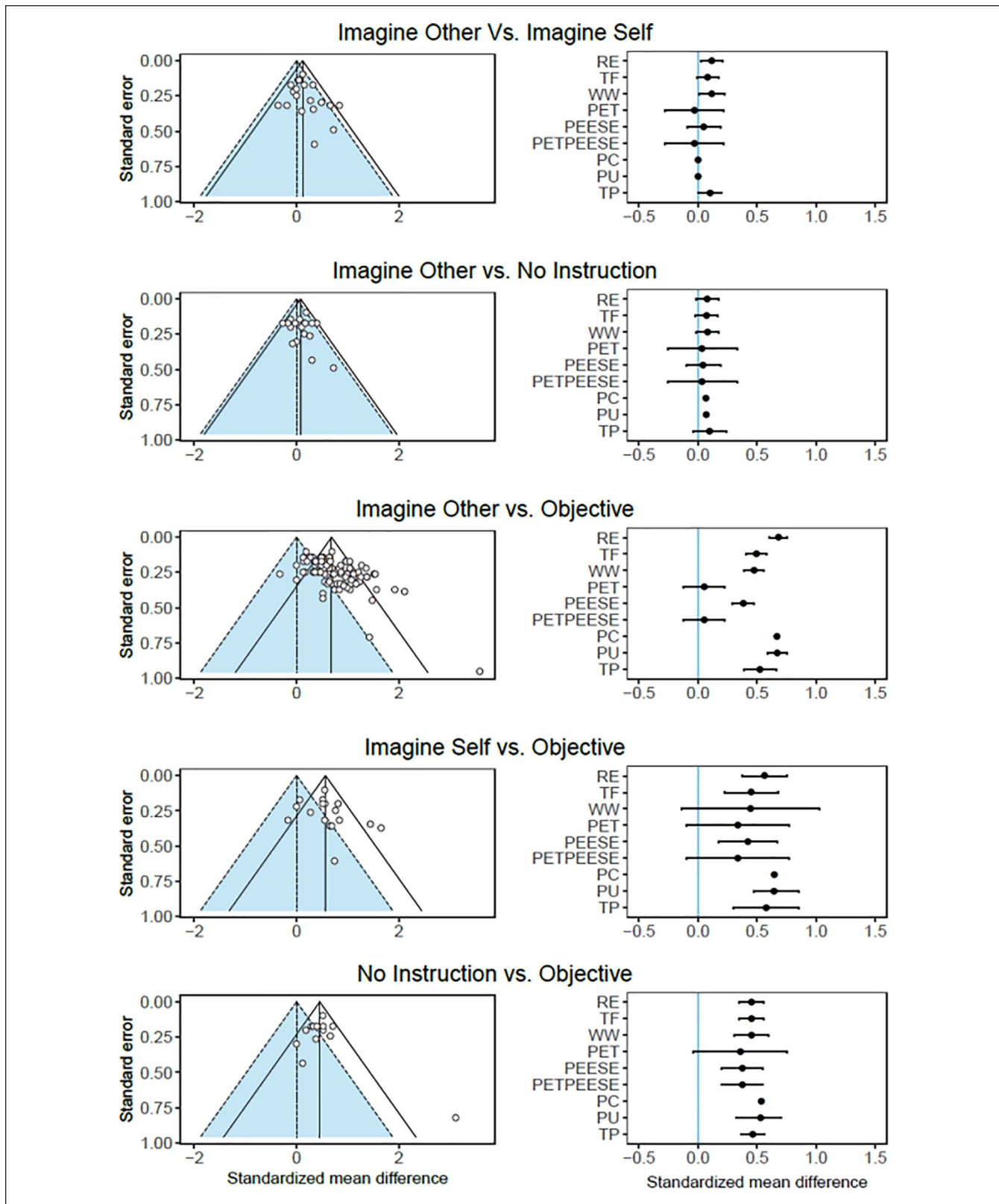
Figure 1. Funnel plots and forest plots from all meta-analytic estimators.
Note. Error bars = 95% confidence intervals. RE = random-effects; TF = trim-and-fill; WW = weighted least squares–weighted average of adequately powered studies; PET = Precision Effect Test; PEESE = Precision Estimation Estimate With Standard Error; PC = *p*-curve; PU = *p*-uniform; TP = three-parameter selection model.

Table 3. Mixed-Effect Models.

Comparisons	Moderators	Parameter estimates			Residual heterogeneity	
		b	95% CI	*p*	Q_e	*p*
Imagine-Other vs. Imagine-Self	Intercept	.19	[−.13, 0.50]	.250	15.49	.278
	Publication	.05	[−.28, 0.38]	.760		
	Out-Group	−.10	[−.42, 0.22]	.532		
	Visual	−.03	[−.45, 0.39]	.901		
	Audio	−.27	[−.57, 0.03]	.077		
Imagine-Other vs. No Instructions	Intercept	.19	[−.07, 0.45]	.161	10.30	.504
	Publication	.01	[−.25, 0.27]	.938		
	Out-Group	−.36	[−.34, −.05]	.024		
	Visual	.04	[−.26, 0.35]	.771		
	Audio	−.16	[−.36, 0.03]	.105		
Imagine-Other vs. Remain-Objective	Intercept	.53	[.35, 0.72]	<.001	287.89	<.001
	Publication	.14	[−.05, 0.34]	.150		
	Out-Group	−.03	[−.21, 0.15]	.781		
	Visual	−.13	[−.38, 0.11]	.268		
	Audio	.18	[.01, 0.35]	.042		
Imagine-Self vs. Remain-Objective	Intercept	.67	[.25, 1.10]	.002	28.45	.004
	Publication	−.12	[−.57, 0.33]	.598		
	Out-Group	−.29	[−.72, 0.14]	.191		
	Visual	NA	NA	NA		
	Audio	.47	[−.10, 1.03]	.106		
No Instructions vs. Remain-Objective	Intercept	.44	[.20, 0.68]	<.001	17.03	.018
	Publication	.04	[−.24, 0.31]	.784		
	Out-Group	−.10	[−.47, 0.27]	.588		
	Visual	−.38	[−1.07, 0.31]	.284		
	Audio	.07	[−.24, 0.38]	.659		

Note. CI = confidence interval.

imagine-other instructions as a Type I error. Second, the effect of imagine-other instructions versus remain-objective instructions was larger when participants received information about the victim aurally rather than reading about his or her plight. One possible explanation of this effect is that aural information triggers more empathic concern than does written information, which increases the scope for remain-objective instructions to reduce empathic concern. The effect of hearing about the victim's plight was not replicated in the no-instructions versus remain-objective comparison, however, perhaps because of its lower statistical power. In addition, the effect was only marginally significant ($b = 0.14$, 95% CI = [−0.03, 0.31], $p = .100$) when we excluded effect sizes that we had coded as zero because we did not have access to enough data to code the actual effect sizes. Overall, the moderators we coded for do not clearly account for variation in effect sizes across studies. Given the evidence that some of the effect size variation in the imagine-self versus remain-objective and imagine-other versus remain-objective datasets is systematic rather than random, we assume that there are moderating factors for which we did not code.

Because including gender composition as a moderator was unplanned and we could not code the gender composition of several unpublished effect sizes by the same author (and missing values on any variable results in listwise deletion from the entire model), we conducted this moderator analysis separately. Gender composition had a significant positive effect in the imagine-other versus remain-objective dataset, $b = .01$, 95% CI = [0.00, 0.01], $p = .002$, suggesting that samples with more women had larger effect sizes. However, the effect of gender composition was nil in the imagine-other versus imagine-self dataset ($b = .00$, 95% CI = [−0.00, 0.01], $p = .791$) and nonsignificantly negative in the imagine-other versus no-instructions dataset ($b = −.00$, 95% CI = [−0.01, 0.01], $p = .545$), imagine-self versus remain-objective dataset ($b = −.00$, 95% CI = [−0.01, 0.01], $p = .688$), and no-instructions versus remain-objective dataset ($b = −.00$, 95% CI = [−0.02, 0.01], $p = .329$). Thus, we were not able to confirm our hypotheses that samples with more women would show larger effects of remain-objective instructions and samples with relatively more men would show larger effects of imagine-other and imagine-self instructions. Possibly, the null findings were due to low statistical power in these smaller datasets. Alternatively, the significant finding in the imagine-other versus remain-objective dataset may have been a false positive. Indeed, in the

Table 4. Study-Level Power and Sample Sizes.

Comparison	Total sample size quantiles (25%, 50%, 75%)	Statistical power quantiles (25%, 50%, 75%)	k (prop.) power > 0.80	k (prop.) p < .05	TES
Imagine-Other vs. Imagine-Self	39, 52, 119	0.06, 0.07, 0.10	0 (0)	2 (.11)	.485
Imagine-Other vs. No Instructions	62, 125, 152	0.06, 0.07, 0.08	0 (0)	1 (.06)	.738
Imagine-Other vs. Remain-Objective	51, 77, 127	0.52, 0.70, 0.91	43 (.40)	90 (.84)	<.001
Imagine-Self vs. Remain-Objective	40, 54, 111	0.36, 0.47, 0.78	3 (.18)	9 (.53)	.617
No Instructions vs. Remain-Objective	61, 112, 136	0.41, 0.66, 0.74	1 (.08)	7 (.54)	.703

Note. TES = Test for Excessive Significance.

absence of a reason to anticipate that the effect of gender composition would matter more for one type of instructions than another, one could argue that the opposing effects of gender on imagine instructions and remain-objective instructions would cancel out.

Sample Size, Study-Level Power, and Excess Significance

Several surveys of meta-analyses in fields such as neuroscience (Button et al., 2013), economics (Ioannidis et al., 2017; Stanley et al., 2017), and psychology (Stanley, Carter, & Doucouliagos, 2018) indicate that individual studies tend to have surprisingly low statistical power. This is an important issue because results from underpowered studies are less likely to reflect the true underlying effect that they purportedly measure. We assessed the study-level statistical power in each of our five datasets by calculating each study's statistical power to detect the WLS estimate from the first step of the WAAP estimator (Stanley et al., 2018). Table 4 displays quantiles of the distributions of both study-level sample sizes and statistical power for each dataset.

Table 4 also shows the number and proportion of studies in each dataset that were adequately powered (i.e., power ≥ 0.80) and the number and proportion of studies that were statistically significant. These two values can be compared in the sense that the average power of a meta-analytic dataset is an estimate of the number of statistically significant effects that should be expected. If the observed number of significant studies is higher than what would be expected given the average power, this may indicate the influence of bias. One way to test for statistical significance produced by biased reporting of underpowered studies is the Test for Excessive Significance (TES; Ioannidis & Trikalinos, 2007), which involves comparing the expected number of significant results (i.e., the average power) to the observed number of significant results using a binomial test. The *p* values from this test can be interpreted as the probability that there is *not* an excess of significant effects—that is, smaller *p* values are more consistent with bias. This *p* value is shown in the final column of Table 4.

As can be seen in Table 4, the study level sample sizes in each dataset tend to have similar interquartile ranges (IQR ≈

40 to 130). The datasets differ the most in their median sample sizes, with the imagine-other versus no-instructions and no-instructions versus remain-objective datasets having a median sample size of more than 100, and the imagine-other versus imagine-self and imagine-self versus remain-objective datasets having a median sample size of approximately 50. The datasets also differed in their statistical power: The imagine-other versus imagine-self and imagine-other versus no-instruction datasets have lower power (IQR ≈ 0.06 to 0.11), the imagine-self versus remain-objective and no-instruction versus remain-objective datasets cluster together with somewhat higher power (IQR ≈ 0.34 to 0.80), and the imagine-other versus remain-objective dataset stands alone with the highest power (IQR = [0.52, 0.91]). Only the imagine-other versus remain-objective dataset has a substantial number of adequately powered studies; however, as indicated by TES, the higher power in this dataset is still less than the relatively large number of significant effect sizes. This result is consistent with the presence of bias in the imagine-other versus remain-objective dataset.

Meta-Analytic Estimators

The point estimates and confidence intervals from each of the nine estimators for each of the five datasets are displayed in Figure 1 and Table 5. Notably, in four of the five datasets (all but imagine-other vs. remain-objective), the point estimates are fairly consistent. For example, the variances in the point estimates for the imagine-other versus imagine-self, imagine-other versus no-instructions, imagine-self versus remain-objective, and no-instructions versus remain-objective data are 0.004, 0.001, 0.014, and 0.003, respectively. In contrast, at 0.062, the variance in the point estimates for the imagine-other versus remain-objective data is six to nearly 60 times larger. Furthermore, when examining results from the methods that produced confidence intervals (see Figure 1 and Table 5), there was substantial overlap in confidence intervals for each method in all but the imagine-other versus remain-objective data. Thus, even though the imagine-other versus remain-objective comparison was based on much more data than the other comparisons, there is more uncertainty about to what extent selective reporting played a role in our random-effects estimate of its effect size.

Table 5. The Point Estimates and 95% Confidence Intervals for Each Estimator for Each Dataset.

	Imagine-other vs. imagine-self	Imagine-other vs. no instructions	Imagine-other vs. remain-objective	Imagine-self vs. remain-objective	No instructions vs. remain-objective
RE	.12 [.02, .21]	.08 [−.02, .17]	.68 [.61, .76]	.56 [.37, .75]	.45 [.35, .55]
TF	.08 [.00, .17]	.07 [−.03, .17]	.49 [.40, .58]	.45 [.23, .68]	.45 [.35, .55]
WW	.12 [.01, .22]	.08 [−.02, .18]	.47 [.39, .56]	.44 [−.14, 1.03]	.45 [.30, .60]
PET	−.03 [−.28, .22]	.03 [−.26, .33]	.05 [−.12, .23]	.34 [−.10, .77]	.36 [−.04, .76]
PEESE	.05 [−.09, .19]	.04 [−.10, .19]	.38 [.29, .48]	.42 [.17, .67]	.38 [.20, .55]
PP	−.03 [−.28, .22]	.03 [−.26, .33]	.05 [−.12, .23]	.34 [−.10, .77]	.38 [.20, .55]
PC	.00 [NA, NA]	.07 [NA, NA]	.67 [NA, NA]	.65 [NA, NA]	.53 [NA, NA]
PU	.00 [NA, NA]	.07 [−1.65, .71]	.67 [.59, .75]	.64 [.48, .85]	.53 [.32, .70]
TP	.10 [.00, .20]	.10 [−.04, .24]	.52 [.39, .66]	.58 [.30, .85]	.47 [.37, .56]

Note. RE = Random-effects; TF = trim-and-fill; WW = weighted least squares–weighted average of adequately powered studies; PET = Precision Effect Test; PEESE = Precision Estimation Estimate With Standard Error; PP = Precision Effect Test–Precision Estimation Estimate With Standard Error; PC = *p*-curve; PU = *p*-uniform; TP = three-parameter selection model.

Three general patterns of point estimates emerged across the five datasets. First, the effect size estimates for both the imagine-other versus imagine-self and imagine-other versus no-instructions datasets tend to cluster around zero. For example, the maximum distance of an estimate from zero for both datasets is 0.12. Second, for both the imagine-self versus remain-objective and no-instructions versus remain-objective datasets, estimates tend to cluster around a medium to large effect size—the range of estimates goes from 0.34 to 0.65. The third and final pattern is a lack of agreement in the estimates for the imagine-other versus remain-objective comparison. As mentioned, the variance in these point estimates is the highest among all of the datasets—the range goes from 0.05 to 0.68—and there is the least amount of overlap in the 95% confidence intervals.

The PET estimates are outliers in that they yielded the lowest values and were statistically nonsignificant for all five datasets. Nearly the same pattern holds for the estimates from the conditional estimator PET-PEESE, save for the no-instructions versus remain-objective dataset. However, when excluding nonsignificant effect sizes we coded as zero for lack of better information, PET-PEESE was significant for the imagine-self versus remain-objective comparison (*b* = 0.45, 95% CI = [0.21, 0.70]), but nonsignificant for the no-instructions versus remain-objective dataset (*b* = 0.30, 95% CI = [−0.09, 0.70]). At the opposite extreme, the random-effects estimates are statistically significant in all but the imagine-other versus no-instruction dataset. Thus, our results are consistent with both the pessimistic conclusion that perspective-taking instructions have no significant effect on empathic concern, and the optimistic conclusion that they—or at least remain-objective instructions—have a medium-to-large effect on empathic concern.

Interestingly, the *p*-curve, *p*-uniform, and the three-parameter selection models all produced higher estimates than the uncorrected random-effects model in some datasets, implying that somehow bias led to the underestimation of the true underlying effect. These three estimators are similar in that they all assume an explicit selection function relating the probability of a study based on whether the *p* value exceeds the typical cutoff of *p* < .05 for the primary effect of interest for publication. This model of selective reporting may not apply to the present meta-analysis, as the effect of perspective-taking instructions on empathic concern served as a manipulation check rather than the primary outcome for most of the included studies. Thus, the *p*-curve, *p*-uniform, and the three-parameter selection model may have yielded nonsensical corrections in some cases because they are based on an incorrect model of how publication bias affected the included studies. Research into whether psychology experiments remain unpublished because they yielded nonsignificant manipulation checks inform whether *p*-curve, *p*-uniform, and the three-parameter selection model are appropriate for meta-analyses of manipulation checks.

Discussion

The present study reported a series of meta-analytic tests to examine whether participants who receive instructions to imagine the feelings of a distressed victim experience more empathic concern than do participants who receive no instructions or who receive instructions to remain objective when learning about the victim's circumstances. We examined five meta-analytic datasets, each of which compared a unique pair of instructions. We applied nine meta-analytic estimators to each of these five datasets, which differ primarily in terms of how each protects from the influence of bias. Although many simulation studies have been conducted to compare these estimators, the general conclusion is that no estimator is reliably better than all others in all cases (Carter et al., 2019). As a result, we try to avoid conclusions that rely solely on a single estimator. We summarize our tentative conclusions based on the preponderance of the evidence in Table 6.

Our first conclusion is that the imagine-other versus imagine-self and the imagine-other versus no-instruction

Table 6. Tentative, *Qualitative* Conclusions About Each Meta-Analytic Dataset.

Comparison	Mean effect size?	Heterogeneity?	Cause for concern?
Imagine-Other vs. Imagine-Self	$g \approx 0$	Uncertain.	None
Imagine-Other vs. No Instructions	$g \approx 0$	Uncertain.	None
Imagine-Other vs. Remain-Objective	Medium	Substantial.	Disagreement in estimators. Too many significant studies.
Imagine-Self vs. Remain-Objective	Medium	Substantial, with some uncertainty. Unexplained.	None
No Instructions vs. Remain-Objective	Medium	Uncertain.	None

datasets reveal point estimates that are indistinguishable from zero. For both datasets, nearly all estimators converged on this conclusion. The exceptions in each case returned values that were very close to zero.

Second, the imagine-self versus remain-objective and no-instructions versus remain-objective datasets both yielded medium-sized effects that are distinguishable from zero. In both datasets, estimates that failed to reach statistical significance were still nonzero, which is a different pattern than in the other three datasets where nonsignificant estimates are also nearly zero. Furthermore, the number of statistically significant findings is not dramatically different from what would be expected given the average statistical power in these two datasets (see the TES results in Table 4). Overall, this pattern makes us willing to believe the results converged on by the trim-and-fill estimator, PEESE, *p*-curve, *p*-uniform, and the three-parameter selection model—that is, a medium-sized true underlying effect (following Cohen's qualitative description scheme; Cohen, 1988). However, as mentioned above, *p*-curve, *p*-uniform, and the three-parameter selection model may be misspecified for these data, so this conclusion should not be held too confidently.

Third, we found several indications that the imagine-other versus remain-objective dataset is biased upward: (a) Both PET and PET-PEESE show a near-zero estimate that is statistically nonsignificant, (b) the numerous studies with large effects created funnel plot asymmetry, (c) and there were more statistically significant studies than would be expected given the dataset's average power. What remains to be determined is whether overestimation due to bias is so strong that these data reflect a true underlying effect of zero. This conclusion relies on believing that PET-PEESE is more appropriate for these data than any other estimator that corrects for bias.

One can also try to draw inferences for the imagine-other versus remain-objective dataset by comparing it to the other four datasets. Recall that imagine-other instructions did not seem to increase empathic concern above and beyond imagine-self or no instructions (i.e., g ≈ 0). Also, in comparison with remain-objective instructions, both imagine-self instructions and the no-instructions conditions increased empathic concern (i.e., a medium-sized true underlying effect). Because imagine-self and no-instructions conditions show a

level of empathic concern that is indistinguishable from the level induced by imagine-other instructions, and because remain-objective instructions seem to evince less empathic concern than imagine-self instructions and no-instructions conditions, it follows that remain-objective instructions should create less empathic concern than imagine-other instructions (i.e., $g > 0$). This logic is more convincing to us than the argument that PET and PET-PEESE are the only estimators performing adequately for the imagine-other versus remain-objective comparison, especially because PET is particularly aggressive in correcting for publication bias (Stanley et al., 2017). Therefore, we think it is likely that the true underlying effect in the imagine-other versus remain-objective dataset is similar to the true underlying effect in the two other datasets that make use of a remain-objective condition (i.e., a medium-sized true underlying effect). This conclusion dovetails with the only preregistered experiment focused on the questions under investigation in the present meta-analysis (McAuliffe, Forster, et al., 2018). We also provisionally conclude that estimators like the random-effects model overestimate the effect of imagine-other versus remain-objective instructions due to bias, and that *p*-curve, *p*-uniform, and the three-parameter selection model possibly overestimate it due to misspecification.

Between-study heterogeneity was clearly present only in datasets that involved remain-objective instructions (see Table 2). However, we are unable to draw firm conclusions about how much heterogeneity exists. First, in the no-instructions versus remain-objective dataset, any heterogeneity is clearly due to a single data point that is both the smallest study and the largest effect size in all of the data we examine here. It would seem prudent to ignore this outlier-driven result (which was only marginally significant anyhow). Second, although the nonzero heterogeneity in the imagine-self versus remain-objective dataset is not due to a single data point, it is neither explained by the moderators we tested nor clearly estimated in terms of magnitude (Table 3). These uncertain findings likely reflect the fact that the dataset for the imagine-self versus remain-objective comparison is small. Finally, the imagine-other versus remain-objective dataset has a much more precise estimate of between-study heterogeneity than in either of the other datasets. However, the heterogeneity estimate itself may be

influenced by the bias that we believe exists in this sample. Given how common heterogeneity is in psychology research (Stanley et al., 2018; van Erp, Verhagen, Grasman, & Wagenmakers, 2017), it seems likely that systematic influences are responsible for some of the between-study variation we observed. Nevertheless, future work is needed to determine which factors moderate the influence of remain-objective instructions on empathic concern.

With these conclusions in hand, we can provide tentative answers to the three questions we posed in the introduction: Do people spontaneously empathize with those in distress? *Yes, empathic concern may be the most typical inclination of people who have paid attention to the plight of a needy person.* Can they experience even more empathic concern if they deliberately engage in perspective taking? *No, encouraging people to engage in perspective taking might not make them much more concerned about others than they are when given no instructions at all.* And do moderators such as the identity of the victim affect the extent to which people spontaneously empathize or successfully increase empathic concern via deliberate effort? *There is no consistent evidence of moderation.* We now turn to the theoretical implications of each of our answers.

Implications

People spontaneously empathize with those in distress. Even enthusiastic supporters of the idea that empathic concern reflects altruistic motivation (the "empathy-altruism hypothesis") have been reluctant to make definitive claims about the prevalence of altruistic motivation outside of the laboratory (Batson, 2011; Eisenberg & Miller, 1987), as any one kind act in an uncontrolled setting could be due to either altruistic motivation or simply following egoistic incentives (avoiding appearing callous, avoiding negative affect, etc.). But the fact that instructions to remain objective are responsible for condition differences in empathic concern also implies that its motivational consequence—altruistic motivation to relieve another's need—is not merely a phenomenon that can be conjured up in the laboratory but is in fact responsible for everyday expressions of compassion and helpfulness. That is, when people hear or observe distressing news on their newsfeed, on television, while listening to the radio, or in conversation with a close friend, they do not receive instructions to take the perspective of the victim but nevertheless are likely to experience empathic concern. Although the presence of empathic concern does not guarantee that helping will follow (Eisenberg & Miller, 1987; McAuliffe, Forster, et al., 2018), we can infer that need-based helping that does occur in everyday life was likely preceded by empathic concern. The historical tendency to prioritize self-interested explanations for helping over altruistic explanations until proven otherwise (Sober & Wilson, 1998) therefore loses its empirical basis: It is implausible that altruistic motivation plays no role in most

need-based helping acts, given that its emotional manifestation is typically present. Of course, this conclusion depends on the validity of the research supporting the empathy-altruism hypothesis and would be weakened if the empirical foundation for the empathy-altruism hypothesis was ever undermined.

Although the spontaneous nature of empathic concern may paint a sanguine picture of human motivation, it is important to note that empathic concern is not an unmitigated good. Feigenson (1997) warns that fostering more empathic concern among judges and juries than they already naturally feel would compromise the impartial mind-set that is required to enforce justice. For example, Bornstein (1998) found that mock jurors felt more sympathy for severely harmed plaintiffs than mildly harmed plaintiffs, and as a result were more likely to deem the defendant liable for the harm, even though the presented evidence did not differ across conditions. The present results imply that stakeholders in the legal system could reduce the bias induced by empathic concern toward one party over another by encouraging decision-makings to remain objective. Support for this prediction comes from an experiment which found that participants who received imagine-other instructions were more likely to unfairly benefit a needy person after learning the details of her plight than participants who received remain-objective instructions (Batson, Klein, Highberger, & Shaw, 1995).

At least two process models are consistent with our finding that people feel empathic concern in the absence of instructions to take the perspective of others. First, authors who posit that certain components of empathic responding are engaged automatically (de Waal, 2008; Drayton, Santos, & Baskin-Sommers, 2018) would argue that "imagine" instructions are redundant because perceiving the suffering of others induces congruent emotions without any deliberate effort. A second possibility is that people routinely make an effort to take the perspective of suffering others, even when they have not received "imagine" instructions. This hypothesis is consistent with evidence that cognitive load reduces core components of empathy (Morelli & Lieberman, 2013).

These two alternatives are not mutually exclusive because the empathy system comprises several brain systems (Morelli & Lieberman, 2013) and empathy might happen automatically in some situations but only effortfully in others (Schumann, Zaki, & Dweck, 2014). Even so, the two process models have different implications for how the present findings relate to the broader literature on empathy. In particular, the automaticity of empathic concern could explain why people avoid situations that they believe would cause empathic concern, which they associate with the expenditure of time, energy, and financial resources (Cameron, Harris, & Payne, 2016; Cameron et al., 2019; Cameron & Payne, 2011; Shaw, Batson, & Todd, 1994; Zaki, 2014). If people can foresee that exposure to suffering would automatically impel them to expend resources that they currently do not wish to

forgo, then they could try to avoid exposing themselves to the types of situations that the experiments in the current meta-analysis required them to experience. In essence, the present findings would verify people's lay-belief that learning about a suffering person will induce empathic concern and all of its associated costs.

Last, the fact that inhibiting perspective taking reduced empathic concern may elucidate positive correlations among two of empathy's key dimensions. Davis (1983) conceptualized individual differences in empathy as falling under four dimensions: perspective-taking (PT), empathic concern (EC), personal distress (PD), and fantasy (FS). A recent psychometric investigation of Davis's four-dimensional Interpersonal Reactivity Index (IRI) questionnaire (Murphy et al., 2018) found that the EC and PT subscales are strongly related ($r = .66$, $r = .76$ after correcting for attenuation by our calculations), and in some factor solutions load on the same latent variable. This strong relationship makes sense given the causal relationship between taking another's perspective and experiencing empathic concern for that person. Indeed, perhaps the only reason that PT and EC do not correlate at unity is that perspective taking is a cognitive ability that is applicable to all instances of inferring another's thoughts and feelings, not just others' distress. Alternatively, PT and EC may vary somewhat independently because EC can be activated independently of PT. An interesting question for future research is whether EC would vary independently of PT if the items on the IRI PT sub-scale all related to exposure to suffering.

Instructions to imagine a suffering person's plight does not increase empathic concern. The unimpressive effect of imagine-other and imagine-self instructions on empathic concern implies that increasing helping at the societal level would not be as simple as encouraging people to take the perspective of others when they encounter needy others. A potentially more promising method by which nonprofit organizations, community leaders, and policymakers could increase empathic concern is by magnifying its cognitive antecedents, such as the perceived impact of and anticipated positive affect from helping (Butts, Lunt, Freling, & Gabriel, 2019). An alternative approach to increasing helping could involve circumventing empathic concern altogether, such as by encouraging people to reason about their moral obligations to needy others or the self-interested benefits of helping needy others, regardless of how much empathic concern their plights arouse (Comunian & Gielen, 1995; Hoffman, 2000).

There are also multiple ways one could interpret the fact that imagine-self or imagine-other instructions did not increase empathic concern beyond default levels. The most straightforward possibility is that people are incapable of empathizing more than they currently do. However, we refrain from drawing such a strong conclusion because "failures" to empathize can reflect a lack of motivation or a (current, but surmountable) lack of ability (Keysers & Gazzola,

2014). The clearest evidence for both motivation and ability deficits comes from brain imaging research on incarcerated populations who are high in psychopathic traits. Studies indicate that psychopaths are capable of imagining *themselves* in another's shoes (Decety, Chen, Harenski, & Kiehl, 2013), but do not spontaneously do so (Meffert, Gazzola, Den Boer, Bartels, & Keysers, 2013), suggesting a lack of motivation but not ability. In contrast, psychopaths neither spontaneously imagine *others*' point of view (Drayton et al., 2018) nor have the typical empathic response when they attempt to do so (Decety et al., 2013), suggesting a lack of ability and perhaps motivation as well.

Although it is possible that the nonclinical populations we studied have an insurmountable inability to increase empathic concern via perspective taking, it could be that perspective-taking instructions lack efficacy simply because participants have little incentive in experiments to try hard to comply with them. It is also possible that the standard imagine-other and imagine-self instructions are simply far too weak to exert the effects that more potent interventions might create. For instance, researchers have observed that participants assigned to receive compassion meditation training showed increased empathic concern and helping relative to control groups (Ashar et al., 2016; Leiberg et al., 2011; Weng et al., 2013) relative to control groups. On the contrary, these effects may not manifest when using active control groups (Ashar, Andrews-Hanna, Halifax, Dimidjian, & Wager, 2019; Kreplin, Farias, & Brazil, 2018), consistent with the possibility that demand characteristics are responsible for the positive effects of meditation training.

Study characteristics did not moderate the observed effects. Although we did not find that any one moderator had significant effects across more than one dataset, we refrain from concluding that deliberately taking the perspective of others more than one normally would never increase empathic concern, or that deliberately remaining objective would reduce empathic concern for everyone. Indeed, we view the present study as informative not only because it reveals the average effects of perspective-taking instructions in the typical paradigms which they are used, but also in revealing substantial gaps in our knowledge about *when* deliberate attempts to upregulate or downregulate empathic concern would be effective. There are at least six dimensions along which it is reasonable to theorize that the efficacy of imagine instructions and remain-objective instructions may differ from the pattern that we reported here. We hope researchers will systematically vary these dimensions in future experiments to clarify the generalizability of the present results.

First, assessing moderation by gender composition may have had limited evidential value because coding characteristics at the study level that in fact vary at the individual level substantially reduces the statistical power of mixed-effect models (Lambert, Sutton, Abrams, & Jones, 2002; Riley,

Lambert, & Abo-Zaid, 2010). Furthermore, effects of gender composition could be confounded with study characteristics that predict gender composition (e.g., research group, research location, the year the study was conducted, etc.). We expect that the increasing public availability of original data will allow future meta-analysts to examine the influence of gender on the perspective taking–empathic concern link using a multilevel meta-analysis framework.

Second, all of the studies in the present meta-analysis measured empathic concern, the broad construct that reflects altruistic motivation. However, empathic concern encompasses at least two narrower constructs, tenderness and sympathy (Lishner, Batson, & Huss, 2011; López-Pérez, Carrera, Oceja, Ambrona, & Stocks, 2019; Niezink, Siero, Dijkstra, Buunk, & Barelds, 2012). People feel tenderness in response to targets that are potentially vulnerable even if they are not in current distress, such as babies. People feel sympathy in response to targets that are currently needy, even if they are not particularly vulnerable (e.g., a strong person suffering an injury). Sympathy is more strongly connected to short-term helping, whereas tenderness is linked more strongly to long-term caring behavior. Although some situations elicit tenderness but not sympathy, the need situations that laboratory participants faced in the present dataset probably elicit both emotions because perspective-taking targets are typically both needy and vulnerable. Because the included studies aggregated across both emotions to create an index of empathic concern (e.g., by averaging words like "sympathy" and "compassionate" with words like "tenderness" and "softhearted"), we cannot ascertain whether the pattern of results we observed applies to tenderness, sympathy, or both. We suspect that our findings primarily hold for sympathy given that perspective-taking instructions typically do cause empathy-based helping (Batson, 2011; McAuliffe, Forster, et al., 2018), but future experiments that measure tenderness and sympathy separately are necessary to confirm this assumption.

Third, only a few relevant studies had participants learn about more than one victim. A recent study indicates that imagine-other and imagine-self instructions may ameliorate the well-known decline in empathic concern that occurs with an increasing number of victims (Ministero, Poulin, Buffone, & DeLury, 2018). Perhaps deliberate perspective-taking efforts lack efficacy only in situations that are not felicitous for experiencing empathic concern.

Fourth, perspective-taking targets were uniformly depicted as deserving of help. Even experiments that featured members of stigmatized groups presented them as either victims or remorseful perpetrators (Batson, Polycarpou, et al., 1997). Imagine-other and imagine-self instructions may have larger effects when directed at needy persons that do not normally elicit empathic concern.

Fifth, the experiments were mostly conducted in North America and Western Europe, and for the most part using college students. Effortful perspective taking would likely also have a larger effect in societies where people's sphere of concern typically do not extend beyond those near and dear to them (Schwartz, 2007). More generally, inquiry into the universality of the observed effects awaits experimentation in cultures with less western influence (Henrich, Heine, & Norenzayan, 2010).

Finally, we did not include personality traits as moderators in the mixed-effects models because most studies did not include individual differences measures. However, some experiments have found that imagine-other instructions increase empathic concern primarily among subjects who are low in agreeableness or trait empathic concern (Graziano et al., 2007; Habashi, Graziano, & Hoover, 2016; McAuliffe, Forster, et al., 2018). People who are low in these traits are likely a minority in that they do not by default take the perspectives of others, and thus can receive a boost in empathic concern by following perspective-taking instructions. Overall, until future research shows otherwise, it is prudent to generalize our conclusions only to situations in which relatively agreeable westerners are compelled to consider the plight of a single needy person who is presented in a sympathetic light.

Conclusion

How do people react, and how *could* they react, when they observe another person in distress? We conducted a series of meta-analytic tests on the effect of perspective-taking instructions on empathic concern to answer these questions. We found that participants generally empathize with others unless they actively refrain from considering their perspective. Participants who deliberately took the perspective of others did not experience significantly more empathic concern than persons who behaved naturally. These observations were for the most part robust to corrections for publication bias. However, our efforts to explain between-study variance in these effects were not particularly successful, and alarmingly few of the studies were adequately powered. Future researchers should focus especially on conducting large-scale experiments in diverse populations testing whether deliberate perspective taking increases empathic concern when relatively callous persons observe relatively unsympathetic victims. The first 50 years of research that employed perspective-taking instructions used them as a means to observing the effects of empathic concern on prosocial behavior and intergroup relations. We envision a new wave of research that employs perspective-taking instructions to study when and to what extent people experience empathic concern for the needy and distressed in the first place.

Declaration of Conflicting Interests

ORCID iD

William H. B. McAuliffe (ID) https://orcid.org/0000-0002-7680-1686

Notes

1. We use terms like "spontaneous," "natural," and "default tendency" to refer generically to what people do in the absence of an explicit request or incentive (see Davis et al., 2004, for a similar definition). We remain agnostic about whether such tendencies are "natural" in the sense that they are innate or are "spontaneous" in the sense of reflecting automatized rather than deliberate processes.

2. In treating the no-instructions condition as a control group, we do not assume that the behavior of participants who receive no perspective-taking instructions are a perfect proxy for how people behave in non-laboratory settings. Rather, we treat the no-instructions condition as comparable to a "real-life" setting in the limited sense that participants are not receiving an intervention that is designed to alter how they respond to observing others in need. We assess the generalizability of our findings to daily life in the present study by examining study-level moderators, and in the discussion point out several future moderators that should be manipulated to further probe the generalizability of our results.

References

Ashar, Y., Andrews-Hanna, J. R., Halifax, J., Dimidjian, S., & Wager, T. (2019). Effects of compassion training on brain responses to suffering others. *bioRxiv*. Retrieved from https://www.biorxiv.org/content/10.1101/616029v1

Ashar, Y. K., Andrews-Hanna, J. R., Yarkoni, T., Sills, J., Halifax, J., Dimidjian, S., & Wager, T. D. (2016). Effects of compassion meditation on a psychological model of charitable donation. *Emotion*, *16*, 691-705.

Balliet, D., & Gold, G. (2005, January). *Perspective-taking, level of action construal, and altruistic motivation*. Poster presented at the annual meeting of the Society for Personality and Social Psychology, *New Orleans, LA*.

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

Batson, C. D., Ahmad, N., Yin, J., Bedell, S. J., Johnson, J. W., & Templin, C. M. (1999). Two threats to the common good: Self-interested egoism and empathy-induced altruism. Personality and Social Psychology Bulletin, *25*, 3-16.

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

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

Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality*, *55*, 19-39.

Batson, C. D., Klein, T. R., Highberger, L., & Shaw, L. L. (1995). Immorality from empathy-induced altruism: When compassion and justice conflict. *Journal of Personality and Social Psychology*, *68*, 1042-1054.

Batson, C. D., Polycarpou, M. P., Harmon-Jones, E., Imhoff, H. J., Mitchener, E. C., Bednar, L. L., . . . Highberger, L. (1997). Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of Personality and Social Psychology*, *72*, 105-118.

Bekkers, R. H. F. P., Ottoni-Wilhelm, M., & Verkaik, D. J. (2015). *Altruism, warm glow, and charitable giving: Three experiments*. Retrieved from https://renebekkers.files.wordpress.com/2011/10/15_09_30_spi_bekkers_verkaik_ottoni-wilhelm.pdf

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, *61*, 27-41.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. Sussex, UK: Wiley Books.

Bornstein, B. H. (1998). From compassion to compensation: The effect of injury severity on mock jurors' liability judgments. *Journal of Applied Social Psychology*, *28*, 1477-1502.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365-376.

Butts, M. M., Lunt, D. C., Freling, T. L., & Gabriel, A. S. (2019). Helping one or helping many? A theoretical integration and meta-analytic review of the compassion fade literature. *Organizational Behavior and Human Decision Processes*, *151*, 16-33.

Cameron, C. D., Harris, L. T., & Payne, B. K. (2016). The emotional cost of humanity: Anticipated exhaustion motivates dehumanization of stigmatized targets. *Social Psychological and Personality Science*, *7*, 105-112.

Cameron, C. D., Hutcherson, C. A., Ferguson, A. M., Scheffer, J. A., Hadjiandreou, E., & Inzlicht, M. (2019). Empathy is hard work: People choose to avoid empathy because of its cognitive costs. *Journal of Experimental Psychology: General*, *148*, 962-976.

Cameron, C. D., & Payne, B. K. (2011). Escaping affect: How motivated emotion regulation creates insensitivity to mass suffering. *Journal of Personality and Social Psychology*, *100*, 1-15.

Carter, E., Schönbrodt, F., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*, 115-144.

Cikara, M., Bruneau, E. G., & Saxe, R. R. (2011). Us and them: Intergroup failures of empathy. *Current Directions in Psychological Science*, *20*, 149-153.

Coburn, K. M., & Vevea, J. L. (2017). Weightr: Estimating weight-function models for publication bias (R Package Version 1.1.2). Retrieved from https://CRAN.R-project.org/package=weightr

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Comunian, A. L., & Gielen, U. P. (1995). Moral reasoning and prosocial action in Italian culture. *The Journal of Social Psychology*, *135*, 699-706.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113-126.

Davis, M. H., Soderlund, T., Cole, J., Gadol, E., Kute, M., Myers, M., & Weihing, J. (2004). Cognitions associated with attempts to empathize: How do we imagine the perspective of another? *Personality and Social Psychology Bulletin*, *30*, 1625-1635.

Decety, J., Chen, C., Harenski, C., & Kiehl, K. A. (2013). An fMRI study of affective perspective taking in individuals with psychopathy: Imagining another in pain does not evoke empathy. *Frontiers in Human Neuroscience*, *7*, Article 489.

Del Re, A. C. (2010). compute.es: Compute effect sizes (Version R Package Version 0.2). Available from http://www.R-project.org

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

Drayton, L. A., Santos, L. R., & Baskin-Sommers, A. (2018). Psychopaths fail to automatically take the perspective of others. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 3302-3307.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455-463.

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

Feigenson, N. R. (1997). Sympathy and legal judgment: A psychological analysis. *Tennessee Law Review*, *65*, 1-78.

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

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

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, *78*, 708-724.

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

Graziano, W. G., Habashi, M. M., Sheese, B. E., & Tobin, R. M. (2007). Agreeableness, empathy, and helping: A person × situation perspective. *Journal of Personality and Social Psychology*, *93*, 583-599.

Habashi, M. M., Graziano, W. G., & Hoover, A. E. (2016). Searching for the prosocial personality: A Big Five approach to linking personality and prosocial behavior. *Personality and Social Psychology Bulletin*, *42*, 1177-1192.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*, 29.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539-1558.

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

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

Ioannidis, J. P., Stanley, T. D., & Doucouliagos, H. (2017). The power of bias in economics research. *The Economic Journal*, *127*, 236-265.

Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, *4*, 245-253.

Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem. *Statistical Science*, *3*, 109-117.

Karmali, F., Kawakami, K., & Page-Gould, E. (2017). He said what? Physiological and cognitive responses to imagining and witnessing outgroup racism. *Journal of Experimental Psychology: General*, *146*, 1073-1085.

Keysers, C., & Gazzola, V. (2014). Dissociating the ability and propensity for empathy. *Trends in Cognitive Sciences*, *18*, 163-166.

Kreplin, U., Farias, M., & Brazil, I. A. (2018). The limited prosocial effects of meditation: A systematic review and meta-analysis. *Scientific Reports*, *8*(1), Article 2403.

Lambert, P. C., Sutton, A. J., Abrams, K. R., & Jones, D. R. (2002). A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology*, *55*, 86-94.

Leiberg, S., Klimecki, O., & Singer, T. (2011). Short-term compassion training increases prosocial behavior in a newly developed prosocial game. *PLoS ONE*, *6*(3), e17798.

Lishner, D. A., Batson, C. D., & Huss, E. (2011). Tenderness and sympathy: Distinct empathic emotions elicited by different forms of need. Personality and Social Psychology Bulletin, *37*, 614-625.

Loewenstein, G., & Small, D. A. (2007). The Scarecrow and the Tin Man: The vicissitudes of human sympathy and caring. *Review of General Psychology*, *11*, 112-126.

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

López-Pérez, B., Carrera, P., Oceja, L., Ambrona, T., & Stocks, E. (2019). Sympathy and tenderness as components of dispositional empathic concern: Predicting helping and caring behaviors. *Current Psychology*, *38*, 458-468.

Marsh, A. A. (2019). The caring continuum: Evolved hormonal and proximal mechanisms explain prosocial and antisocial extremes. *Annual Review of Psychology*, *70*, 347-371.

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion*, *5*, 175-190.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, *23*, 209-237.

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*, 493-506.

McAuliffe, W. H. B., Pedersen, E. J., & McCullough, M. E. (2018, March). *Cognitive models underlying inconsistent self-reported gender differences in empathy*. Paper presented at the annual meeting of the Society for Personality and Social Psychology, Atlanta, GA.

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, *11*, 730-749.

Meffert, H., Gazzola, V., Den Boer, J. A., Bartels, A. A., & Keysers, C. (2013). Reduced spontaneous but relatively normal deliberate vicarious representations in psychopathy. *Brain*, *136*, 2550-2562.

Ministero, L. M., Poulin, M. J., Buffone, A. E., & DeLury, S. (2018). Empathic concern and the desire to help as separable components of compassionate responding. *Personality and Social Psychology Bulletin*, *44*, 475-491.

Morelli, S. A., & Lieberman, M. D. (2013). The role of automaticity and attention in neural processes underlying empathy for happiness, sadness, and anxiety. *Frontiers in Human Neuroscience*, *7*, Article 160.

Mullen, B., & Hu, L. T. (1989). Perceptions of ingroup and outgroup variability: A meta-analytic integration. *Basic and Applied Social Psychology*, *10*, 233-252.

Murphy, B. A., Costello, T. H., Watts, A. L., Cheong, Y. F., Berg, J. M., & Lilienfeld, S. O. (2018). Strengths and weaknesses of two empathy measures: A comparison of the measurement precision, construct validity, and incremental validity of two multidimensional indices. *Assessment*. Advance online publication. doi:10.1177/1073191118777636

Niezink, L. W., Siero, F. W., Dijkstra, P., Buunk, A. P., & Barelds, D. P. (2012). Empathic concern: Distinguishing between tenderness and sympathy. *Motivation and Emotion*, *36*, 544-549.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Pedersen, E. J., McAuliffe, W. H. B., & McCullough, M. E. (2018). The unresponsive avenger: More evidence that disinterested third parties do not punish altruistically. *Journal of Experimental Psychology: General*, *147*, 514-544.

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org/

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *British Medical Journal*, *340*, Article c221.

Rumble, A. C., Van Lange, P. A., & Parks, C. D. (2010). The benefits of empathy: When empathy may sustain cooperation in social dilemmas. *European Journal of Social Psychology*, *40*, 856-866.

Schumann, K., Zaki, J., & Dweck, C. S. (2014). Addressing the empathy deficit: Beliefs about the malleability of empathy predict effortful responses when empathy is challenging. *Journal of Personality and Social Psychology*, *107*, 475-493.

Schwartz, S. H. (2007). Universalism values and the inclusiveness of our moral universe. *Journal of Cross-Cultural Psychology*, *38*, 711-728.

Shaw, L. L., Batson, C. D., & Todd, R. M. (1994). Empathy avoidance: Forestalling feeling for another in order to escape the motivational consequences. *Journal of Personality and Social Psychology*, *67*, 879-887.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*, 666-681.

Slovic, P., Västfjäll, D., Erlandsson, A., & Gregory, R. (2017). Iconic photographs and the ebb and flow of empathic response to humanitarian disasters. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 640-644.

Smith, K. D., Keating, J. P., & Stotland, E. (1989). Altruism reconsidered: The effect of denying feedback on a victim's status to empathic witnesses. *Journal of Personality and Social Psychology*, *57*, 641-650.

Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Stanley, T. D., Carter, E. C., & Doucouliagos, C. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin*, *144*, 1325-1346.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*, 60-78.

Stanley, T. D., Doucouliagos, H., & Ioannidis, J. (2017). Finding the power to reduce publication bias. *Statistics in Medicine*, *36*, 1580-1598.

Stotland, E. (1969). Exploratory investigations of empathy. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (*Vol. 4*, pp. 271-313). New York, NY: Academic Press.

Stürmer, S., Snyder, M., Kropp, A., & Siem, B. (2006). Empathy-motivated helping: The moderating role of group membership. *Personality and Social Psychology Bulletin*, *32*, 943-956.

Sun, B., Li, W., Lou, B., & Lv, L. (2011). The relation of perspective taking and helping behavior: The role of empathy and group status. In *Proceedings of the 2011 international conference of social science and humanity*. Retrieved from https://pdfs.semanticscholar.org/5d76/54d8f158b1e2c7255520bcc9d1159032aa02.pdf

Takaku, S. (2001). The effects of apology and perspective taking on interpersonal forgiveness: A dissonance-attribution model of interpersonal forgiveness. *The Journal of Social Psychology*, *141*, 494-508.

Toi, M., & Batson, C. D. (1982). More evidence that empathy is a source of altruistic motivation. *Journal of Personality and Social Psychology*, *43*, 281-292.

van Aert, R. C. (2018). puniform package for R. Retrieved from https://rdrr.io/github/RobbievanAert/puniform/man/puniform.html

van Assen, M. A., van Aert, R., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, *20*, 293-309.

van Erp, S., Verhagen, J., Grasman, R. P., & Wagenmakers, E. J. (2017). Estimates of between-study heterogeneity for 705 meta-analyses reported in *Psychological Bulletin from 1990-2013*. *Journal of Open Psychology Data*, *5*, 4.

Veerbeek, M. (2005). *Ben ik belangrijker dan jij? Een studie naar de invloed van empathie op prosociaal gedrag* [Am I more important than you are? A study on the influence of empathy on prosocial behavior]. Unpublished paper, VU University, Amsterdam, The Netherlands.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1-48.

Weng, H. Y., Fox, A. S., Shackman, A. J., Stodola, D. E., Caldwell, J. Z., Olson, M. C., . . . Davidson, R. J. (2013). Compassion training alters altruism and neural responses to suffering. *Psychological Science*, *24*, 1171-1180.

Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science*, *14*, 131-134.

Zaki, J. (2014). Empathy: A motivated account. *Psychological Bulletin*, *140*, 1608-1647.