UNIVERSITY OF CALIFORNIA,
IRVINE

A meta-analysis of religious prosociality

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF ARTS

in Social Ecology

by

John Michael Kelly

Thesis Committee:
Professor Peter Ditto, Chair
Associate Professor Azim Shariff
Assistant Professor Paul Piff

2019

DEDICATION

To

my wife Kaitie

for supporting me in my interminable trip through academia, and my unwillingness to get a normal job

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The work reported in this thesis is also reported in the manuscript "Religiosity predicts prosociality (at least when measured with self-reports): A meta-analysis," which is under peer review at the time this thesis is being submitted. Much of that manuscript is reproduced within this thesis. I would like to sincerely express my gratitude to my co-authors on that manuscript, Stephanie Kramer and Azim Shariff, who have put years of their life into this project. Without their contribution, neither this thesis nor the manuscript on which it is based would be a reality.

I additionally want to offer my gratitude to the chair of my committee, Pete Ditto, who has taken me under his wing this last year, and continues to offer his support even while on sabbatical. I also wish to thank the other member of my committee, Paul Piff, whose intellectual contributions have been enormously helpful, and from whom I've drawn a great deal of inspiration.

I further would like to thank my many friends and classmates at UCI, who have provided a tremendous amount of support, both intellectual and moral. My graduate experience would be far less productive and enjoyable without them.

I must also thank my parents, who instilled my intellectual curiosity and love of learning from a young age, and whose love and support has been immeasurable. And finally I offer my thanks to my wife Kaitie, who has never left my side, and who I love so very much.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1839285

ABSTRACT OF THE THESIS

A meta-analysis of religious prosociality

By

John Michael Kelly

Master of Arts in Social Ecology

University of California, Irvine, 2019

Professor Peter Ditto, Chair

This meta-analysis explores the longstanding and heavily debated question of whether religiosity is associated with prosocial and antisocial behavior. In an analysis of 179 effects across 89 samples, encompassing 167,508 participants, a significant relationship of r = .10 was found between religiosity and prosociality. However, substantial heterogeneity of methods was identified, and several potential moderators of this relationship were explored. The effect was most powerfully moderated by the type of measurement used to assess prosocial or antisocial behavior. Religiosity significantly correlated with self-report measures of prosociality at r = .15, but among samples using a behavioral measure of prosociality, the effect was only a marginally significant r = .04. Three possible explanations of this moderation are discussed, namely that 1) lab-based methods do not accurately capture religious prosociality; 2) the self-report effect is explained by religious self-enhancement; or 3) both religiosity and self-reported prosociality are explained by self-enhancement. Recommendations for future research are discussed that may help resolve these possible explanations. Other potential moderators are analyzed, including whether religiosity is differently related to prosocial and antisocial behavior.

INTRODUCTION

All morality proceeds from religion, for religion is simply the formula of morality.

–Dostoevsky, translated in Scanlan, 2010, p. 94

[E]thics and morality are quite independent of faith, and cannot be derived from it.

–Christopher Hitchens, *God is Not Great*, 2007

For nearly as long as we have record of human writing on the subject of ethics and morality, morality and religion have been perceived to be intimately linked. The first law code of which we have record, the Sumerian Code of Ur-Nammu, makes appeals to the gods of their people, Utu and Nanna, to establish moral and legal claims (Finkelstein, 1968). This pattern of divine command of moral laws is repeated time and time again in ancient law codes like the Code of Hammurabi (Prince, 1904) and the Law of Moses. The discussion of Euthyphro with Socrates reveals that such a link between religion and morality continued into the era of Ancient Greece (Plato, trans. 2002). And this link between religion and morality is certainly not limited to Western and Near Eastern thought. The ancient concept of Dharma, central to Hindu religion, has alternatively been translated into English as "religion," "morality," and "justice" (Holredge, 2004).

Yet while the perceived link between religion and morality is widespread and enduring, the question remains whether those who are religious indeed act more morally than those who are not. The belief that this may be so is by no means constrained to ancient people. In a 2014 survey, a majority of those sampled internationally believed that belief in God was not only conducive to being a good person, but *necessary* for morality (Pew Research Center, 2014). To take these findings literally, these respondents suggest that

non-believers could not be moral. Even in the increasingly secular United States, a large minority of respondents suggest this to be true (Smith 2017). Of course, others strongly contest this claim. Secular humanist thinkers like Christopher Hitchens and Richard Dawkins have argued that religion is, at best, unrelated to morality and, at worst, harmful to true morality. Yet, while the question of whether the religious behave more morally than the nonreligious has enjoyed decades of empirical scrutiny, a firm answer has remained elusive. Here, I discuss some of the literature that has aimed to address this question and report a meta-analysis investigating the relationship between religiosity and two facets of morality: prosocial and antisocial behavior.

History

While the question has been long-debated, earnest attempts to empirically investigate the relationship between religiosity and morality (and, more specifically, prosocial and antisocial behavior) were until recently quite spotty. Early work in this area primarily investigated the effects of religiosity on prejudice or acceptance, and many studies seemingly came away with contradictory results. Some studies found little relationship between religiosity and humanitarian concerns (Cline & Richards, 1965; Rokeach, 1970), while others found that religiosity positively predicted prejudice (Allport & Kramer, 1946; Gough, 1951; Kirkpatrick, 1949) or positively predicted acceptance (at least for certain forms of religiosity; Allen & Spilka, 1967; Allport & Ross, 1967). Furthermore, some suggested that a simple linear relationship between religiosity and prosociality was not the proper way of investigating these questions. In a modified version of the Milgram study, Bock & Warren (1972) suggested an inverted U relationship, such that moderately religious people were more likely to administer "shocks," when compared

with people on either end of the religious spectrum. Additionally, "religiosity" was proposed to not exist as one homogenous phenomenon, but rather a broader phenomenon composed of differentiable motivations and behaviors, which differentially predict prosociality (Allport & Ross, 1967).

Despite these early disagreements and challenges, investigations into religion's relationship with morality have only accelerated over the last decade or so. A recent meta-analysis of religious priming found a reliably positive association between prosociality and the momentary activation of religious thinking (Shariff, Willard, Andersen, & Norenzayan, 2015). But findings about whether prosociality relates to *dispositional* religiosity—that is, whether the everyday religious person acts more prosocially than the everyday non-religious person—have remained mixed. Different studies have found that religiosity may promote prosocial behavior (e.g. Shepperd, Miller, & Smith, 2015), may promote forms of antisocial behavior (e.g. Jackson & Gray, 2018), or may be unrelated to prosociality (e.g. Eckel & Grossman, 2004). As a result, the literature would benefit from an aggregation of these results, providing a comprehensive view of findings, and investigating potential systematic sources of discrepancies.

Prosocial Behavior

While the association between religiosity and goodness often focuses on morality, broadly construed, here the analysis is limited to prosocial behavior (and its antithesis, antisocial behavior). It is difficult, if not impossible, to compare overall moral behavior across levels of religiosity, as religiosity itself determines which issues are placed in the moral domain (McKay & Whitehouse, 2015). Many such disagreements on what constitute moral questions arise from differences in doctrinal beliefs, yet disagreements exist even

among issues that are not strictly relevant to doctrine (Shariff, Piazza, & Kramer, 2014). While nearly all people see issues of justice and harm as being morally relevant, religious people are more likely than the nonreligious to place issues pertaining to ingroup loyalty, respect for authority, and purity in the moral domain (Graham & Haidt, 2010).

Despite these discrepancies in moral judgment, there are areas in which the religious and nonreligious tend to agree (Shariff et al., 2014). Prosocial behaviors, which Batson (2012) describes as "a broad range of actions intended to benefit one or more people other than oneself—behaviors such as helping, comforting, sharing, and cooperating" (p. 242) are morally valued by religious and non-religious people alike. As such, prosocial behaviors are an area of morality, albeit a limited one, in which people may legitimately be compared across the religious spectrum. Conversely, it is also possible to investigate religion's relationship with antisociality, which, here, is roughly operationalized as the opposite of prosociality: self-interested behavior at the expense of others. Throughout this manuscript, the general terms "prosociality" and "prosocial behavior" refer both to true prosociality and the avoidance of antisociality, unless otherwise specified, but it is important to recognize that these are separable constructs. In the meta-analysis to follow, religiosity's association with both prosociality and antisociality is explored.

Why might one expect the religious to be more prosocial?

There are a number of reasons to expect the religious to be more prosocial than the nonreligious. All of the world's major religions emphasize prosociality. Jesus' ministry was characterized by his selfless sacrifice for the good of the world and the call for his followers to do good in a similar manner. Judaism and Islam both emphasize the importance of loving

one's neighbor as oneself. Love, compassion, and Karma are fundamental to Buddhism. Often, these religions teach of divine injunctions promoting moral behavior. Omniscient gods or forces such as Karma are believed to reward good behavior and punish immorality. Fear of divine punishment is associated with the avoidance of counter-normative behavior (Johnson & Krüger, 2004). It stands to reason that followers of these religions would thus seek to behave in a manner consistent with the values of their religious doctrines.

Furthermore, religion is associated with strong communities and compliance to group norms (Baumeister, Bauer, & Lloyd, 2010). Religious communities may bind groups of religious people into "moral communities" that encourage charitability and prosociality more broadly (Graham & Haidt, 2010). Certainly, community effects on prosociality are not limited to religious communities; even commitment to secular groups like bowling leagues may increase charitable giving (Putnam, 2000). Yet religion may provide particularly strong and accessible communities with a strong moral focus.

Religiosity is also correlated with personality factors associated with prosociality. Religious people tend to be higher in agreeableness and conscientiousness (Saroglou, 2002). Agreeableness predicts prosocial values (Carlo, Okun, Knight, & de Guzman, 2005), volunteerism (Lodi-Smith & Roberts, 2007) and prosocial strategies for resource control (Hawley, 2003). Conscientiousness is associated with self-regulation (Roberts, Chernyshenko, Stark, & Goldberg, 2005), self-reported empathy (Barrio, Aluja, & García, 2004), and workplace integrity (Murphy & Lee, 1994). Both agreeableness and conscientious are negatively associated with antisociality (Jones, Miller, & Lynam, 2011) and aggression (Murphy & Lee, 1994). It is therefore reasonable to expect these

dispositional correlates of religiosity to be associated with relatively high levels of prosociality and low levels of antisociality.

Lay beliefs tend to be consistent with the expectation that religiosity leads to increased prosociality. As mentioned above, a great many people worldwide suggest belief in God is necessary for morality (Pew Research Center, 2014). While the proportion of Americans agreeing with this assertion has declined in recent years, a sizable minority still think that morality without belief in God is impossible (Smith, 2017). Atheists, due to a perceived lack of supernatural beliefs safeguarding against immoral behavior, are largely seen as untrustworthy (Gervais, Shariff, & Norenzayan, 2011). The average theist is seen more favorably than the average atheist, and these perceptions shape views of theists' and atheists' morality (Brown-Iannuzzi, McKee, & Gervais, 2018). Even American atheists themselves tend to believe that religion generally works to promote moral behavior (Gervais, 2014).

Finally, some empirical evidence also points towards a positive effect of religiosity on prosocial behavior. Priming research shows that experimentally manipulating the salience of religion has a positive effect on the prosocial behavior of religious (but not irreligious) participants (Shariff et al., 2015). While these findings say little about the effects of *dispositional* religiosity on prosocial behavior (the relationship of interest in this paper), they nonetheless demonstrate religion's potential for a positive effect on prosociality. Perhaps more indicative of a relationship between dispositional religiosity and prosociality is a meta-analysis conducted by Fehr and colleagues which demonstrated a positive relationship between religiosity and tendency to forgive (Fehr, Gelfand, & Nag, 2010). Following McCullough, Pargament, and Thoresen's (2001) definition of forgiveness

as "prosocial change toward a perceived transgressor," this appears to be a clear positive effect of religiosity on at least a limited aspect of prosocial behavior.

Why might one *not* expect the religious to be more prosocial?

Critics note that, while religious doctrine may advocate prosociality, this does not necessarily mean such doctrine is actively and consistently followed (Dawkins, 2016; Hitchens, 2007; Slone, 2007). In Darley and Batson's (1973) classic study, the parable of the Good Samaritan did not encourage helping behavior even when made directly salient. Indeed, there are reasons to believe that dispositional religiosity has no relationship (or even a negative relationship) with prosocial behavior. First, although religious doctrines and institutions may, in fact, positively contribute to prosocial behavior, secular alternatives may do so as well, and may sometimes do so even more effectively—thereby either erasing or reversing any religious advantage. Non-religious philosophies (such as humanism and utilitarianism) and secular institutions (such as police and the court system) similarly encourage prosocial behavior and discourage antisocial behavior. Nations low in religiosity but high in confidence in these non-religious institutions, such as Japan and Norway, do not seem to suffer from widespread antisociality (Crabtree, 2010; Norris & Inglehart, 2004; Paul, 2005). In the United States, a decrease in religiosity has corresponded with a decrease in violent crime (Zuckerman, 2009). Certainly correlation does not imply causation, but this relationship does serve as evidence that decreases in antisociality can be found in the absence of religion.

Furthermore, there exists empirical evidence suggesting that religiosity may increase antisociality towards outgroups. A meta-analysis revealed a significant correlation between religiosity and racism (Hall, Matz, & Wood, 2010). Religious groupishness seems

to have broadly divergent consequences for ingroup versus outgroup members, providing support and boosting health for the ingroup, while encouraging denigration of the outgroup (Shariff et al., 2014). The coalitional element of religiosity, as measured by frequency of religious service attendance, has been found to correlate with increased hostility toward religious outgroups cross-culturally (Ginges, Hansen, & Norenzayan, 2009). Asking participants to recall their recent religious service attendance – though not their acts of personal religious devotion – even increased support for a suicide attack on the outgroup. Further, Bushman, Ridge, Das, Key and Busath (2007) reported that when participants read a passage depicting coalitional violence, those who were told that it was from the Bible, rather from a more neutral ancient scroll, acted more aggressively in a later task. Similarly, religious people may be more likely than the non-religious to engage in passive immorality when the immoral act can be construed to be allowed or sanctioned by God (Jackson & Gray, 2018), or active immorality when the behavior could be forgiven by a merciful God (DeBono, Shariff, Poole & Muraven, 2017).

Finally, some skepticism may be warranted towards empirical findings demonstrating a positive relationship between religiosity and prosociality when those studies use a self-reported measure of prosociality. Researchers have long noted with concern that studies investigating prosociality tend to use self-report measures (e.g. Batson, Neifeh, & Pate, 1978). Prosociality and the avoidance of antisociality are socially desirable behaviors, and thus self-reports of these behaviors are highly susceptible to self-enhancement biases. This is especially troubling when one considers that religious people show a greater tendency towards self-enhancement and socially desirable responding than do the non-religious (Crandall & Gozali, 1969; Sedikides & Gebauer, 2010). Therefore, if a

relationship between religiosity and self-reported prosociality is found, it is difficult to know whether this finding should be attributed to a difference in true prosociality or to a difference in socially desirable responding.

The existing literature

Research and arguments like those outlined above have allowed the debate about religiosity and prosociality to persist without resolution. Studies directly testing the relationship have been similarly mixed in their conclusions, with effect sizes spanning a broad range, showing a relationship between religiosity and prosociality ranging from around $r = -0.4$ to $r = 0.6$ Much of the diversity in these findings seems to follow from a similar level of diversity in methods (Galen, 2012; Oviedo, 2016). Given this method heterogeneity, any meta-analytical investigation of the religiosity-prosociality link should be sensitive to how different methods may profoundly impact results. Thus I describe below the major dimensions of these differences.

Measures of prosociality. Importantly, the literature contains diverse measures of prosociality—some studies using self-report measures and others using behavioral measures. It has long been acknowledged that differences between these types of measures may lead to systematic and confounding discrepancies in recorded prosociality, but this issue has not yet been resolved (Batson, 1976; Galen, 2012). Self-report measures of prosociality include subjects' reports of their own helping behavior (Blogowska & Saroglou, 2011; Saroglou, Pichon, Trompette, Verschueren, & Dernelle, 2005), charitable donations (e.g. Brooks, 2003), and forgiveness (e.g. Brown, Barnes, & Campbell, 2007), as well as self-reports of anti-social behaviors like aggression (Saroglou et al., 2005). All such self-report measures are subject to self-enhancement and social desirability bias (John &

Robins, 1994; Maccoby & Maccoby, 1954; Nederhof, 1985). As noted above, there is also a positive association between measured self-enhancement and measured religiosity (at least intrinsic religiosity; Batson, Naifeh, & Pate, 1978; Sedikides & Gebauer, 2010; Trimble, 1997). As such, self-enhancement may systematically and problematically confound the relationship between religiosity and prosocial behavior.

Behavioral measures of prosocial behavior are less subject to self-serving biases, but they are not without their own flaws. Behavioral measures in this literature include such measures as dictator games (e.g. Benjamin, Choi, & Fisher, 2010; Eckel & Grossman, 2004; Paciotti et al., 2011), public goods games (e.g. Ahmed & Salas, 2009; Anderson, Mellor, & Milyo, 2010), the helping of confederates (e.g. Greer, Berman, Varan, Bobrycki, & Watson, 2005), and measures of antisocial behavior such as allocating hot sauce to punish a disliked individual (e.g. Blogowska et al., 2013). While these controlled, behavioral measures suffer relatively less from self-enhancement biases, and have higher internal validity than do self-report measures, they are lacking somewhat in external validity. Although several studies do show that commonly-used laboratory measures of prosociality and antisociality are predictive of "real-world" behavior out in the field (Anderson & Bushman, 1997; Benz & Meier, 2008; Cohn & Maréchal, 2016), many of these measures nonetheless diverge considerably from the naturalistic prosocial behavior that may be associated with religiosity. Prototypical religious prosociality includes missionary or charity work, charitable donations, and general "niceness" and "neighborliness." All such forms of prosociality are difficult to replicate naturalistically in the lab. Therefore, there are strengths and weaknesses to both self-report and behavioral measures, but there are fundamental differences among these methods that may lead to differing measured effects.

It is thus essential to investigate the type of prosociality measure as a potential moderator of the religiosity-prosociality link, as is done in the meta-analysis to follow.

Measures of religiosity. Just as the literature contains many measures of prosociality, it also features different means of measuring religiosity. These can roughly be separated these into three categories: affiliation, attendance, and scale measures. Unlike the measures of prosociality, these are all self-report, but they are still different in important ways. Affiliation questions simply inquire what religion (if any) participants are affiliated with (e.g. Benjamin et al., 2010; Clobert & Saroglou, 2013), or, in some cases, they simply ask whether or not the participant is religious in a dichotomous fashion (e.g. Ahmed & Salas, 2009). Attendance measures ask how often participants attend religious services (e.g. Bloodgood, Turnley, & Mudrack, 2008; Greer et al., 2005). Scale measures investigate a broad range of religious components, including beliefs, religious importance, and attitudes (e.g. Brooks, 2003; Galen et al., 2014; Xygalatas et al., 2016). Although all these methods are valid measures of religiosity, they vary in meaningful ways and capture different aspects of the religious experience. Certainly, this coding structure does not capture the full heterogeneity of the diverse construct of "religiosity" but investigating the type of measurement as a moderator may nonetheless provide an initial look at whether any such differences exist.

Sample. While the literature is rather heterogeneous in its measures, it is unfortunately more homogeneous in its samples. While some studies have made commendable efforts to collect cross-cultural samples (e.g. Xygalatas, 2012), most studies use WEIRD samples (Western, Educated, and from Industrialized, Rich, and Democratic countries; Henrich, Heine, & Norenzayan, 2010). As the nature of religiosity and cultural

perceptions of religiosity vary greatly across cultures, it would be valuable to have a larger collection of non-WEIRD studies to improve generalizability. Nonetheless, this limited WEIRD status variation will be included as a moderator within the meta-analysis.

While broad-level cultural diversity is lacking in the included samples, there is a wider diversity of student vs. non-student samples. While there is certainly value in using college student samples (Greenberg, 1987), there remains an enduring concern that college students samples are not representative of the larger population (Gordon, Slade, & Schmitt, 1986; Peterson, 2001). Indeed, students and non-students are importantly different in a number of areas, including prosociality and religiosity-relevant effects (Peterson, 2001; M. Zuckerman, Silberman, & Hall, 2013). As such, it is important to test whether any relationship found between religiosity and prosociality is generalizable, or whether it is an artifact of the student status of the sample. Thus, student status in included as another potential moderator.

The Present Study

In the present study, a meta-analysis is conducted to investigate the overall relationship between religiosity and prosociality, across the diversity of methods. Differences among methods are also systematically tested as moderators of the religiosity-prosociality relationship and explanations are proposed for the discrepancy in findings along the lines of these moderators.

In recognition of some of the concerns raised by Galen (2012), Oviedo (2016), and others in the field, priming effects were excluded from the data for this analysis and as many relevant moderators were included as possible. Studies investigating prejudice and prejudice-motivated antisociality were also excluded. While this exlusion unfortunately

prohibits the analysis of a great deal of early work on the relationship between prosociality and religiosity, which heavily emphasized prejudice, this exclusion was made to avoid contaminating the construct of religious prosociality with confounding factors like conservatism. Overall, every attempt was made to answer these questions: Is religiosity, at baseline, positively related to prosociality? And further, what factors detected in the literature moderate this potential relationship?

Method

Search and Selection

Google Scholar and PsycNET were used to search for relevant papers published any time before 2018 using the terms *relig*, believers, Christians, Jews, Muslims, Buddhists, Hindus* and *theists* for religious, combined with terms related to prosociality (*prosocial*, giving, charity, volunteer*, altruism*) or antisociality (*cheating, aggression, hostility, antisocial**). The reference sections of these papers produced further pertinent papers. My collaborator and I also posted a call requesting unpublished papers and data on spsp.org.

Studies were selected for analysis if they included a measure of religiosity at the individual level and one or more dependent variables related to prosociality or antisociality. To be analyzed, either the effect size (Pearson's *r*) for bivariate relations, or enough information to calculate them, must have been accessible. When these data were missing, my collaborator or I contacted authors for bivariate correlations or means and standard deviations (in the case of categorical measures of religiosity). However, several requests for data were not met with responses, and thus several studies were excluded for lack of necessary information. Studies that included only a categorical measure of religiosity and had fewer than 10 nonbelieving or unaffiliated participants were also

excluded, since having so few participants made it impossible to make meaningful comparisons between theists and nontheists. In total, 179 effects were included in the analysis across 89 samples (total unique N = 167,508). See Table 1 for a list of all studies included in the analysis.

Information Gathering Procedure

Effect sizes and information related to moderators were independently collected by both my collaborator and me. Discrepancies were resolved through discussion between the two coders. Because independent coding between two coders and resolution of discrepancies through discussion has been found to be a reliable method (Rosenthal, 1987), quantitative measurement of reliability was not required.

Behavioral or self-report measures of prosociality or antisociality. Each study's measure of prosociality or antisociality was coded as either a behavioral measure or a self-report measure. In most instances, this was a simple distinction to make. If the measure was a direct observation or measurement of behavior, this was operationalized as a behavioral measure. Common examples included dictator game donations (e.g. Miyatake & Higuchi, 2017), one's choice to volunteer (e.g. Lin, Tong, Lee, Low, & Gomes, 2016), and cheating behavior (e.g. Bloodgood et al., 2008). On the other hand, if the measure of prosocial or antisocial behavior was comprised of the participant's report of their own choices, behaviors, or tendencies, these were coded as self-report. Examples of self-report measures include self-reported volunteering (e.g. Gutierrez & Mattis, 2014), intention to donate in the future (Lehmann et al., 2017), and aggression scales (e.g. Smokowski, Guo, Cotter, Evans, & Rose, 2016).

Prosociality and antisociality. Additionally, each dependent variable measure was coded as either representing prosocial behavior or antisocial behavior. To be included, outcome variables had to appear repeatedly in the literature (i.e. show widespread agreement) as representative of our intended sociality constructs. Self-reports of charitable giving, volunteerism, dispositional altruism, willingness or intentions to help, forgiveness and hypothetical money allocation were found to be common operationalizations of prosociality. Self-reports related to vengeance, cheating, harm done to others, and aggression were operationalized as anti-sociality.

Similarly, behavioral measures of pro- or antisociality needed to be widely regarded as valid operationalizations of prosocial or antisocial behavior in the literature. The majority of these were economic games in which players cooperate to share a resource, or defect to keep more for themselves. These included dictator, ultimatum, trust, and public goods games. Other studies employed confederates to solicit help from participants. One compared self-reported interest in volunteering for specific charitable organizations to actual sign-ups when participants were later contacted by those organizations. The amount of electric shock or hot sauce participants intended to force on another participant were behavioral measures of antisociality.

Religiosity. As noted above, religiosity measures fit within three broad types: measures of affiliation, measures of attendance, or scale measures. Affiliation items either asked about what religion or denomination a participant belonged to, if any, or simply asked whether or not a participant was religious. In the studies included in this meta-analysis, even the former were typically treated as a dichotomous religious/non-religious variable. Measures of attendance evaluated a participant's self-reported attendance of

religious services or activities (usually in Likert scale form). Scale measures included a variety of different scales assessing a range of religiously-relevant attitudes and beliefs, including level of intrinsic religiosity, religious orthodoxy and alignment with religious prototypes. Spirituality is conceptually distinct from religiosity in important ways, and thus was excluded from this meta-analysis. It lacks the specific moral injunctions that theoretically contribute to religious prosociality, and is distinct from religiosity with regard to relevant personality correlates (McCullough & Willoughby, 2009).

Demographics. We additionally coded for several demographic characteristics of each study. First, each sample was coded for whether or not it came from a WEIRD culture. Second, each sample was coded for whether or not it was comprised of students. Third, we collected the average age of the sample. Finally, we collected the percentage of the sample that was female. When any of this demographic information was unavailable for a particular sample, that particular variable was left blank, and the study was not included in a moderator analysis investigating that variable.

Extraction and Coding

Effect sizes (Pearson's *r*) were extracted for each relevant bivariate correlation. Importantly, since the purpose of the present study was to examine the contributions of baseline religiosity on pro- and antisociality, rather than the impacts of religious situations, effects of experimental manipulations were excluded. Thus, for example, in any study that included a religious priming manipulation, we only analyzed the religion-prosociality relationship in the study's control condition. As mentioned, we also excluded conditions that assessed prejudice, as such studies may conflate general prosociality or antisociality with doctrinally-relevant characteristics.

Studies using two or more independent samples were coded as separate studies. The sign was reversed for effect sizes between religiosity and antisocial reports or behaviors, so that positive effect sizes apply to religiosity being related to either greater prosociality or lesser antisociality. Studies that included more than one measure of religiosity (e.g. a measure of affiliation and a scale of religiosity) and/or more than one relevant dependent variable completed by the same participants, were included as separate effects, but were coded as belonging to the same sample.

Analysis

Data were transformed into Fisher's r-to-z transformed correlation coefficients (Fisher, 1921) for analysis, using the *escalc* function of the *metafor* package for R (Viechtbauer, 2015). Corresponding sampling variances and confidence intervals were calculated using this same function. Unless otherwise noted, outputs were then converted back to Pearson's correlations for presentation in this paper.

I identified outliers using the recommendations of Viechtbauer and Cheung (2010). Specifically, outliers were identified if they had a standardized residual of greater than 3 in absolute magnitude. These studies were identified using the *metaoutliers* function of the *altmeta* function for R (Lin & Chu, 2016). This method yielded only one outlier. All analyses held robustly regardless of whether this outlier was excluded. Therefore, the analyses reported here do not have this outlier removed.

Meta-analyses are typically conducted using one of two types of models: a fixed-effect model or a random-effects model. Fixed-effect models assume that one true effect size exists and that all variance is due to sampling error. Random-effects models, on the other hand, allow there to be a true distribution of effect sizes (Borenstein, Hedges, Higgins,

& Rothstein, 2010). Because we hypothesized effects to vary as a function of method and sample, a random-effects model was used. Furthermore, as many of the effects analyzed in this study were nested within the same samples, a multi-level analysis was used. I conducted a three-level random-effects analysis using the recommendations of Assink and Wibbelink (2016), which accounts for random effects at the level of each individual effect and each individual sample. Whereas traditional methods of dealing with dependency in effect sizes simply ignore that dependency, average the effect sizes, or ignore one or more of the effect sizes, this method allows for dependency (such as two separate effects within one sample) to be analyzed by considering variance across all three levels. I furthermore conducted moderation analyses, also using the recommendations of Assink and Wibbelink (2016). These analyses were computed using the rma.mv function of the metafor package for R (Viechtbauer, 2015). Additionally, following the recommendation Assink and Wibbelink (2016), a Knapp and Hartung (2003) adjustment was applied for all analyses, which applies wider confidence intervals than do traditional methods, and which reduces the risk of unjustified significant results found in some analyses that use an unmodified Z distribution in assessing significance (see Li, Shi, & Roth, 1994; Ziegler, Koch, & Victor, 2001).

We expected publication bias to be a relatively small issue in this meta-analysis compared with many other meta-analyses for three reasons: We collected as many unpublished effects as we could; many of the studies from which the effects were obtained had competing hypotheses; and many of the included effects were not the primary effect of interest in the study from which they were obtained (and indeed, in many cases, the effect we extracted was not statistically significant). Nonetheless, I assessed publication bias by

two means. First, I visually inspected funnel plots of all included effects (Light & Pillemer, 1984). Funnel plots provide a graphical representation of each effect size and its corresponding standard error. Hypothetically, in a normal situation with little publication bias, highly powered effects with low standard error will cluster around the effect mean, whereas those with greater standard error will spread out from the mean symmetrically in each direction, producing a funnel-shaped plot. If asymmetry is found, this could be indicative of publication bias or a biased inclusion of studies.

Additionally, I conducted a sensitivity analysis using a priori weight functions, as recommended by Vevea and Woods (2005). As per the authors' recommendations, I did not use this analysis to provide a direct "test" or "correction" for publication bias, but rather to assess how robustly the effects hold under various assumptions with varying degrees of bias. The authors argue that corrections for publication bias require certain assumptions about the bias landscape that are fundamentally unknowable. Therefore, it is a preferable to investigate how the estimated effect is impacted by various assumptions of publication, rather than a single estimated effect—an argument echoed by McShane, Böckenholt, and Hansen (2016). These analyses apply a pre-selected weight to each effect based on its respective p-value—each weight representing a potential likelihood of that study being selected or included compared to a specified baseline. These analyses were run using the shinyapp provided by Vevea and Coburn at http://vevealab.shinyapps.op/WeightFunction/Model. Notably, this model of applying weights based on p-values is similar in some respects to other commonly used bias assessment methods like p-curve (Simonsohn, Nelson, Simmons, 2014) and p-uniform (van Assen, can Aert, & Wicherts, 2014). However, the Vevea and Woods model is preferable in

this context, as it is not limited by the requirement that all effects included in the analysis be statistically significant, and indeed many of the included effects are not statistically significant (McShane et al., 2016). Additionally, simulation studies (e.g. Carter, Schönbrodt ,Gervais, & Hilgard, 2017) have found weight function models like the one proposed by Vevea and Woods (2005) to produce less biased estimates than do p-curve, p-uniform, or other bias-correction methods like trim-and-fill (Duval & Tweedie, 2000) or PET-PEESE (Stanley & Doucouliagos, 2014).

Results

Overall Random-Effects

A random-effects analysis of all studies suggests a small but significant positive relationship between religiosity and prosociality, $r = .10$, $t(178) = 7.33$, $p < .001$, 95%CI [0.08, 0.13]. A forest plot of all effects, including the overall summary effect, can be seen in Figure 1. I next tested for heterogeneity of within-study variance and between-study variance. Likelihood-ratio tests suggest that the fit of the three-level model is significantly better than a model in which variance at level 2 or level 3 is fixed to 0. Therefore, there is greater variability in effect sizes than would be expected from sampling variance alone. Furthermore, I estimated the distribution over the three levels of the meta-analytic model using the formulas given by Cheung (2014). From these formulas, I derived that only 2.74% of the variance can be attributed to sampling variance, while 67.80% can be attributed to differences between effect sizes within samples and 29.46% can be attributed to variance between samples. Hunter and Schmidt (1990) suggest that heterogeneity may be regarded as substantial if less than 75% of the total variance can be attributed to sampling variance. The low sampling variance within this study is therefore indicative of a great deal of

heterogeneity. Given this heterogeneity, it was appropriate to search for moderators that may partially explain the high degree of variance not attributable to sampling variance.

Behavioral vs. Self-Report Measurement

Given previous research on religiosity and moral hypocrisy (Batson, 1976; Galen, 2012), it was hypothesized religiosity would be positively related to self-report measures, but not behavioral measures, of prosocial and antisocial behavior. Consistent with this literature, DV measurement method (self-report or behavioral) was a significant moderator of this effect, $F(1,175) = 18.54$, $p < .001$. Religiosity still positively predicted prosociality as measured by self-report, $k = 105$, $r = .15$, 95% CI [0.12, 0.18], an effect significantly greater than zero, $t(175) = 8.63$, $p < .001$. On the other hand, when measured using behavioral methods, the link between religiosity and prosocial behavior was substantially smaller ($k = 72$, $r = .04$, 95% CI [-0.00, 0.08]), and only marginally deviated from 0, $t(175) = 1.76$, $p = .08$. A summary of effects, broken down by measurement type, may be found in Figure 1. Residual heterogeneity remained high when accounting for this moderator, $QE(175) = 5{,}841.20$, $p < .001$, suggesting that heterogeneity is not fully accounted for by this moderator.

Pro vs. antisocial measures

I next compared religiosity's association to measures of prosocial behavior versus antisocial behavior. This moderation was not significant ($F(1, 177) = 2.07$, $p = .15$. However, while the relationship between religiosity and prosocial behavior was significant ($k = 151$, $r = .11$, $t(177) = 7.41$, $p < .001$, 95% CI [.08, .14]), the relationship between religiosity and the avoidance of antisociality was only marginally significant ($k = 28$, $r =$

.06, $t(177) = 1.83$, $p = .07$, 95% CI [-.01, .13]). It is worth noting, however, that unlike with the relatively well-balanced numbers of self-report and behavior-based studies, the meta-analysis had a much smaller number of studies investigating antisociality when compared to the collection of studies investigating prosociality. This lack of statistical power makes it difficult to draw conclusions about whether the lack of moderation and lack of religiosity-antisociality relationships are due to true non-effects or simply due to insufficient power. Therefore, I caution against drawing overly strong conclusions from this analysis about how religiosity relates to antisociality, and whether the religiosity-antisociality relationship significantly varies from the religiosity-prosociality relationship. When accounting for this potential moderator, residual heterogeneity remained high ($QE(2, 177) = 5{,}801.20$, $p < .001$).

Religiosity Measure

I next tested type of religiosity measure as a moderator. An omnibus test revealed no significant moderation ($F(2, 172) = 0.28$, $p = .76$). A significant relationship between religiosity and prosociality was found, regardless of how religiosity was measured (Table 2).

Student status

The student status of the sample was not found to be a moderator of the relationship, $F(1, 167) = 0.83$, $p = .36$. Religiosity was related to prosociality in both non-student samples, $k = 83$, $r = .12$, $t(167) = 5.91$, $p < .001$, 95% CI [.08, .15], and in student samples, $k = 86$, $r = .09$, $t(167) = 3.83$, $p < .001$, 95% CI [.04, .13].

WEIRD Status

Culture, as operationalized by WEIRD status, did not moderate the relationship between religiosity and prosocial behavior ($F(1, 174) = 0.88$, $p = .35$). Yet, due to the relative lack of studies using non-WEIRD samples (25 effects, vs 151 effects from WEIRD samples), it is again difficult to conclude whether the lack of significance is due to a true lack of moderation or due to a lack of statistical power. Nonetheless, the mean effect is significantly greater than zero both in WEIRD samples ($k = 151$, $r = .11$, $t(174) = 7.03$, $p < .001$, 95% CI [.08, .14]) and in non-WEIRD samples ($k = 25$, $r = .07$, $t(174) = 2.01$, $p = .046$, 95% CI [.00, .14].

Demographics

I next analyzed gender of the sample, operationalized as the percentage of the sample that was female, as a moderator. No significant moderation was found, $F(1, 151) = 0.07$, $p = .79$. Average age of the sample was also not a significant moderator of the effect, $F(1,121) = 0.01$, $p = .92$.

Meta-Regression

Finally, I conducted a meta-regression entering all potential moderators of interest as covariates (Table 3). Here, the self-report/behavioral measure distinction remains a significant moderator. The prosociality/antisociality distinction also emerges as a significant moderator, such that religiosity predicts prosociality more strongly than it predicts the avoidance of antisociality, holding all other covariates constant. Even accounting for all covariates, a great deal of residual heterogeneity remained, indicating that a substantial portion of the variance in the religiosity-prosociality relationship is

explained by factors for which the moderators are unable to account ($QE(104) = 704.01$, $p < .001$).

Tests for Publication Bias

I first sought to assess publication bias using funnel plots. The funnel plot of all included studies yielded a somewhat unusual pattern: highly-powered studies saw about as much variability from the mean as did less-powered studies (Figure 2a). Such a pattern may be explained by a high degree of heterogeneity, as variance in effects may be more fully explained by variation between samples or measures than it is by sampling variance (Sterne et al., 2011). Therefore, this funnel plot provides relatively little information about bias in our collection of studies, but it does provide further evidence of heterogeneity.

To partially address this heterogeneity, I split the funnel plots by the greatest source of heterogeneity: self-report versus behavioral measures. A funnel plot of self-report measures alone yields a similar pattern to that of the overall collection of studies (Figure 2b). Here again, a high degree of heterogeneity hinders the ability to assess symmetry. A funnel plot of behavioral effects, however, yields a more traditional funnel shape, perhaps suggesting relatively less heterogeneity in these measures (Figure 2c). A visual inspection of this plot does not suggest heavy asymmetry, providing little evidence of bias.

I next conducted a sensitivity analysis using a priori weight functions, as recommended by Vevea and Woods (2005). The models run for this sensitivity analysis are not properly able to deal with the fact that effects were nested within samples. As such, unlike in the overall analysis and moderator analyses, effects were averaged within samples, to yield only one aggregate effect per sample (or, in the moderator analysis, separate averages of within-sample effects were made for report and behavioral

measures). I applied weights based on pre-specified p-value cut-points, representing various assumptions regarding the likelihood that effects with varying p-values were represented in the sample. An unadjusted model in which effects of all p-values were regarded as equally likely to be published estimated a religiosity-prosociality relationship of $r = .11$, relatively consistent with the previous analysis. All adjusted models reduced a somewhat smaller effect, but under most assumptions of bias, the effect held reasonably robustly, typically varying from $r = .07$ to $r = .10$. Only under models that assumed extreme one-tailed bias (i.e. models that assumed a strong bias towards effects that show a significant positive relationship between religiosity and prosociality) did the estimated effect approach 0. However, I again argue that such an extreme one-tailed bias is highly unlikely, as we obtained a great deal of studies that did not show a positive relationship between religiosity and prosociality at the $\alpha = .05$ level. Additionally, theoretical arguments can and have been made that religiosity can, at times, be negatively related to prosociality or positively related to antisociality (or that they are simply unrelated). Therefore, it is unlikely that there is a strong bias in the field towards publishing or sharing positive religiosity-prosociality relationship effects over negative effects. I thus suggest that it is likely that bias has slightly inflated the estimated effect, as is typical of meta-analytic estimates, but the effect is unlikely to be fully explained by bias.

I next investigated how the behavioral/reported prosociality moderation held up under various a priori weight functions. While the magnitude of religiosity-prosociality relationship varied across assumptions, the moderation held up quite robustly. Across all tested assumptions, the religiosity-prosociality relationship, as indicated by Pearson's *r*, was at least .09 larger in magnitude in studies that used reported measures of prosociality

than in those that used behaviorally-measured prosociality. Under assumptions of one-tailed bias, the extent of the moderation was even greater than in the unadjusted model. Again, specific findings of these analyses can be found in the supplementary materials.

Finally, I considered the possibility that studies using self-report and behavioral measures may be under somewhat different selection pressures. I therefore ran a priori weight function models on each subset of the sample independently. Under most assumptions, estimates of the effect size among studies using behavioral measures held robustly, varying between r = .03 and r = .05. However, under a severe one-tailed bias, the effect reversed, reaching as low as r = -.02. Nonetheless, for the aforementioned reasons, I suggest that a severe one-tailed bias is less likely than a more moderate bias or a two-tailed bias. The effect of studies using report measures also held robustly. Here, even under severe one-tailed bias, the effect did not approach 0, reaching only as low as r = .07.

Discussion, Limitations, and Recommendations

Across 179 effects, I meta-analytically tested the prediction that dispositional religiosity is associated with prosocial behavior and the avoidance of antisocial behavior. Indeed, a positive association was found such that, across studies, religious people scored higher on measures of prosocial behavior and lower on measures of antisocial behavior than did the non-religious. However, the effect was somewhat small (r = .10), and significant heterogeneity was found across effects, making it difficult to determine the "true" effect. Two potential sources of this heterogeneity were identified and will be discussed in more detail: whether behavioral or report measures of the dependent variable were used and whether prosocial or antisocial outcomes were measured.

Additionally, I conducted sensitivity analyses to investigate how robustly the estimated effects hold up to different assumptions of publication or selection bias. I found that the estimated effects were somewhat mitigated under the assumption that some bias is present and greatly mitigated if substantial one-tailed bias is present. Fortunately, substantial one-tailed bias is unlikely in the included studies, as the collection of effects included unpublished effects, effects with competing hypotheses, and many effects that were not a study's primary finding of interest. Nonetheless, it is likely that the effects estimated in the random-effects models were slightly inflated by bias, as is typical of meta-analytical estimates.

While an initial goal of this meta-analysis was to conclusively answer the question of how strongly dispositional religiosity and prosocial behavior are related, it is unfortunately impossible to draw ironclad conclusions. This is primarily due to heterogeneity, and, to a lesser degree, because of the potential of bias. However, recent commentaries and simulation studies of meta-analyses have suggested that even in the absence of such heterogeneity, one should be cautioned against over-extrapolation from any one single point estimate of effect size (e.g. Carter et al., 2017; McShane et al., 2016; Vevea & Woods, 2005). The value of meta-analyses primarily derives from their ability to establish bounds of what this effect size may reasonably be expected to be, based on the existing literature, as well as to establish how characteristics of the existing literature may affect the estimated effect size. Following these recommendations, I will discuss the sources of heterogeneity identified in the meta-analysis and will provide recommendations for future research that may address these sources of heterogeneity.

Prosociality vs. Antisociality

While I did not have a firm hypothesis about how religiosity would differentially affect prosociality and antisociality, it was nonethless important to investigate this as a moderator. Exhibition of prosocial behavior and avoidance of antisocial behavior are meaningfully different constructs, even if they both fall under the umbrella of "moral behavior." In some circumstances, antisocial behavior and prosocial behavior have been found to be only weakly correlated (Krueger, Hicks, & McGue, 2001; Sage, Kavussanu, & Duda, 2006; Veenstra, Lindenberg, Oldehinkel, Winter, Verhulst, & Ormel, 2008) and predictors of moral behavior may differentially affect prosociality and antisociality (Gollwitzer, Schmitt, Schalke, Maes, & Baer, 2005; Malti & Krettenauer, 2013; Olthof, 2012).

While the distinction between prosociality and antisociality did not on its own emerge as a significant moderator of the effect, there was a trend in the data such that religiosity was more strongly associated with prosociality than the avoidance of antisociality. Additionally, while the relationship between religiosity and prosociality was significant, the relationship between religiosity and the avoidance of prosociality was only marginal. Complicating the interpretation of these findings, there was a relative lack of studies investigating antisociality in this dataset. Such lack of statistical power would make it difficult to detect a moderation, if one exists, and it would also make it difficult to capture a relation between religion and antisociality, if that exists. Nonetheless, despite this lack of power, the prosocial/antisocial distinction emerged as a significant moderator when entered into a model with all other covariates. Thus, there is some evidence that religiosity may be more strongly associated with prosociality than the avoidance of antisociality. This is not altogether surprising, as a substantial literature has found that certain aspects of religiosity, like fundamentalist beliefs, predict prejudice and hostility, particularly towards

outgroups (e.g. Allport & Kramer, 1946; Johnson, Rowatt, & LaBouff, 2012; Koopmans, 2015). While we excluded studies that directly and intentionally measured prejudice in order to avoid contamination from political attitudes, it is certainly feasible that such outgroup hostility may have still factored into the measures of antisociality. Furthermore, the fact that we did not directly include measures of prejudice suggests that this moderation may actually underestimate religiosity's differential effect on prosociality and antisociality, when all forms of behaviors are considered. Given these findings, I recommend that prosociality and antisociality continue to be explored as separable constructs within the context of religiosity research, rather than merely opposite sides of the same coin. Further research on how, when, and why religiosity might differentially act upon these constructs is warranted.

Behavioral vs. Self-Report Measures of Prosociality

The strongest moderator of the religion-prosociality relationship, and one that was hypothesized, was the distinction between religiosity's effect on behavioral and self-report measures of prosociality. Most participants of studies included in the analysis completed self-report measures of prosocial or antisocial behavior, usually in large surveys. However, a sizable minority were evaluated by behavioral measures—usually economic games, or other tightly-controlled experimental protocols. It has long been argued that measures of the relationship between religiosity and prosocial behavior are differentially affected by these fundamentally different methods (Batson et al., 1978; Galen, 2012; Saroglou, 2006). This meta-analysis offers clear empirical support for this observation. Among studies that used a self-report measure of prosociality, the relationship between religiosity and prosociality was highly significant. However, among studies using a behavioral measure of

prosociality, the effect was much smaller, and only marginally significant. Indeed there is theoretical reason to doubt whether even this small effect of religiosity on prosociality is due to religiosity *per se*, as recent research has found the relationship between religiosity and economic game decisions to be heavily mediated by intuitive thinking (Stagnaro, Arechar, & Rand, 2018) and the assumption that the recipient will be a member of one's religious ingroup (Kelly & Shariff, 2019). Therefore, if such factors were accounted for, the moderation by measurement technique may be even greater than was estimated. Furthermore, the moderation by DV measurement is quite robust, holding up even under assumptions of heavy publication or selection bias. Nonetheless, despite the strength and robustness of this moderation, the analyses reported here are incapable of explaining exactly *why* this discrepancy occurs. It may be either that the behavioral or self-report measures are capturing the "true" relationship. Here, I offer three possible explanations for this disparity (any or all of which might be at play) and propose future lines of research to address these issues.

1. The limitations of lab-based methods. Lab-based, behavioral methods have several benefits over self-report methods. They have a high degree of control and context-independence (anonymizing participants, limiting information, and taking place in neutral settings), thereby leading to a high degree of internal validity. But what these methods offer in regard to internal validity may come at the expense of external validity. It is possible that the lab-based methods that have been used to assess this relationship have failed to fully capture a true religious prosociality advantage that genuinely occurs outside of the lab. Perhaps self-report measures are indicative of a real behavioral difference (rather than a mere artefact of reporting biases), to which lab-based methods are not

adequately sensitive. When one considers prototypical examples of religious prosociality, one might imagine such activities as missionary work, work in soup kitchens, or donation drives. It is reasonable to assume that these types of behaviors do not perfectly correlate with behavior like contributions in a lab-based dictator game.

Furthermore, as previously noted, lab-based methods typically go to great lengths to ensure anonymity; however, *who* the recipient is may be of importance in religious prosociality. Saroglou (2006) suggested that religious people may be less willing to donate to those who are not part of their group or who threaten their identity. Therefore, ambiguity of the recipient's identity may lead to hesitancy to donate. Lab-based studies also tend to be individualistic experiences. People tend to participate in studies alone, or, if they participate with someone else, it is usually a stranger. Religious prosociality, on the other hand, is often a community activity, or is the result of the systematically-structured opportunities offered by one's membership in a church (Park & Smith, 2000; Saroglou, 2006). Finally, it may be that religious participants truly are more prosocial, but they "max out" their prosociality in their everyday activities. When taking part in a relatively artificial laboratory experiment, they may feel little obligation to give more of their money or effort than they already have.

Recommendations. In the lab, researchers should take care to make assessments of prosociality resemble real-world prosocial decisions as much as possible. One specific means of measuring prosociality within the lab is to assess charitable donations, while giving a choice of where to donate. In situations where the lack of identifying recipient information is hindering donation behavior, participants may be more likely to donate if charities are directly identifiable, and if they have an option to pick a charity matching their

interests. Furthermore, to test the prediction that religious people will be more prosocial, but only in a parochial fashion, ingroup or outgroup affiliation of the recipient could be made salient.

Researchers may also move outside the artificial settings of the lab, while still assessing behavior. An experience sampling method gives researchers access to participants' behavior in their everyday lives (Csikszentmihalyi & Larson, 2014). Using this method, researchers could evaluate time spent volunteering or performing other prosocial actions as a function of religiosity. One experience sampling-based study has indeed tested the relationship between religiosity and moral behavior and found no difference between the religious and the non-religious on reported morality or immorality (Hofmann, Wisneski, Brandt, & Skitka, 2014). However, one limitation of this design was that participants were simply asked if they had engaged in a moral or immoral behavior within the last hour. As religiosity is conflated with the types of behaviors that are placed within the moral behavior, I recommend follow-up research that tests the relationship between religiosity and specific prosocial or antisocial behaviors. Alternatively, researchers could evaluate the religiosity of those who respond to genuine donation requests, assessing donation likelihood and amount as a function of religiosity (see, e.g. Malhotra, 2010). This method has its limitations as well, however, as the types of organizations or causes people will donate to vary as a function of religiosity.

2. Religious self-enhancers. The next possibility is simply that self-report measures are failing to capture what they are intended to capture, while being highly susceptible to self-enhancement. While lab-based measures of prosociality may suffer from issues of validity, self-report measures of ethical and unethical behavior suffer from greater

threats to validity. Indeed, self-ratings of ethical behavior often fail to capture true levels of ethical or unethical behavior (Alem, Eggert, Kocher, & Ruhinduka, 2018).

On the other hand, while self-report measures of ethicality may fail to capture true ethicality, they unfortunately capture one's propensity for socially desirable responding or self-enhancement (Fernandes & Randall, 1992). Troublingly, there is also an extensive literature arguing that religious people self-enhance more than do nonbelievers (Allport, 1966; Batson, 1976; James, 1902; Saroglou, 2006). Sedikides and Gebauer (2010) demonstrated religiosity's relationship with self-enhancement meta-analytically and have even argued that religion may be a form of satisfying a dispositional drive to self-enhance. Gebauer, Sedikides, & Schrade (2017) further demonstrated that religious people (or at least Christians) most significantly overestimate themselves on areas closely linked to their self-concept, a category that certainly includes prosocial behavior. As such, it may be that behavioral measures' tiny effect represents the true association. Conversely, self- reported prosociality may simply capture a religious self-enhancement effect. In other words, religious people may not be being honest about their true level of prosocial behavior, or they are deluding themselves into believing they are more prosocial than they are.

Recommendations. Controlled laboratory studies are an effective, and commonly used, method of avoiding the problems characteristic of self-report measures. However, as previously noted, such methods are not without their flaws when used in the domain of religiosity and prosociality. Nonetheless, many of the aforementioned recommendations for avoiding the issues of lab-based methods would also be useful to address this concern. Assessing actual donation or volunteering behavior eliminates the problem of dishonesty. Similarly, experience sampling methods lessen the effect of self-serving biases. While they

are still subject to some self-enhancement or social desirability risk, as participants can simply lie about what they were doing at a given time, it seems relatively less likely a participant would lie outright than to simply enhance or exaggerate their response on a self-report item.

Additionally, researchers could lessen the risk of self-enhancing biases by taking peer-reports of prosocial behavior, rather than self-reports. One study using this method demonstrated that religious people were seen to be more altruistic and empathic by their peers than were nonreligious people (Saroglou et al., 2005). While less influenced by self-enhancement, this method is, however, also far from perfect. High self-enhancers may be more likely to tell others about their prosocial behavior or inflate how prosocial they appear to others, which would allow peers to have greater knowledge of their prosociality than would peers of low self-enhancers. Furthermore, peers of religious people are more likely to be religious themselves and may self-enhance by exaggerating the goodness of their peers. As such, while peer ratings may be a valuable contribution to the literature, I suggest the best solution to avoid self-serving biases is to collect representative and generalizable behavioral measures of prosociality.

3. The tautological explanation. Just as self-reports of prosociality are subject to self-enhancement, so too are self-reports of religiosity itself. Measures of church attendance are notoriously exaggerated (Hadaway, Marler, & Chaves, 1998), and even simple reports of theism vs. atheism often substantially overestimate belief (Gervais & Najle, 2018). As religion is still highly normative in much of the world, and belief is seen as necessary for morality by many (Pew Research Center, 2014), it is no wonder that people are inclined to exaggerate their belief. However, this presents a significant problem for the

measurement of religiosity. All studies in this meta-analysis, and virtually all religiosity studies in the literature, use self-report measures of religiosity. Thus, it remains possible that many studies demonstrating a positive relationship between religiosity and some positive trait measured by self-report (including prosociality) are at least partially the product of a tautological measurement overlap; those who self-enhance more (on a measure of religiosity), self-enhance more (on the trait of interest). I therefore dub this potential explanation the "tautological explanation".

Recommendations. I recommend investigating the effect of self-enhancement on reported religiosity. If indeed self-enhancement is related to reported religiosity more than actual religiosity, one might expect religiosity to be more closely related to self-enhancement in populations where religiosity is highly normative than where it is not. While their explanation of the phenomenon is different, Sedikides and Gebauer (2010) offer support for this prediction. Intrinsic religiosity was found to be more strongly related to self-enhancement in the United States than in the United Kingdom or in Canada. The relationship was also stronger in American Christian universities than in American secular universities. A stronger way to test this prediction would be to directly manipulate the normativity or desirability of religiosity within a controlled setting and investigate religiosity's relationship with other positive, self-reported traits as a function of religious normativity and self-enhancement. Simply assessing religiosity in a face-to-face manner vs. an anonymous manner would also provide insight into its susceptibility to social desirability. While anonymous responses do not eliminate social desirability concerns, the effect is less strong there than in more direct interactions (Joinson, 1999).

I also recommend the direct assessment of religiosity and self-reported prosociality relationships across varying degrees of social desirability in naturalistic settings. If the tautological explanation is accurate, one would expect the religion-prosociality link to be stronger in places in which religiosity is socially desirable than where it is not. I am aware of one paper that has directly used this type of method: Stavrova and Siegers (2014) assessed the relationship between religiosity, reported prosociality, and reported antisociality across 70 countries. However, they found the opposite effect of what might be predicted by the tautological explanation. There was a significantly greater religious advantage in reported prosociality and antisociality in countries in which religion is *not* societally enforced than where it is. This paper therefore does argue that religious people are prone to report being more prosocial, but only when those people are intrinsically, rather than extrinsically, motivated to participate in one's religion.

Nonetheless, if self-enhancement is in fact affecting measures of religiosity, it is prudent to measure religiosity by means less subject to self-serving biases. Peer reports of religiosity are a potential solution, but peer knowledge of one's religiosity is dependent on the peer's knowledge of the subject's belief, attendance, or practice, which is, in turn, subject to how much the participant reveals about these factors. The unmatched count technique is a promising method of capturing more honest assessments of religious belief (Gervais & Najle, 2018). This method allows completely anonymous reporting of whether some socially undesirable belief or attitude is true (e.g. "I do not believe in God"). This method requires large samples, however, and only offers group-level information. Therefore, its value in directly investigating the link between religiosity and prosociality is limited.

Researchers could further pursue behavioral measures of religiosity. Again, an experience sampling method (Csikszentmihalyi & Larson, 2014) would be of use. By this method, researchers could more directly assess frequency of religious attendance or practice. Alternatively, they may assess those leaving worship services and compare their results on measures of prosociality or other variables of interest to those who were not at the service, thus assessing a true measure of religious attendance, rather than a self-report one.

Finally, experimental methods are also valuable. While manipulating one's dispositional religiosity is challenging, if not impossible, it is possible to manipulate religious context or salience. Religious priming work does exactly this (Willard, Shariff & Norenzayan, 2016). While this method does not directly address the effect of dispositional religiosity (the variable of interest within this meta-analysis), it does provide evidence of the impact religion may have on prosocial behavior, independent of self-enhancement concerns. Furthermore, religious priming only has an effect on prosociality for those who are dispositionally religious (Shariff et al., 2015), providing evidence that dispositional religiosity is at least indirectly related to prosocial behavior. However, given that there are many routes to prosociality, some of which may be more effective at encouraging the non-religious than the religious (e.g. Saslow et al., 2013), the mere fact that the religious act more prosocially in religious situations gives no strong indication that they act more prosocially overall.

Conclusions regarding the self-report/behavior discrepancy. Determining the "true" relationship between dispositional religiosity and prosociality depends on resolving the question of whether behavioral or self-report measures better estimate real-world

prosocial behavior. If behavioral measures are more reliable, the relationship may be quite small indeed. However, if self-report measures are more valid, the relationship may be as high as r = .18 (the 95% confidence upper bound on the relationship between religiosity and self-report measures of prosociality). While this correlation may still seem small, it is a stronger correlation than many other meta-analytically estimated religious effects (Table 4) and only slightly smaller than the average effect in social psychology (r = .21; Richard, Bond, & Stokes-Zoota, 2003). Unfortunately, however, the cause of the self-report/behavioral measure discrepancy remains unresolved. Behavioral measures may obfuscate a true religiosity-prosociality relationship, or self-enhancement may lead to a spurious relationship, either as an intrinsic correlate of religiosity, or as a mutually-confounding third variable. A satisfactory answer to whether religious people are indeed meaningfully more prosocial awaits resolution of this discrepancy. The above recommendations may help bring about such a resolution and allow greater clarity on whether the dispositionally religious are in fact more prosocial than those who are not.

Other limitations and recommendations

Parochiality. The distinction between religious ingroup and outgroup is crucial to theories of religious prosociality, and evidence regarding parochial religious prosociality and antisociality has been mixed (Dunkel & Dutton, 2016; Ginges, Sheikh, Atran, & Argo, 2016; Preston & Ritter, 2013). As such, one desired goal of this meta-analysis was originally to investigate the groupish nature of religious prosociality, or lack thereof. However, in most cases, we were unable to confidently code whether recipients of prosociality were perceived to be ingroup or outgroup members. Even in papers that more clearly delineated ingroup versus outgroup members, group membership was rarely defined by religious

affiliation. Instead, participants would report helping intentions towards, for example, friends versus unknown community members. Examining whether religious individuals differentially help close others addresses an important question, but it does not tap into the religious ingroup-outgroup distinction at the heart of religious prosociality theories. Therefore, it was impossible to reasonably compare whether, and to what degree, religious-ingroup prosociality differs from prosociality directed toward the outgroup. At the moment, this is an intriguing question that remains unanswered and ripe for future research.

Representativeness. This analysis is limited by the general lack of religious diversity of participants and the Western-centric focus of the bulk of the available literature. When I refer to religiosity in the included studies and my own results, I assume that religion's relation to prosociality is reasonably stable across faith groups. This is consistent with theory and some research. However, there are myriad differences between Western Christians, who make up the vast majority of religious participants in this analysis, and people from other cultures and faith traditions that may very well be relevant. For example, while it is true and important that all world religions emphasize prosocial values, the social norms and the enforcement mechanisms of religious values are drastically different in Eastern contexts such as China (Sun, 2013). In China, which holds about as much of the human population as North America and Europe combined, people tend to practice religious rituals but do not identify as religious, partially because claiming an affiliation is often seen as akin to claiming perfection and is therefore socially undesirable. Certainly, the fact that China is now an officially atheist country has further decreased the social desirability of claimed religious affiliation (Dillon, 2001). This pattern is in stark

contrast to the tendency among American participants to over-report both affiliation (Gervais & Najle, 2018) and service attendance (Hadaway et al., 1998) because both are socially desirable. Muslims make up the second largest religious group in the world but its nearly 2 billion adherents are all but absent from this meta-analysis. Finally, cross-cultural research that has compared prosociality in a variety of religions across eight small-scale societies has found considerable variation that systematically covaried with the types of religious beliefs (Purzycki, 2016). This suggests that broader between-religious differences (a topic on which the Psychology of Religion has been generally quite quiet) could be significant, meaningful and informative.

Unfortunately, logistical and other challenges have discouraged Western psychologists of religion from extensive investigation of religious prosociality among non-Christians and non-WEIRD populations. Due to their relative paucity in the literature, current understanding of how non-Christian religions or non-Western contexts influence the relationship between religion and prosociality is not yet sophisticated enough to draw firm conclusions about the generalizability of these findings. It is therefore essential to make a diversity of samples a priority in future research, particularly in the context of the Psychology of Religion.

Unexplained heterogeneity. Even after accounting for all potential moderators of interest, residual heterogeneity in the model remained quite high. This is not at all surprising. Both religiosity and prosociality are broad, complex, overarching constructs. It is quite reasonable to assume that different aspects of religiosity differentially act upon different aspects of prosociality, and that these effects are highly dependent on situation. I fully expect that there are meaningful moderators of the religiosity-prosociality

relationship that are not accounted for in this analysis. Therefore, further research into the more granular aspects of religiosity and prosociality are not only warranted, but crucial.

Conclusion

In the present work, I conducted a meta-analysis to investigate how dispositional religiosity relates to prosociality. A significant relationship was found, such that religiosity predicted increased prosociality. However, interpretation of this finding was complicated by significant heterogeneity across effects beyond mere sampling variance. One substantial contributor to this heterogeneity was identified: the distinction between self-report and behavioral measures of prosociality. A significant relationship of $r = .15$ was estimated between religiosity and self-report measures of prosociality, but only a very small, marginally-significant effect was found between prosociality and behavioral measures of religiosity. Interpretation of this moderation is challenging, but I have provided three possible, non-mutually-exclusive interpretations of this discrepancy and have outlined recommendations for future research. Specifically, this moderation may be due to the behavioral measures' inability to capture a true effect, self-report measures' susceptibility to religious self-enhancement, or a spurious correlation between religiosity and prosociality which are both cause by self-enhancement. In summary, despite this comprehensive view, the current status of the literature still does not allow a confident conclusion as to whether religious people are meaningfully more prosocial than the non-religious. However, this meta-analysis identifies a likely upper bound of the strength of this relationship, if one exists, (approximately $r = .18$) and elucidates systematic limitations within the current literature, while also providing prospective remedies and future directions.

References

References marked with an asterisk indicate studies included in the meta-analysis

*Ahmed, A. M., & Salas, O. (2008). In the back of your mind: Subliminal influences of religious concepts on prosocial behavior. *Working Papers in Economics,* 331, 1-25.

*Ahmed, A. M., & Salas, O. (2009). Is the hand of God involved in human cooperation? *International Journal of Social Economics*, 36(1/2), 70–80.

Alem, Y., Eggert, H., Kocher, M. G., & Ruhinduka, R. D. (2018). Why (field) experiments on unethical behavior are important: Comparing stated and revealed behavior. *Journal of Economic Behavior & Organization,* 156, 71-85.

Allen, R. O., & Spilka, B. (1967). Committed and consensual religion: A specification of religion-prejudice relationships. *Journal for the Scientific Study of Religion*, 191-206.

Allport, G. W. (1966). The religious context of prejudice. *Journal for the Scientific Study of Religion*, 5(3), 447–457.

Allport, G. W., & Kramer, B. M. (1946). Some roots of prejudice. *The Journal of Psychology*, 22(1), 9-39.

Allport, G. W., & Ross, J. M. (1967). Personal religious orientation and prejudice. *Journal of personality and social psychology*, 5(4), 432-443.

Anderson, C. A., & Bushman, B. J. (1997). External validity of "trivial" experiments: The case of laboratory aggression. Review of General Psychology, 1(1), 19-41.

*Anderson, L., Mellor, J., & Milyo, J. (2010). Did the devil make them do it? The effects of religion in public goods and trust games. *Kyklos*, 63(2), 163-175.

Assink, M., Wibbelink, C. J., & others. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, 12(3), 154–174.

Baier, C. J., & Wright, B. R. (2001). "If you love me, keep my commandments": A meta-analysis of the effect of religion on crime. *Journal of Research in Crime and Delinquency*, *38*(1), 3-21.

Barrio, V. del, Aluja, A., & García, L. F. (2004). Relationship between empathy and the Big Five personality traits in a sample of Spanish adolescents. *Social Behavior and Personality: An International Journal*, 32(7), 677–681.

Batson, C. D. (1976). Religion as Prosocial: Agent or Double Agent? *Journal for the Scientific Study of Religion*, 15(1), 29–45. https://doi.org/10.2307/1384312

Batson, C. D. (2012). A history of prosocial behavior research. In A.W. Kruglanski & W. Stroebe (Eds.) *Handbook of the history of social psychology* (pp. 243-264). Retrieved from https://books.google.com/books?hl=en&lr=&id=MOlUSV2koMcC&oi=fnd&pg=PA242&ots=PMqAjayK05&sig=5apl5KopbsVFjEUJ2CuJWq3fuS0#v=onepage&q&f=false

Batson, C. D., Naifeh, S. J., & Pate, S. (1978). Social desirability, religious orientation, and racial prejudice. *Journal for the Scientific Study of Religion*, 31–41.

*Batson, C. D., Oleson, K. C., & Weeks, J. L. (1989). Religious prosocial motivation: is it altruistic or egoistic?. *Journal of Personality and Social Psychology*, 57, 873-884.

Baumeister, R. F., Bauer, I. M., & Lloyd, S. A. (2010). Choice, free will, and religion. *Psychology of Religion and Spirituality*, 2(2), 67-82.

*Benjamin, D. J., Choi, J. J., & Fisher, G. W. (2010). *Religious identity and economic behavior* (No. w15925). National Bureau of Economic Research.

*Berger, I. E. (2006). The influence of religion on philanthropy in Canada. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 17(2), 110-127.

Benz, M., & Meier, S. (2008). Do people behave in experiments as in the field?—evidence from donations. *Experimental Economics*, 11(3), 268–281.

Bergin, A. E. (1983). Religiosity and mental health: A critical reevaluation and meta-analysis. *Professional Psychology: Research and Practice*, *14*(2), 170-184.

*Blogowska, J., Lambert, C., & Saroglou, V. (2013). Religious prosociality and aggression: It's real. *Journal for the Scientific Study of Religion*, 52(3), 524–536.

*Blogowska, J., & Saroglou, V. (2011). Religious fundamentalism and limited prosociality as a function of the target. *Journal for the Scientific Study of Religion*, 50(1), 44–60.

*Bloodgood, J. M., Turnley, W. H., & Mudrack, P. (2008). The influence of ethics instruction, religiosity, and intelligence on cheating behavior. *Journal of business ethics*, 82(3), 557-571.

*Bock, D. C., & Warren, N. C. (1972). Religious belief as a factor in obedience to destructive commands. *Review of Religious Research*, 185–191.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97–111.

*Brañas-Garza, P., Espín, A. M., & Neuman, S. (2014). Religious pro-sociality? Experimental evidence from a sample of 766 Spaniards. *PloS one*, *9*(8), e104685.

*Brooks, A. C. (2003). Religious faith and charitable giving. *Policy Review*, (121), 39-51.

*Brown, R. P., Barnes, C. D., & Campbell, N. J. (2007). Fundamentalism and forgiveness. *Personality and Individual Differences*, 43(6), 1437–1447.

Brown-Iannuzzi, J. L., McKee, S., & Gervais, W. M. (2018). Atheist horns and religious halos: Mental representations of atheists and theists. *Journal of Experimental Psychology: General*, 147(2), 292-297. http://dx.doi.org/10.1037/xge0000376

Bushman, B. J., Ridge, R. D., Das, E., Key, C. W., & Busath, G. L. (2007). When God sanctions killing: Effect of scriptural violence on aggression. *Psychological Science*, 18(3), 204–207.

Carlo, G., Okun, M. A., Knight, G. P., & de Guzman, M. R. T. (2005). The interplay of traits and motives on volunteering: Agreeableness, extraversion and prosocial value motivation. *Personality and Individual Differences*, 38(6), 1293–1305.

Carter, E., Schönbrodt, F., Gervais, W. M., & Hilgard, J. (2017). Correcting for bias in psychology: A comparison of meta-analytic methods. Pre-print. Retreived from https://osf.io/jn6x5/

Cline, V. B., & Richards, J. M. (1965). A factor-analytic study of religious belief and behavior. *Journal of Personality and Social Psychology*, *1*(6), 569-578.

*Clobert, M., & Saroglou, V. (2013). Intercultural non-conscious influences: Prosocial effects of Buddhist priming on Westerners of Christian tradition. *International Journal of Intercultural Relations*, 37(4), 459–466.

*Clobert, M., Saroglou, V., & Hwang, K. K. (2015). Buddhist concepts as implicitly reducing prejudice and increasing prosociality. *Personality and Social Psychology Bulletin*, *41*(4), 513-525.

Cohn, A., & Maréchal, M. A. (2016). Laboratory measure of cheating predicts school misconduct. *The Economic Journal*, *128*(615), 2743-2754.

Crabtree, S. (2010). Religiosity highest in world's poorest nations: United States is among the rich countries that buck the trend. Retrieved from https://news.gallup.com/poll/142727/religiosity-highest-world-poorest-nations.aspx

Crandall, V. C., & Gozali, J. (1969). The social desirability responses of children of four religious-cultural groups. *Child development*, *40*(3), 751-762.

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. *In Flow and the foundations of positive psychology* (pp. 35–54). Springer. Retrieved from http://link.springer.com/10.1007/978-94-017-9088-8_3

*Curlin, F. A., Dugdale, L. S., Lantos, J. D., & Chin, M. H. (2007). Do religious physicians disproportionately care for the underserved?. *The Annals of Family Medicine*, 5(4), 353-360.

Darley, J. M., & Batson, C. D. (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, 27(1), 100–108.

Dawkins, R. (2016). *The god delusion*. Random House.

DeBono, A., Shariff, A. F., Poole, S., & Muraven, M. (2017). Forgive us our trespasses: Priming a forgiving (but not a punishing) god increases unethical behavior. *Psychology of Religion and Spirituality*, 9(S1), S1-S10.

Dillon, M. 2001. *Religious minorities and China*, London: Minority Rights Group.

*Dugas, M., Bélanger, J. J., Moyano, M., Schumpe, B. M., Kruglanski, A. W., Gelfand, M. J., ... & Nociti, N. (2016). The quest for significance motivates self-sacrifice. *Motivation Science*, *2*(1), 15-32.

Dunkel, C. S., & Dutton, E. (2016). Religiosity as a predictor of in-group favoritism within and between religious groups. *Personality and Individual Differences*, 98, 311–314.

*Durfeld, C., Martin, R. Washburn, A. Wilson, A. (2016). An analysis of prosocial behavior in college students. *Metamorphosis*. Retrieved from https://metamorphosis.coplac.org/index.php/metamorphosis/article/view/21

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(2), 455-463.

*Eckel, C. C., & Grossman, P. J. (2004). Giving to secular causes by the religious and nonreligious: An experimental test of the responsiveness of giving to subsidies. *Nonprofit and Voluntary Sector Quarterly*, 33(2), 271–289.

*Fehr, R., Gelfand, M. J., & Nag, M. (2010). The Road to Forgiveness: A Meta-analytic Synthesis of Its Situational and Dispositional Correlates. *Psychological Bulletin*, *136*(5), 894-914.

Fernandes, M. F., & Randall, D. M. (1992). The Nature of Social Desirability Response Effects in Ethics Research. *Business Ethics Quarterly*, *2*(2), 183-205.

Finkelstein, J. J. (1968). The Laws of Ur-Nammu. *Journal of cuneiform studies*, *22*(3-4), 66-82.

Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3-32.

*Furrow, J. L., King, P. E., & White, K. (2004). Religion and positive youth development: Identity, meaning, and prosocial concerns. *Applied Developmental Science*, 8(1), 17–26.

Galen, L. W. (2012). Does religious belief promote prosociality? A critical examination. *Psychological Bulletin*, 138(5), 876-906.

*Galen, L. W., Sharp, M., & McNulty, A. (2014). Nonreligious group factors versus religious belief in the prediction of prosociality. *Social Indicators Research*, *122*(2), 411-432.

Gebauer, J. E., Sedikides, C., & Schrade, A. (2017). Christian Self-enhancement. *Journal of Personality and Social Psychology*, *113*(5), 786-809.

Gervais, W. M. (2014). Everything is permitted? People intuitively judge immorality as representative of atheists. *PloS One*, 9(4), e92302.

Gervais, W. M., & Najle, M. B. (2018). How many atheists are there?. *Social Psychological and Personality Science*, *9*(1), 3-10.

Gervais, W. M., Shariff, A. F., & Norenzayan, A. (2011). Do you believe in atheists? Distrust is central to anti-atheist prejudice. *Journal of Personality and Social Psychology*, 101(6), 1189.

Ginges, J., Hansen, I., & Norenzayan, A. (2009). Religion and support for suicide attacks. *Psychological Science*, 20(2), 224–230.

Ginges, J., Sheikh, H., Atran, S., & Argo, N. (2016). Thinking from God's perspective decreases biased valuation of the life of a nonbeliever. *Proceedings of the National Academy of Sciences*, 113(2), 316–319.

Gollwitzer, M., Schmitt, M., Schalke, R., Maes, J., & Baer, A. (2005). Asymmetrical effects of justice sensitivity perspectives on prosocial and antisocial behavior. *Social Justice Research*, 18(2), 183-201.

*Gomes, C. M., & McCullough, M. E. (2015). The effects of implicit religious primes on dictator game allocations: A preregistered replication experiment. *Journal of Experimental Psychology: General*, 144(6), e94.

Gordon, M. E., Slade, L. A., & Schmitt, N. (1986). The "science of the sophomore" revisited: From conjecture to empiricism. *Academy of Management Review*, 11(1), 191–207.

Gough, H. G. (1951). Studies of social intolerance: I. Some psychological and sociological correlates of anti-Semitism. *The Journal of Social Psychology*, 33(2), 237-246.

Graham, J., & Haidt, J. (2010). Beyond Beliefs: Religions Bind Individuals Into Moral Communities. *Personality and Social Psychology Review*, 14(1), 140–150. https://doi.org/10.1177/1088868309353415

Greenberg, J. (1987). The college sophomore as guinea pig: Setting the record straight. *Academy of Management Review*, 12(1), 157–159.

*Greer, T., Berman, M., Varan, V., Bobrycki, L., & Watson, S. (2005). We are a religious people; we are a vengeful people. *Journal for the Scientific Study of Religion*, 44(1), 45–57.

*Gutierrez, I. A., & Mattis, J. S. (2014). Factors predicting volunteer engagement among urban-residing African American women. *Journal of Black Studies*, 45(7), 599-619.

Hackney, C. H., & Sanders, G. S. (2003). Religiosity and mental health: A meta–analysis of recent studies. *Journal for the Scientific Study of Religion*, *42*(1), 43-55.

Hadaway, C. K., Marler, P. L., & Chaves, M. (1998). Overreporting church attendance in America: Evidence that demands the same verdict. *American Sociological Review*, 63(1), 122–130.

Hall, D. L., Matz, D. C., & Wood, W. (2010). Why don't we practice what we preach? A meta-analytic review of religious racism. *Personality and Social Psychology Review*, 14(1), 126–139.

*Hansen, D. E., Vandenberg, B., & Patterson, M. L. (1995). The effects of religious orientation on spontaneous and nonspontaneous helping behaviors. *Personality and Individual Differences*, 19(1), 101–104.

*Hardy, S. A., Nadal, A. R., & Schwartz, S. J. (2017). The Integration of Personal Identity, Religious Identity, and Moral Identity in Emerging Adulthood. *Identity*, *17*(2), 96-107.

Hawley, P. H. (2003). Prosocial and Coercive Configurations of Resource Control in Early Adolescence: A Case for the Well-Adapted Machiavellian. *Merrill-Palmer Quarterly*, *49*(3), 279-309.

Henrich, J., Heine, J. S., & Norenzayan, A. (2010). The weirdest people in the world. *Behavioral and Brain Sciences*, 33, 61–83.

Hitchens, C. (2007). God is not great: How religion poisons everything. Toronto: McClelland & Steward.

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, *345*(6202), 1340-1343.

Holdrege, B. A. (2004). Dharma. In *The Hindu World* (pp. 225-260). Routledge.

*Horstmann, K. (2017). [Religiosity and charitable giving]. Unpublished raw data.

*Hunsberger, B., & Platonow, E. (1986). Religion and helping charitable causes. *The Journal of Psychology*, 120(6), 517-528.

*Jackson, J. C., & Gray, K. (2018). When a Good God Makes Bad People: Testing a Theory of Religion and Immorality. *Journal of Personality and Social Psychology*.

James, W. (1902). The varieties of religious experience: A study in human nature. Cambridge: Harvard University Press.

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology*, 66(1), 206.

Johnson, D., & Krüger, O. (2004). The Good of Wrath: Supernatural Punishment and the Evolution of Cooperation. *Political Theology*, 5(2), 159–176. https://doi.org/10.1558/poth.2004.5.2.159

Johnson, M. K., Rowatt, W. C., & LaBouff, J. P. (2012). Religiosity and prejudice revisited: In-group favoritism, out-group derogation, or both?. *Psychology of Religion and Spirituality*, 4(2), 154.

Joinson, A. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers*, 31(3), 433–438.

Jones, S. E., Miller, J. D., & Lynam, D. R. (2011). Personality, antisocial behavior, and aggression: A meta-analytic review. *Journal of Criminal Justice*, 39(4), 329–337.

Kelly, J.M., & Shariff, A. (2019). [Ingroup assumptions as mediation of religious dictator game giving]. Write-up of findings. https://osf.io/hvuc3/

Kirkpatrick, C. (1949). Religion and humanitarianism: A study of institutional implications. *Psychological Monographs: General and Applied*, 63(9).

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in medicine*, 22(17), 2693-2710.

Koopmans, R. (2015). Religious fundamentalism and hostility against out-groups: A comparison of Muslims and Christians in Western Europe. *Journal of Ethnic and Migration Studies*, 41(1), 33-57.

Krueger, R. F., Hicks, B. M., & McGue, M. (2001). Altruism and antisocial behavior: Independent tendencies, unique personality correlates, distinct etiologies. *Psychological Science*, 12(5), 397–402.

*Kulow, K., & Kramer, T. (2016). In pursuit of good karma: When charitable appeals to do right go wrong. *Journal of Consumer Research*, 43(2), 334-353.

*Lang, M., Mitkidis, P., Kundt, R., Nichols, A., Krajčíková, L., & Xygalatas, D. (2016). Music As a Sacred Cue? Effects of Religious Music on Moral Behavior. *Frontiers in psychology*, *7*, 814-814.

*Leach, M. M., Berman, M. E., & Eubanks, L. (2008). Religious activities, religious orientation, and aggressive behavior. *Journal for the Scientific Study of Religion*, 47(2), 311–319.

*Lehmann, C., Heller, M. Jewe, K., McDowell, M. (2017) [Religiousness and generosity]. Raw data.

*Li, K.-K., & Chow, W.-Y. (2015). Religiosity/spirituality and prosocial behaviors among Chinese Christian adolescents: The mediating role of values and gratitude. *Psychology of Religion and Spirituality*, 7(2), 150-161.

Li, Y., Shi, L., & Roth, D. H. (1994). The bias of the commonly-used estimate of variance in meta-analysis. *Communications in statistics-theory and methods*, 23(4), 1063-1085.

Light, R. J., & Pillemer, D. B. (1984). Quantitative procedures. In *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Lin, L. & Chu, H. Altmeta: Alternative meta-analysis methods. Retrieved from https://cran.r-project.org/web/packages/altmeta/index.html

*Lin, P. K., Tong, E. M., Lee, L. N., Low, A. H., & Gomes, D. (2016). The prosocial impact of God concept priming on God believers. *Psychology of Consciousness: Theory, Research, and Practice*, 3(1), 93.

*Liu, J., Gong, P., Li, H., & Zhou, X. (2017). A field study of the association between CD38 gene and altruistic behavior: Empathic response as a mediator. *Psychoneuroendocrinology*, 85, 165-171.

Lodi-Smith, J., & Roberts, B. W. (2007). Social investment and personality: A meta-analysis of the relationship of personality traits to investment in work, family, religion, and volunteerism. *Personality and Social Psychology Review*, 11(1), 68–86.

Maccoby, E. E., & Maccoby, N. (1954). The interview: A tool of social science. *In G. Lindzey (Ed.), Handbook of social psychology* (pp. 449–488). Cambridge, MA: Addison-Wesley.

Mahoney, A., Pargament, K. I., Tarakeshwar, N., & Swank, A. B. (2001). Religion in the Home in the 1980s and 1990s: A Meta-analytic Review and Conceptual Analysis of Links Between Religion, Marriage, and Parenting. *Journal of Family Psychology*, *15*(4), 559-596.

Malhotra, D. (2010). "Sunday Effect" on pro-social behavior. *Judgment and Decision Making*, 5(2), 138-143.

Malti, T., & Krettenauer, T. (2013). The relation of moral emotion attributions to prosocial and antisocial behavior: A meta-analysis. *Child development*, 84(2), 397-412.

*Martin, A. (2013). Does Religion Buffer Cheating? (Doctoral dissertation). Northern Illinois University, De Kalb, Illinois.

McCullough, M. E., Pargament, K. I., & Thoresen, C. E. (2001). *Forgiveness: Theory, research, and practice*. Guilford Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=bXZbpTaorg4C&oi=fnd&pg=PA1&dq=mccullough+pargament+thoresen&ots=m76GsDxqfW&sig=dSDSUn2OAEvz2qJRdc57UTYshFI

McCullough, M. E., & Willoughby, B. L. (2009). Religion, self-regulation, and self-control: Associations, explanations, and implications. *Psychological Bulletin*, 135(1), 69-93.

McKay, R., & Whitehouse, H. (2015). Religion and morality. *Psychological Bulletin*, 141(2), 447–473. https://doi.org/10.1037/a0038455

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(5), 730-749.

*Minton, E. A., Kahle, L. R., Jiuan, T. S., & Tambyah, S. K. (2016). Addressing Criticisms of Global Religion Research: A Consumption-Based Exploration of Status and Materialism, Sustainability, and Volunteering Behavior. *Journal for the Scientific Study of Religion*, *55*(2), 365-383.

*Miyatake, S., & Higuchi, M. (2017). Does religious priming increase the prosocial behaviour of a Japanese sample in an anonymous economic game?. *Asian Journal of Social Psychology*, 20(1), 54-59.

*Morgan J. (2017) [Morality Survey]. Raw data.

Murphy, K. R., & Lee, S. L. (1994). Personality variables related to integrity test scores: The role of conscientiousness. *Journal of Business and Psychology*, 8(4), 413–424.

*Musick, M. A., Wilson, J., & Bynum Jr, W. B. (2000). Race and formal volunteering: The differential effects of class and religion. *Social Forces*, *78*(4), 1539-1570.

*Nadal, A. R., Hardy, S. A., & Barry, C. M. (2016). Understanding the roles of religiosity and spirituality in emerging adults in the united states. *Psychology of Religion and Spirituality*, 10(1), 30-43.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3), 263–280.

*Nelson, M. F., James, M. S., Miles, A., Morrell, D. L., & Sledge, S. (2017). Academic Integrity of Millennials: The impact of religion and Spirituality. *Ethics & Behavior*, *27*(5), 385-400.

Norris, P., & Inglehart, R. (2011). *Sacred and Secular: Religion and Politics Worldwide* (Cambridge Studies in Social Theory, Religion and Politics). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511894862

Olthof, T. (2012). Anticipated feelings of guilt and shame as predictors of early adolescents' antisocial and prosocial interpersonal behaviour. *European Journal of Developmental Psychology*, 9(3), 371-388.

Oviedo, L. (2016). Religious attitudes and prosocial behavior: A systematic review of published research. *Religion, Brain & Behavior*, *6*(2), 169-184.

*Paciotti, B., Richerson, P., Baum, B., Lubell, M., Waring, T., McElreath, R., … Edsten, E. (2011). Are religious individuals more generous, trusting, and cooperative? An experimental test of the effect of religion on prosociality. In *The economics of*

religion: Anthropological approaches (pp. 267–305). Emerald Group Publishing Limited. Retrieved from http://www.emeraldinsight.com/doi/abs/10.1108/S0190-1281(2011)0000031014

Park, J. Z., & Smith, C. (2000). 'To Whom Much Has Been Given...': Religious Capital and Community Voluntarism Among Churchgoing Protestants. *Journal for the Scientific Study of Religion*, 39(3), 272-286.

Paul, G. (2005). Cross-national correlations of quantifiable societal health with popular religiosity and secularism in the prosperous democracies. *Journal of Religion and Society*, 7.

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*, 28(3), 450–461.

*Petrovic, K. (2017). [Volunteering behavior]. Raw data.

Pew Research Center. (2014). Worldwide, Many See Belief in God as Essential to Morality. Retrieved from http://www.pewglobal.org/2014/03/13/worldwide-many-see-belief-in-god-as-essential-to-morality/.

*Pichon, I., & Saroglou, V. (2009). Religion and helping: Impact of target thinking styles and just-world beliefs. *Archive for the Psychology of Religion*, *31*(2), 215-236.

*Piff (2016). [Religion and prosocial behavior]. Raw data.

*Poulin, M. (2006). [9/11 Longitudinal study]. Raw data.

Preston, J. L., & Ritter, R. S. (2013). Different effects of religion and God on prosociality with the ingroup and outgroup. *Personality and Social Psychology Bulletin*, 0146167213499937.

Prince, J.D.. (1904). The code of Hummurabi, *The Journal of Theology*, 8(3), 601-609.

Purzycki, B. G., Apicella, C., Atkinson, Q. D., Cohen, E., McNamara, R. A., Willard, A. K., ... & Henrich, J. (2016). Moralistic gods, supernatural punishment and the expansion of human sociality. *Nature*, 530(7590), 327-330.

Putnam, R. D. (2000). Bowling alone: America's declining social capital. In *Culture and politics* (pp. 223-234). Palgrave Macmillan, New York.

*Regnerus, M. D., & Smith, C. (1998). Who Gives to the Poor? The Influence of Religious Tradition and Political Location on the Personal Generosity of Americans Toward the Poor. *Journal for the Scientific Study of Religion*, *37*(3), 481-493.

Richard, F. D., Bond Jr, C. F., & Stokes-Zoota, J. J. (2003). One Hundred Years of Social Psychology Quantitatively Described. *Review of General Psychology*, *7*(4), 331-363.

Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology*, 58(1), 103–139.

Rokeach, M. (1970). Faith, hope and bigotry. *Psychology Today*, 3(11), 33–37.

Rosenthal, R. (1987). *Judgment studies: design, analysis, and meta-analysis*. New York: Cambridge University Press.

*Ruffle, B., & Richard, S. (2007). Does It Pay To Pray? Costly Ritual and Cooperation. *The BE Journal of Economic Analysis & Policy*, 7(1), 1-37.

Sage, L., Kavussanu, M., & Duda, J. (2006). Goal orientations and moral identity as predictors of prosocial and antisocial functioning in male association football players. *Journal of Sports Sciences*, 24(05), 455-466.

Sakaluk, J. Make it pretty: Forest and funnel plots for meta-analysis using ggplot2. [Blog post]. Retrieved from https://sakaluk.wordpress.com/2016/02/16/7-make-it-pretty-plots-for-meta-analysis/.

Saroglou, V. (2002). Religion and the five factors of personality: A meta-analytic review. *Personality and Individual Differences*, 32(1), 15–25.

Saroglou, V. (2006). Religion's role in prosocial behavior: Myth or reality? *Religion*, 31(2), 1–66.

*Saroglou, V., Pichon, I., Trompette, L., Verschueren, M., & Dernelle, R. (2005). Prosocial Behavior and Religion: New Evidence Based on Projective Measures and Peer Ratings. *Journal for the Scientific Study of Religion*, 44(3), 323–348.

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

Scanlan, J. (2002). *Dostoevsky the Thinker*. Ithaca, NY. Cornell University Press.

*Scott, M. (2017) [Religiosity and Prosocial Intention]. Unpublished data.

Sedikides, C., & Gebauer, J. E. (2010). Religiosity as self-enhancement: A meta-analysis of the relation between socially desirable responding and religiosity. *Personality and Social Psychology Review*, 14(1), 17–36.

*Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game. *Psychological science*, *18*(9), 803-809.

Shariff, A. F., & Norenzayan, A. (2011). Mean gods make good people: Different views of God predict cheating behavior. *The International Journal for the Psychology of Religion*, 21(2), 85–96.

Shariff, A. F., Piazza, J., & Kramer, S. R. (2014). Morality and the religious mind: why theists and nontheists differ. *Trends in Cognitive Sciences*, 18(9), 439–441. https://doi.org/10.1016/j.tics.2014.05.003

Shariff, A. F., & Rhemtulla, M. (2012). Divergent effects of beliefs in heaven and hell on national crime rates. *PLoS One*, 7(6), e39048.

Shariff, A. F., Willard, A. K., Andersen, T., & Norenzayan, A. (2015). Religious priming: A meta-analysis with a focus on prosociality. *Personality and Social Psychology Review*, *20*(1), 27-48.

*Shepherd, A. M., Schnitker, S. S., Leffel, G. M., Oakes Mueller, R. A., Curlin, F. A., Yoon, J. D., & Greenway, T. (2018). Developing the Good Physician: Spirituality affects the development of virtues and moral intuitions in medical students. *The Journal of Positive Psychology*, *13*(2), 143-154.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2), 534.

Slone, J. (2007). *Theological incorrectness: Why religious people believe what they shouldn't*. Oxford University Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=WcUv6o0oeSkC&oi=fnd&pg=PA2&dq=slone+theological+incorrectness&ots=iybN4yurPx&sig=-fY213C7-qbOKRXTCyPnB79ofms

Smith, G. A. (2017). *A Growing Share of Americans say it's not necessary to believe in God to be moral*. Retrieved from http://www.pewresearch.org/fact-tank/2017/10/16/a-growing-share-of-americans-say-its-not-necessary-to-believe-in-god-to-be-moral/.

*Smokowski, P. R., Guo, S., Cotter, K. L., Evans, C. B., & Rose, R. A. (2016). Multi-level risk factors and developmental assets associated with aggressive behavior in disadvantaged adolescents. *Aggressive behavior*, 42(3), 222-238.

*Stagnaro, M., Arechar, A., & Rand, D. G. (2018). Does belief in God cause prosociality? Manuscript in preparation.

*Stankov, L., & Lee, J. (2016). Nastiness, morality and religiosity in 33 nations. *Personality and Individual Differences*, 99, 56-66.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1), 60-78.

*Stavrova, O., & Siegers, P. (2014). Religious prosociality and morality across cultures: How social enforcement of religion shapes the effects of personal religiosity on prosocial and moral attitudes and behaviors. *Personality and Social Psychology Bulletin*, 40(3), 315-333.

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., ... & Tetzlaff, J. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal*, 343, d4002.

Sun, A. (2013). *Confucianism as a world religion: Contested histories and contemporary realities*. Princeton University Press.

*Thunström, L. (2018). *Thoughts and prayers—Do they crowd out donations?* Manuscript submitted for publication.

*Treboux, D. (2017). [Religiosity and prosocial behavior]. Unpublished analyses.

Trimble, D. E. (1997). The religious orientation scale: Review and meta-analysis of social desirability effects. *Educational and Psychological Measurement*, 57(6), 970–986.

van Assen, M. A., van Aert, R., & Wicherts, J. M. (2014). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, 20(3), 293.

Veenstra, R., Lindenberg, S., Oldehinkel, A. J., De Winter, A. F., Verhulst, F. C., & Ormel, J. (2008). Prosocial and antisocial behavior in preadolescence: Teachers' and parents' perceptions of the behavior of girls and boys. *International Journal of Behavioral Development*, 32(3), 243-251.

Vevea, J. L., & Woods, C. M. (2005). Publication Bias in Research Synthesis: Sensitivity Analysis Using A Priori Weight Functions. *Psychological Methods*, *10*(4), 428-443.

Viechtbauer, W. (2015). Metafor: Meta-analysis package for R version 1.9-2. Retrieved from https://cran.r-project.org/web/packages/metafor/index.html

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research synthesis methods*, 1(2), 112-125.

*Wagener, L. M., Furrow, J. L., King, P. E., Leffert, N., & Benson, P. (2003). Religious involvement and developmental resources in youth. *Review of Religious Research*, 44(3), 271–284.

*White, C. J. M. (2017). [Religioisity and Prosociality]. Analyses from unpublished data.

*White, C. J. M., Kelly, J.M., Shariff, A., & Norenzayan, A. (2018). Supernatural norm enforcement: Thinking about Karma and God reduces selfishness among believers. Submitted for review. https://psyarxiv.com/w4dfp

Witter, R. A., Stock, W. A., Okun, M. A., & Haring, M. J. (1985). Religion and subjective well-being in adulthood: A quantitative synthesis. *Review of Religious Research*, 332-342.

*Xygalatas, D. (2012). Effects of religious setting on cooperative behaviour. A case study from Mauritius, Religion, Brain and Behavior 3(2): 91-102

*Xygalatas, D., Mitkidis, P, Fischer, R., Reddish, P., Skewes, J., Geertz, A. W., Roepstorff, A. & Bulbulia, J. (2013). Extreme Rituals Promote Prosociality. Psychological Science 24(8): 1602–1605

*Xygalatas, D., Klocová, E. K., Cigán, J., Kundt, R., Maňo, P., Kotherová, S., … Kanovsky, M. (2016). Location, Location, Location: Effects of Cross-Religious Primes on Prosocial Behavior. *The International Journal for the Psychology of Religion*, 26(4), 304–319.

Xygalatas, D., Kotherová, S., Maňo, P., Kundt, R., Cigán, J., Klocová, E. K., & Lang, M. (2017). Big Gods in small places: the Random Allocation Game in Mauritius. *Religion, Brain & Behavior*, *8*(2), 243-261.

*Yeung, J. W. (2017). Religion, Volunteerism and Health: Are Religious People Really Doing Well by Doing Good?. *Social Indicators Research*, 1-20.

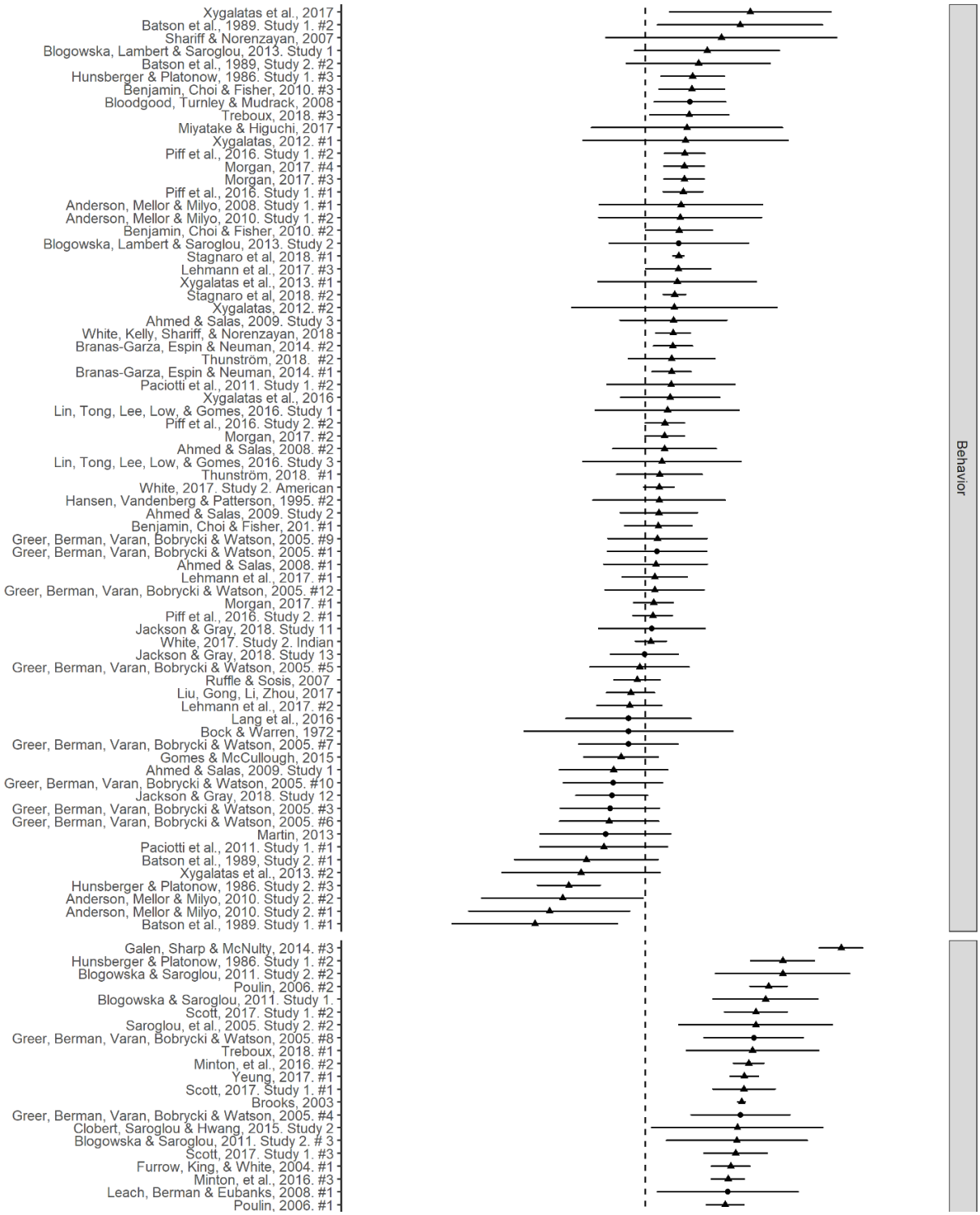
Ziegler, S., Koch, A., & Victor, N. (2001). Deficits and remedy of the standard random effects methods in meta-analysis. *Methods of Information in Medicine*, 40(02), 148-155.

Zuckerman, P. (2009). Atheism, Secularity, and Well-Being: How the Findings of Social Science Counter Negative Stereotypes and Assumptions. *Sociology Compass*, 3(6), 949–971.

Zuckerman, M., Silberman, J., & Hall, J. A. (2013). The relation between intelligence and religiosity: A meta-analysis and some proposed explanations. *Personality and Social Psychology Review*, 17(4), 325–354.

Appendix: Figures and Tables

Figure 1. Forest Plot of all effects



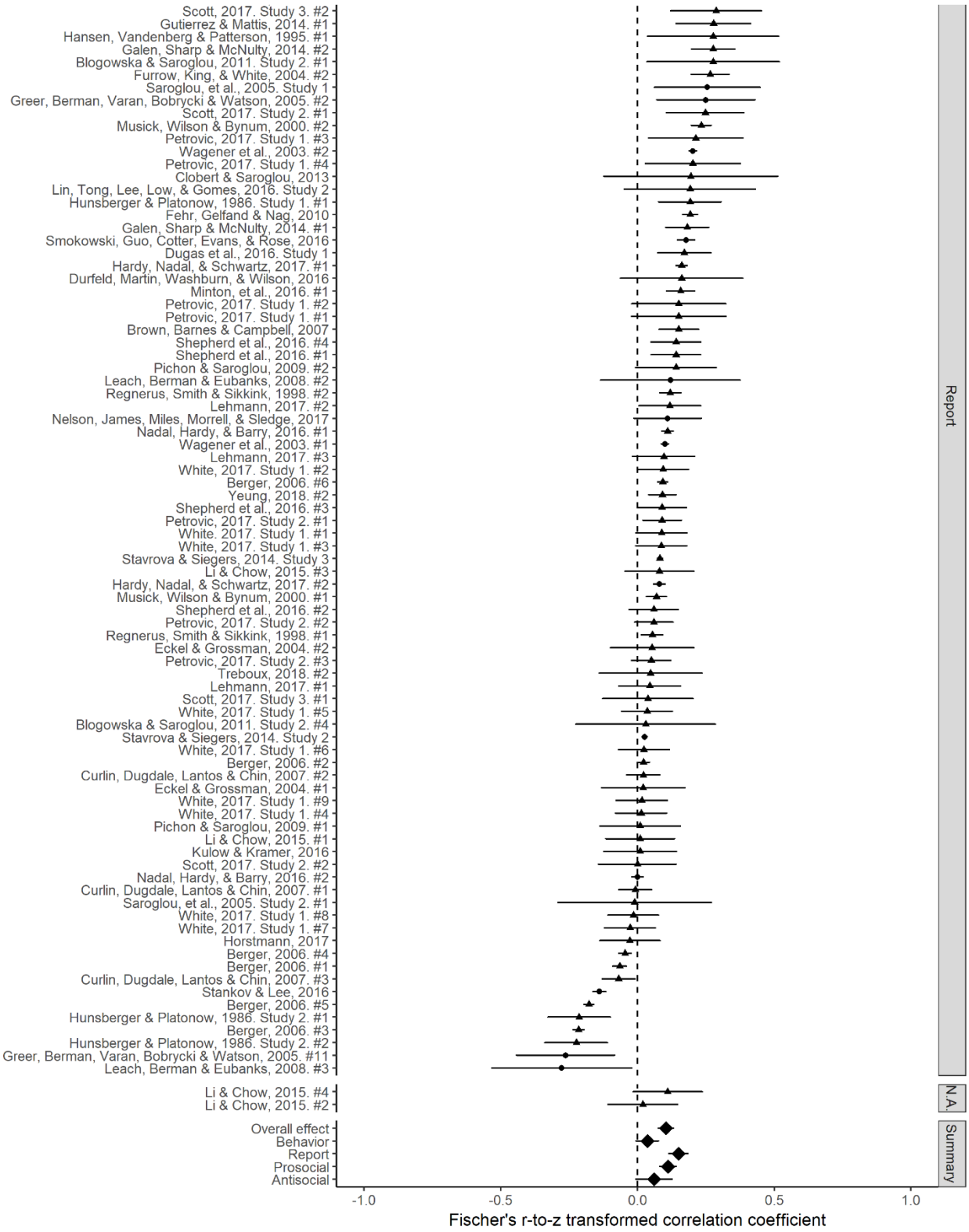
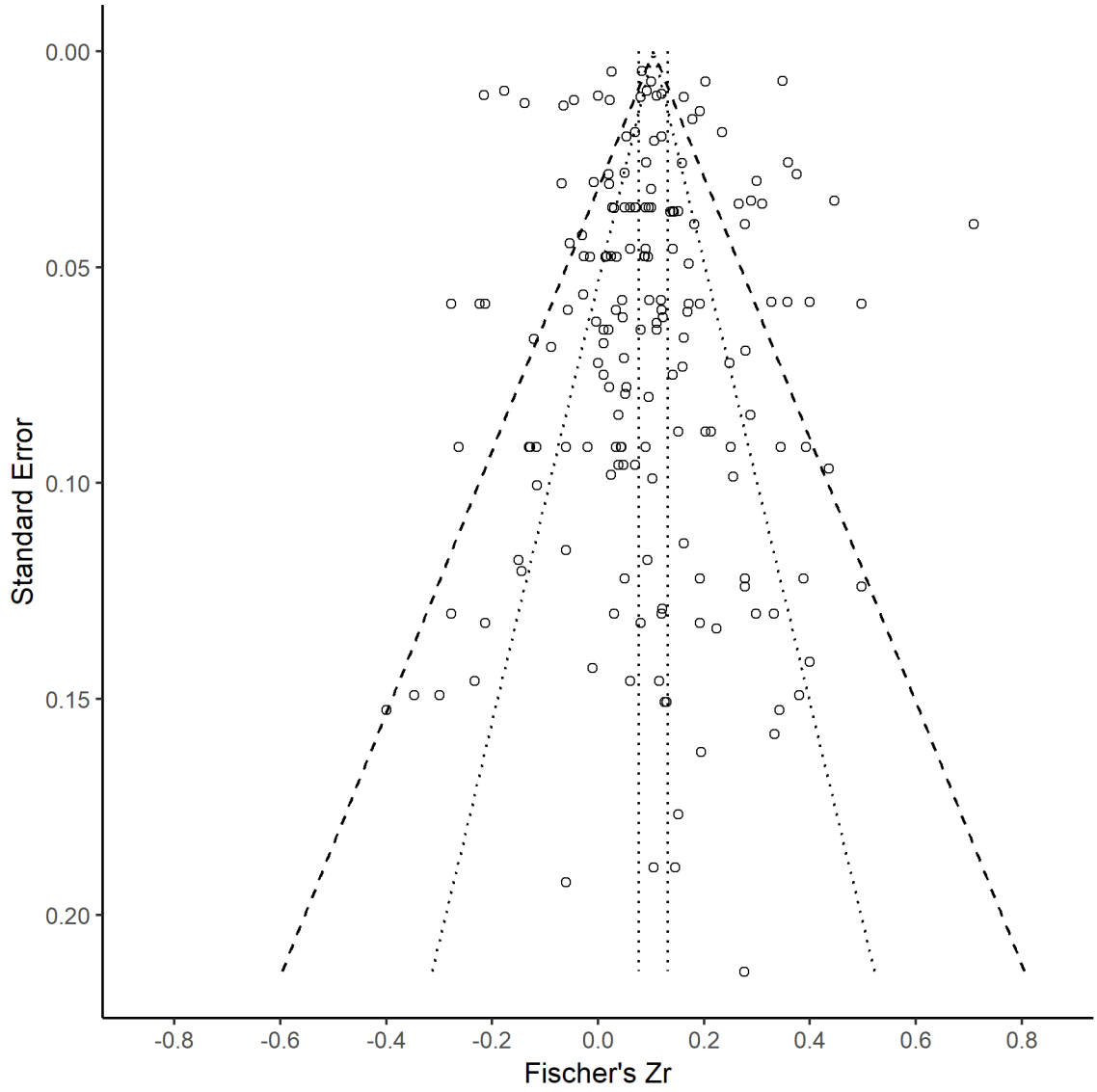
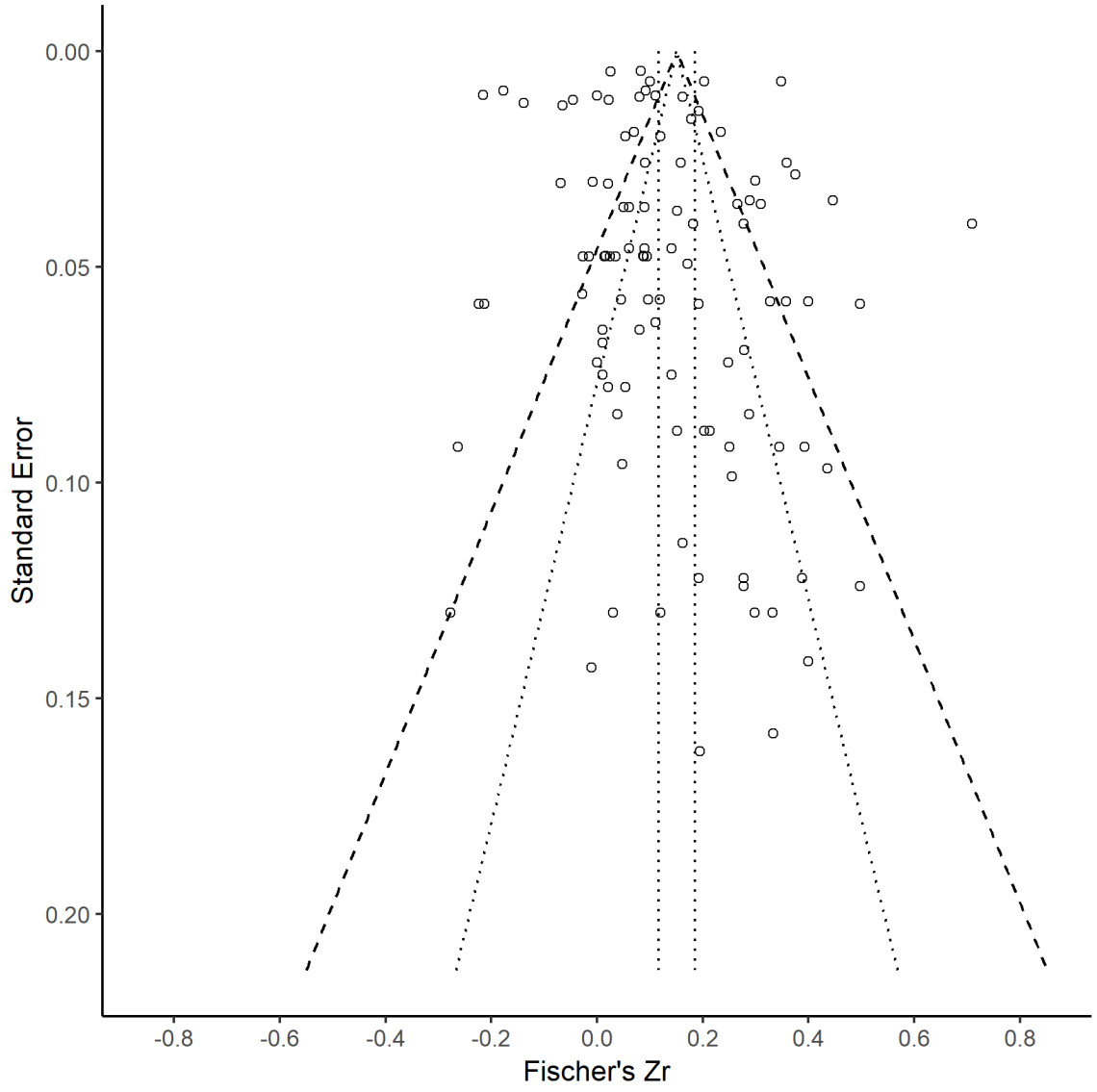

Scott, 2017. Study 3. #2
Gutierrez & Mattis, 2014. #1
Hansen, Vandenberg & Patterson, 1995. #1
Galen, Sharp & McNulty, 2014. #2
Blogowska & Saroglou, 2011. Study 2. #1
Furrow, King, & White, 2004. #2
Saroglou, et al., 2005. Study 1
Greer, Berman, Varan, Bobrycki & Watson, 2005. #2
Scott, 2017. Study 2. #1
Musick, Wilson & Bynum, 2000. #2
Petrovic, 2017. Study 1. #3
Wagener et al., 2003. #2
Petrovic, 2017. Study 1. #4
Clobert & Saroglou, 2013
Lin, Tong, Lee, Low, & Gomes, 2016. Study 2
Hunsberger & Platonow, 1986. Study 1. #1
Fehr, Gelfand & Nag, 2010
Galen, Sharp & McNulty, 2014. #1
Smokowski, Guo, Cotter, Evans, & Rose, 2016
Dugas et al., 2016. Study 1
Hardy, Nadal, & Schwartz, 2017. #1
Durfeld, Martin, Washburn, & Wilson, 2016
Minton, et al., 2016. #1
Petrovic, 2017. Study 1. #2
Petrovic, 2017. Study 1. #1
Brown, Barnes & Campbell, 2007
Shepherd et al., 2016. #4
Shepherd et al., 2016. #1
Pichon & Saroglou, 2009. #2
Leach, Berman & Eubanks, 2008. #2
Regnerus, Smith & Sikkink, 1998. #2
Lehmann, 2017. #2
Nelson, James, Miles, Morrell, & Sledge, 2017
Nadal, Hardy, & Barry, 2016. #1
Wagener et al., 2003. #1
Lehmann, 2017. #3
White, 2017. Study 1. #2
Berger, 2006. #6
Yeung, 2018. #2
Shepherd et al., 2016. #3
Petrovic, 2017. Study 2. #1
White. 2017. Study 1. #1
White, 2017. Study 1. #3
Stavrova & Siegers, 2014. Study 3
Li & Chow, 2015. #3
Hardy, Nadal, & Schwartz, 2017. #2
Musick, Wilson & Bynum, 2000. #1
Shepherd et al., 2016. #2
Petrovic, 2017. Study 2. #2
Regnerus, Smith & Sikkink, 1998. #1
Eckel & Grossman, 2004. #2
Petrovic, 2017. Study 2. #3
Treboux, 2018. #2
Lehmann, 2017. #1
Scott, 2017. Study 3. #1
White, 2017. Study 1. #5
Blogowska & Saroglou, 2011. Study 2. #4
Stavrova & Siegers, 2014. Study 2
White, 2017. Study 1. #6
Berger, 2006. #2
Curlin, Dugdale, Lantos & Chin, 2007. #2
Eckel & Grossman, 2004. #1
White, 2017. Study 1. #9
White, 2017. Study 1. #4
Pichon & Saroglou, 2009. #1
Li & Chow, 2015. #1
Kulow & Kramer, 2016
Scott, 2017. Study 2. #2
Nadal, Hardy, & Barry, 2016. #2
Curlin, Dugdale, Lantos & Chin, 2007. #1
Saroglou, et al., 2005. Study 2. #1
White, 2017. Study 1. #8
White, 2017. Study 1. #7
Horstmann, 2017
Berger, 2006. #4
Berger, 2006. #1
Curlin, Dugdale, Lantos & Chin, 2007. #3
Stankov & Lee, 2016
Berger, 2006. #5
Hunsberger & Platonow, 1986. Study 2. #1
Berger, 2006. #3
Hunsberger & Platonow, 1986. Study 2. #2
Greer, Berman, Varan, Bobrycki & Watson, 2005. #11
Leach, Berman & Eubanks, 2008. #3
Li & Chow, 2015. #4
Li & Chow, 2015. #2
Overall effect
Behavior
Report
Prosocial
Antisocial
Report
N.A.
Summary
-1.0
-0.5
0.0
0.5
1.0
Fischer's r-to-z transformed correlation coefficient

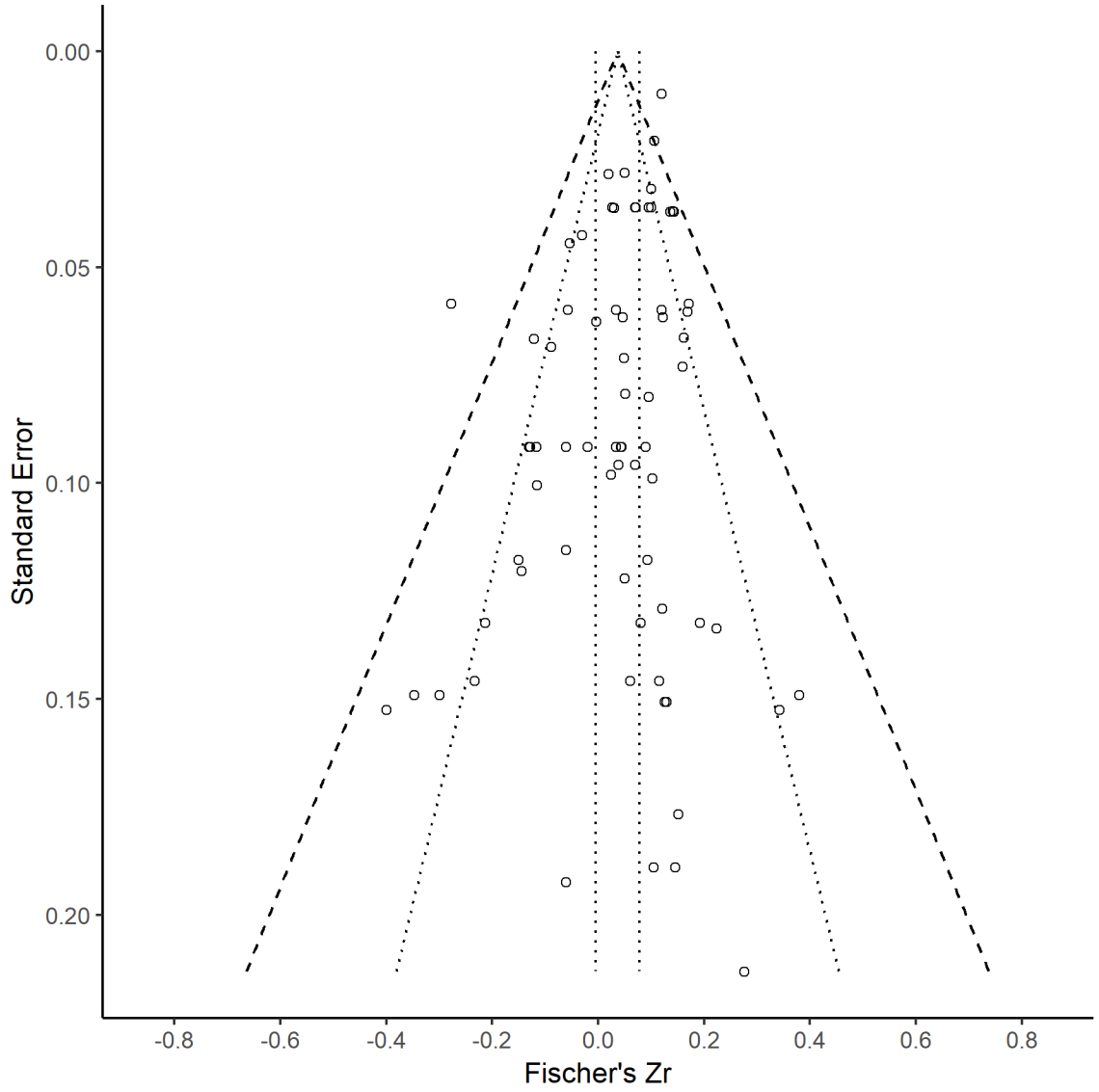
Figure 2.
a. Overall funnel plot



b. Report measure funnel plot



c. Behavior measure funnel plot



Note: The outer diagonal lines represent the 99% confidence interval region, while the inner diagonal lines represent the 95% confidence interval region. Dotted vertical lines represent the 95% confidence interval of the meta-analytic estimate. Code for plots adapted from Sakaluk (2016).

Table 1. Studies included in the meta-analysis

Reference	Study #	r	Sample size	DV type	Relig. Type	Pro/ antisocial	WEIRD status	College Student	Age	% Female
Ahmed & Salas, 2008		0.04	112	behav	aff	pro	non	student	19	29
		0.07	112	behav	aff	pro	non	student	19	29
Ahmed & Salas, 2009	Study 1	-0.11	102	behav	aff	pro	non	student	22	54
	Study 2	0.05	201	behav	aff	pro	non	student	20	64
	Study 3	0.10	105	behav	aff	pro	WEIRD	student	23	48
Anderson, Mellor & Milyo, 2010	Study 1	0.12	47	behav	aff	pro	WEIRD	student	NA	NA
		0.13	47	behav	att	pro	WEIRD	student	NA	NA
	Study 2	-0.33	48	behav	aff	pro	WEIRD	student	NA	NA
		-0.29	48	behav	att	pro	WEIRD	student	NA	NA
Batson et al., 1989	Study 1	-0.38	46	behav	scale	pro	WEIRD	student	NA	46
		0.33	46	behav	scale	pro	WEIRD	student	NA	46
	Study 2	-0.21	60	behav	scale	pro	WEIRD	student	NA	100
		-0.21	60	behav	scale	pro	WEIRD	student	NA	100
Benjamin, Choi & Fisher, 2010		0.05	266	behav	att	pro	WEIRD	student	NA	52
		0.12	266	behav	aff	pro	WEIRD	student	NA	52
		0.16	277	behav	aff	pro	WEIRD	student	NA	52
Berger, 2006	Donations	-0.07	6269	report	att	pro	WEIRD	non	NA	56
		0.02	7867	report	aff	pro	WEIRD	non	NA	56
		-0.21	9613	report	att	pro	WEIRD	non	NA	56
	Volunteer	-0.05	7837	report	scale	pro	WEIRD	non	NA	56
		-0.18	11981	report	scale	pro	WEIRD	non	NA	56
		0.09	12001	report	aff	pro	WEIRD	non	NA	56
	Study 1	0.41	110	report	scale	pro	WEIRD	student	20.0	63

Blogowska & Saroglou, 2011	Study 2	0.32	62	report	scale	pro	WEIRD	student	21.4	63
	Study 2	0.27	68	report	scale	pro	WEIRD	student	21.4	63
	Study 2	0.46	68	report	scale	pro	WEIRD	student	21.4	63
	Study 2	0.03	62	report	scale	pro	WEIRD	student	21.4	63
Blogowska, Lambert & Saroglou, 2013	Study 1	0.22	59	behav	scale	pro	WEIRD	student	19.5	88
	Study 2	0.12	63	behav	scale	anti	WEIRD	student	19.9	88
Bloodgood, Turnley & Mudrack, 2008		0.16	230	behav	att	anti	WEIRD	student	NA	37
Bock & Warren, 1972		-0.06	30	behav	scale	anti	WEIRD	student	NA	NA
Brañas-Garza, Espín & Neuman, 2014		0.09	766	behav	aff	pro	WEIRD	non	37.7	54
		0.10	766	behav	att	pro	WEIRD	non	37.7	54
Brooks, 2003		0.34	20734	report	att	pro	WEIRD	non	NA	59
Brown, Barnes & Campbell, 2007	Study 2	0.15	734	report	scale	pro	WEIRD	student	NA	63
Clobert & Saroglou, 2013		0.19	41	report	aff	pro	non	student	19.6	90
Clobert, Saroglou & Hwang, 2015	Study 2	0.32	43	report	aff	pro	WEIRD	student	20	88
Curlin, Dugdale, Lantos & Chin, 2007		-0.00	1091	report	att	pro	WEIRD	non	49	26
		0.02	1063	report	scale	pro	WEIRD	non	49	26
		-0.07	1072	report	aff	pro	WEIRD	non	49	26
Dugas et al., 2016	Study 1	0.17	415	report	aff	pro	non	non	39.3	51
Durfeld, Martin, Washburn,		0.16	80	report	att	pro	WEIRD	student	22	89

& Wilson, 2016									
Eckel & Grossman, 2004	0.02	168	report	scale	pro	WEIRD	student	23	44
	0.05	168	report	scale	pro	WEIRD	student	23	44
Fehr, Gelfand & Nag, 2010	0.19	5224	report	NA	pro	NA	non	NA	NA
Furrow, King, & White, 2004	0.30	801	report	att	pro	WEIRD	non	15.9	53
	0.26	801	report	scale	pro	WEIRD	non	15.9	53
Galen, Sharp & McNulty, 2014	0.18	628	report	scale	pro	WEIRD	non	48.4	24
	0.27	628	report	scale	pro	WEIRD	non	48.4	24
	0.61	628	report	scale	pro	WEIRD	non	48.4	24
Gomes & McCullough, 2015	-0.09	216	behav	NA	pro	WEIRD	non	NA	46
Greer, et al., 2005	0.04	122	behav	att	anti	WEIRD	student	20.4	71
	0.25	122	report	att	anti	WEIRD	student	20.4	71
	-0.13	122	behav	att	anti	WEIRD	student	20.4	71
	0.33	122	report	att	anti	WEIRD	student	20.4	71
	-0.02	122	behav	att	pro	WEIRD	student	20.4	71
	-0.13	122	behav	att	pro	WEIRD	student	20.4	71
	-.06	122	behav	att	anti	WEIRD	student	20.4	71
	.37	122	report	att	anti	WEIRD	student	20.4	71
	.04	122	behav	att	pro	WEIRD	student	20.4	71
	-.12	122	behav	att	anti	WEIRD	student	20.4	71
	-.26	122	report	att	anti	WEIRD	student	20.4	71
	.03	122	behav	att	pro	WEIRD	student	20.4	71

Gutierrez & Mattis, 2014		0.27	211	report	aff	pro	WEIRD	non	32.4	100
Hansen, Vandenberg & Patterson, 1995		0.27	70	report	scale	pro	WEIRD	student	25.4	54
		0.05	70	behav	scale	pro	WEIRD	student	25.4	54
Hardy, Nadal, & Schwartz, 2017		0.16	8987	report	scale	pro	WEIRD	student	19.8	73
		0.08	8987	report	scale	anti	WEIRD	student	19.8	73
Horstmann, 2017		-0.03	318	report	aff	pro	WEIRD	non	NA	NA
Hunsburger & Platonow, 1986	Study 2	0.19	295	report	scale	pro	WEIRD	student	NA	NA
		0.46	295	report	scale	pro	WEIRD	student	NA	NA
		0.17	295	behav	scale	pro	WEIRD	student	NA	NA
		-0.21	295	report	scale	pro	WEIRD	student	NA	NA
		-0.22	295	report	scale	pro	WEIRD	student	NA	NA
		0.27	295	behav	scale	pro	WEIRD	student	NA	NA
Jackson & Gray, 2018	Study 11	0.02	107	behav	aff	anti	WEIRD	non	37.6	58
	Study 12	-0.12	228	behav	aff	anti	WEIRD	non	35.6	50
	Study 13	-0.00	258	behav	aff	anti	WEIRD	non	36.3	56
Kulow & Kramer, 2016		0.01	222	report	scale	pro	WEIRD	student	21	61
Lang et al., 2016		-.06	78	Behave	Scale	anti	non	non	23.3	42
Leach, Berman & Eubanks, 2008		0.29	62	report	scale	anti	WEIRD	student	20.4	73
		0.12	62	report	scale	anti	WEIRD	student	20.4	73
		-0.27	62	report	scale	anti	WEIRD	student	20.4	73
Lehmann, 2017		0.05	304	report	scale	pro	WEIRD	non	19.1	81
		0.12	304	report	scale	pro	WEIRD	non	19.1	81

		0.10	304	report	scale	pro	WEIRD	non	19.1	81
Lehmann et al., 2017		0.03	281	behav	scale	pro	WEIRD	non	37.4	47
		-0.06	281	behav	scale	pro	WEIRD	non	37.4	47
		0.12	281	behav	scale	pro	WEIRD	non	37.4	47
Li & Chow, 2015		0.01	243	report	scale	pro	non	non	18.1	52
		0.02	243	report	scale	pro	non	non	18.1	52
		0.08	243	report	scale	pro	non	non	18.1	52
		0.11	243	report	scale	pro	non	non	18.1	52
Lin et al., 2016	Study 1	0.08	60	behav	aff	pro	WEIRD	student	21.6	71
	Study 2	0.19	70	report	aff	pro	WEIRD	student	20.9	64
	Study 3	0.06	50	behav	aff	pro	WEIRD	student	22.6	67
Liu, et al., 2017		-0.05	509	behav	aff	pro	non	student	24.4	74
Martin, 2013		-0.14	72	behav	scale	anti	WEIRD	student	19.2	54
Minton et al., 2016		0.16	1499	report	scale	pro	non	non		51
		0.36	1111	report	scale	pro	non	non		51
		0.29	1232	report	scale	pro	non	non		51
Miyatake & Higuchi, 2017		0.15	35	behav	aff	pro	non	student	19.8	71
Morgan, 2017		0.03	763	behav	aff	pro	WEIRD	non	NA	NA
		0.07	764	behav	scale	pro	WEIRD	non	NA	NA
		0.14	729	behav	aff	pro	WEIRD	non	NA	NA
		0.14	731	behav	scale	pro	WEIRD	non	NA	NA

Musick, Wilson & Bynum, 2000		0.07	2867	report	scale	pro	WEIRD	non	46.6	56
		0.23	2867	report	scale	pro	WEIRD	non	46.6	56
Nadal, Hardy, & Barry, 2016		0.11	9495	report	scale	pro	WEIRD	student	19.8	73
		0.00	9495	report	scale	anti	WEIRD	student	19.8	73
Nelson et al., 2017		0.11	256	report	scale	anti	WEIRD	student	NA	NA
Paciotti et al., 2011	Study 1	-0.15	75	behav	NA	pro	NA	student	20.9	56
		0.09	75	behav	NA	pro	NA	student	20.9	56
Petrovic, 2017	Study 1	0.15	132	report	att	pro	WEIRD	non	30.1	65
		0.15	132	report	att	pro	WEIRD	non	30.1	65
		0.21	132	report	scale	pro	WEIRD	non	30.1	65
		0.20	132	report	scale	pro	WEIRD	non	30.1	65
	Study 2	0.09	769	report	aff	pro	WEIRD	non	42.2	68
		0.06	769	report	scale	pro	WEIRD	non	42.2	68
		0.05	768	report	att	pro	WEIRD	non	42.2	68
Pichon & Saroglou, 2009		0.01	181	report	scale	pro	WEIRD	non	29.9	48
		0.14	181	report	scale	pro	WEIRD	non	29.9	48
Piff et al., 2016	Study 1	0.14	727	behav	scale	pro	WEIRD	non	50.2	51
	Study 1	0.14	727	behav	scale	pro	WEIRD	non	50.2	51
	Study 2	0.03	765	behav	scale	pro	WEIRD	non	50.2	51
	Study 2	0.07	765	behav	scale	pro	WEIRD	non	50.2	51
Poulin, 2006		0.28	844	report	att	pro	WEIRD	non	46.5	NA
		0.42	844	report	att	pro	WEIRD	non	46.5	NA
Regnerus, Smith &		0.05	2573	report	aff	pro	NA	non	NA	59

Sikkink, 1998		0.01	2591	report	aff	pro	NA	non	NA	59
Ruffle & Sosis, 2007		-0.03	552	behav	aff	pro	WEIRD	non	48.4	51
Saroglou et al., 2005	Study 1	0.25	106	report	scale	anti	WEIRD	student	19.8	69
	Study 2	-0.01	52	report	scale	pro	WEIRD	student	19.8	100
		0.38	53	report	scale	pro	WEIRD	student	19.8	100
Scott, 2017	Study 1	0.34	300	report	scale	pro	WEIRD	non	34.7	47
		0.38	300	report	scale	pro	WEIRD	non	34.7	47
		0.32	300	report	scale	pro	WEIRD	non	34.7	47
	Study 2	0.24	195	report	aff	pro	WEIRD	student	19.3	57
		0.00	195	report	aff	pro	WEIRD	student	19.3	57
	Study 3	0.04	144	report	aff	pro	WEIRD	student	18.7	65
		0.28	144	report	aff	pro	WEIRD	student	18.7	65
Shariff & Norenzayan, 2007	Study 1	0.27	25	behav	aff	pro	WEIRD	student	21	68
Shepherd et al., 2018		0.14	480	report	scale	pro	WEIRD	non	NA	45
		0.06	480	report	scale	pro	WEIRD	non	NA	45
		0.09	480	report	scale	pro	WEIRD	non	NA	45
		0.14	480	report	scale	pro	WEIRD	non	NA	45
Smokowski, Guo, Cotter, Evans, & Rose, 2016		0.18	4056	report	scale	anti	WEIRD	non	NA	52
Stagnaro et al, 2018		0.12	10184	behav	scale	pro	WEIRD	non	NA	51
		0.11	2334	behav	scale	pro	WEIRD	non	NA	50

Stankov & Lee, 2016		-0.14	6938	report	scale	anti	non	non	22.3	57
Stavrova & Siegers, 2014	Study 2	0.03	43965	report	scale	anti	WEIRD	non	NA	NA
	Study 3	0.08	46859	report	scale	pro	non	non	NA	NA
Thunström, 2018		0.05	162	behav	aff	pro	WEIRD	non	47	52
		0.10	159	behav	aff	pro	WEIRD	non	47	52
Treboux, 2017		0.37	70	report	aff	pro	WEIRD	student	20.8	78
		0.05	112	report	aff	pro	WEIRD	non	35.7	60
		0.16	190	behav	aff	pro	WEIRD	non	36.1	51
Wagener et al, 2003		0.10	20020	report	scale	anti	WEIRD	non	NA	51
		0.20	20020	report	att	anti	WEIRD	non	NA	51
White, 2017	Study 1	0.09	446	report	scale	pro	WEIRD	student	20.1	74
		0.09	445	report	scale	pro	WEIRD	student	20.1	74
		0.09	447	report	aff	pro	WEIRD	student	20.1	74
		0.01	445	report	scale	pro	WEIRD	student	20.1	74
		0.04	444	report	scale	pro	WEIRD	student	20.1	74
		0.02	446	report	aff	pro	WEIRD	student	20.1	74
		-0.03	446	report	scale	pro	WEIRD	student	20.1	74
		-0.02	445	report	scale	pro	WEIRD	student	20.1	74
		0.02	447	report	aff	pro	WEIRD	student	20.1	74
	Study 2, American	0.05	1264	behav	scale	pro	WEIRD	non	37.5	63
	Study 2, Indian	0.02	1233	behav	scale	pro	non	non	31.2	28
White et al., 2018	Study 3	0.10	986	behav	scale	pro	WEIRD	non	35.5	43
Xygalatas, 2012		.15	31	behav	scale	pro	non	non	34.5	47

	.11	31	behav	att	pro	non	non	34.5	47
Xygalatas et al., 2013	.12	50	behav	scale	pro	non	non		
	-.23	50	behav	att	pro	non	non		
Xygalatas et al., 2016	0.09	122	behav	scale	pro	non	non	31.6	44
Xygalatas et al., 2017	.36	48	behave	scale	pro	non	non	36.8	29
Yeung, 2017	0.34	1504	report	scale	pro	WEIRD	non	NA	NA
	0.09	1504	report	scale	pro	WEIRD	non	NA	NA

Table 2. Estimated effect by religiosity measure type

	k	*r*	*t*	*p*	95% Confidence Lower Bound	95% Confidence Upper Bound
Scale	105	.11	6.24	<.001	.08	.15
Attendance	26	.12	3.60	<.001	.05	.18
Affiliation	44	.09	3.63	<.001	.04	.14

Table 3. Meta-regression.

	b	*p*	95% Confidence Lower Bound	95% Confidence Upper Bound
Intercept	-.07	.22	- .19	. 04
Self-Report	.11	.001	.05	.18
Prosocial	.08	.04	.01	.16
WEIRD sample	.05	.36	-.05	.14
Attendance	.02	.73	- .08	.11
Scale	.00	.89	-.06	.07
Student	.01	.84	- .09	.11
Age (mean centered)	.00	.83	- .00	.01
% Female (mean centered)	-.03	.84	- .27	.22

Table 4. Meta-analytical estimates of the relationship between religiosity and selected outcomes.

Citation	Outcome	Estimated effect (*r*)
Present research	Prosociality (overall)	.10
	Prosociality (report)	.15
	Prosociality (behavior)	.04
Baier & Wright (2001)	Crime	- 0.12
Bergin (1983)	Mental health	0.09
Hackney & Sanders (2003)		0.10
Hall, Matz, & Wood (2010)	Racism (*correlated with religious identification*)	0.12
Mahoney et al. (2001)	Divorce (*correlated with religious attendance*)	- 0.13
	Marital satisfaction (*correlated with religious attendance*)	0.07
	Marital commitment	0.19
Saroglou (2002)	Extraversion	0.07
	Agreeableness	0.21
	Conscientiousness	0.18
	Neuroticism	0.02
	Openness	0.04
Sedikides & Gebauer (2010)	Socially desirable responding	0.10
Witter et al. (1985)	Subjective Well-being	0.16
Zuckerman, Silberman, & Hall (2013)	Intelligence	-0.24

Note: Correlations are with overall religiosity, unless otherwise noted.