Article

A Meta-Analytic Review of the Validity of the Tangram Help/Hurt Task

Personality and Social Psychology Bulletin
2024, Vol. 50(3) 436–449

DOI: 10.1177/01461672221127759
journals.sagepub.com/home/pspb
SAGE

Sophie L. Kjærvik[1], Muniba Saleem[2], Gertrudes Velasquez[3], Craig A. Anderson[4], and Brad J. Bushman[1]

Abstract

The Tangram Help/Hurt Task (THHT) allows participants to help another participant win a prize (by assigning them easy tangrams), to hurt another participant by preventing them from winning the prize (by assigning them difficult tangrams), or to do neither (by assigning them medium tangrams) in offline or online studies. Consistent with calls for continued evidence supporting psychological measurement, we conducted a meta-analytic review of the THHT that included 52 independent studies involving 11,060 participants. THHT scores were associated with helping and hurting outcomes in theoretically predicted ways. Results showed that THHT scores were not only associated with short-term (experimental manipulations, state measures) and long-term (trait measures) helping and hurting outcomes, but also with helping and harming intentions. We discuss the strengths and limitations of the THHT relative to other laboratory measures of prosocial behavior and aggression, discuss unanswered questions about the task, and offer suggestions for the best use of the task.

Keywords

Tangram Help/Hurt Task, THHT, tangram, aggression, helping

Received February 24, 2022; revision accepted August 29, 2022

> "If you haven't measured something, you really don't know very much about it."
>
> —Karl Pearson

> "Science is measurement. If I cannot make measurements, I cannot study a problem scientifically."
>
> —William H. George

The first step toward understanding a theoretical concept is to measure it in a valid and reliable manner. Creating and validating independent and dependent variables are central to scientific advancement.[1] Within the social/behavioral sciences, the creation/validation process often is quite complex and involves translating conceptual (theoretical) stimulus and response variables into appropriate empirical realizations (Denson & Anderson, press; Prot & Anderson, 2013). Debates about the validity of particular measures and manipulations are common in psychology and other social/behavioral sciences and often are useful in improving the sciences for several reasons, including improved theorizing, improved measurement, and understanding failures to replicate. In sum, there is a continuing need to examine evidence pertaining to the rigor, strengths, and limitations of measures.

In keeping with this focus on the benefits of measurement validity debates, the present research provides meta-analytic evidence for a laboratory measure assessing two very important theoretical concepts—helping (prosocial behavior) and hurting (aggression)—called the Tangram Help/Hurt Task (THHT; Saleem, Anderson, et al., 2015; Saleem, Prot, et al., 2015; Saleem et al., 2017). In doing so, we discuss the strengths and limitations of the THHT as reflected in existing studies that have used the task, consider questions about the task that remain unknown and would be useful to examine in future research, and provide suggestions and recommendations for scholars interested in using THHT to assess helpful and hurtful behaviors in their own work. Before discussing these details, it is important to consider: (a) definitions of aggressive and prosocial behaviors and (b) existing laboratory paradigms used to assess these behaviors.

What are Aggressive and Prosocial Behaviors?

Hurting and helping are descriptive labels of the social psychological concepts of aggressive and prosocial behaviors. In this research, as in much of existing literature, we consider

[1]The Ohio State University, Columbus, USA
[2]University of California, Santa Barbara, USA
[3]Florida State University, Tallahassee, USA
[4]Iowa State University, Ames, USA

Corresponding Author:
Brad J. Bushman, School of Communication, The Ohio State University, 3016 Derby Hall, 154 North Oval Mall, Columbus, OH, USA.
Email: bushman.20@osu.edu

hurting and helping to be isomorphic with aggressive and prosocial behaviors, respectively. Indeed, scholars in both of these literatures describe specific examples of hurting (e.g., hurting someone's feelings) and helping (e.g., helping those less fortunate) to illustrate aggressive and prosocial behaviors (Baron & Richardson, 1994; Schroeder & Graziano, 2015).

Although there are slight variations in the definition, most scholars agree that aggression is "any form of behavior directed toward the goal of harming or injuring another living being who is motivated to avoid such treatment." (Baron & Richardson, 1994, p. 7). Further distinctions can be made between different forms of aggression (e.g., verbal vs. physical, direct vs. indirect vs. displaced, active vs. passive, Bushman & Huesmann, 2010; Parrott & Giancola, 2007). To our knowledge, hurting and aggressive behaviors are not distinguished in the aggressive behavior literature, although the latter is more common than the former. Overall, there is considerable agreement among aggression scholars on the conceptual definition of aggression, but debates about the validity of the various empirical realizations (aka, operational definitions) frequently arise. What existing measures have established construct, internal, and external validity (Cook & Campbell, 1979; discussed below)? What exactly do they measure? Of course, ethical considerations preventing participants from physically and seriously harming each other, necessarily constrain laboratory measures of aggressive behavior, which ultimately influence concerns relating to construct and external validity. However, these concerns are not specific to measures of aggression (e.g., Bader et al., 2021); they apply to measurement in all sciences.

Scholars interested in studying prosocial behavior have grappled with similar concerns (e.g., Galizzi & Navarro-Martinez, 2019; Wispe, 1972). Overall, there is agreement that prosocial behavior reflects acts of positive social behavior intended to promote the welfare of others (Pfattheicher et al., 2022). Inconsistencies in definitions exist with regard to: (a) who the receiver of the welfare is (individual, group, society), (b) what is meant by welfare (e.g., psychological wellbeing, money), and (c) the role of benefactor's intentions. Further distinctions can be made between prosocial, helping, and altruistic behaviors depending on: (a) consequences of the act from the recipient's perspective, (b) locus of reinforcement (i.e., internal vs. external) for the benefactor, (c) intent of the benefactor, and (d) motivation underlying helping (Bar-Tal & Raviv, 1982). Although there is no unanimous consensus on these distinctions, helping and altruistic behaviors are considered to be subcategories of prosocial behavior, with the latter (but not necessarily the former) involving a specific motivation to act in a prosocial way with an "ultimate goal of increasing another's welfare" (Batson, 2011, p. 20). For simplicity's sake, these labels can be thought of evolving on a continuum of broad to specific forms (i.e., prosocial, helping, altruism) of positive behaviors that promote or are aimed at promoting the welfare of others (Penner et al., 2005). Based on existing evidence, helping as measured within the THHT appears to coincide with the definition of prosocial and helping behaviors but not necessarily altruistic behaviors.

Laboratory Measures of Aggressive and Prosocial Behaviors

Laboratory assessments of aggressive and prosocial behaviors in psychological sciences have a long history in which these constructs have been measured in diverse and creative ways (e.g., Galizzi & Navarro-Martinez, 2019; McCarthy & Elson, 2018; Wispe, 1972). Although many of these measures have been widely used and have provided valuable empirical data (e.g., Burnett Heyes, 2020; Motsenok et al., 2022; Thielmann et al., 2020; Warburton & Bushman, 2019), they are not void of criticisms. Consistent with concerns about experimental research and laboratory paradigms in general (e.g., Lin et al., 2021; Mitchell, 2012), laboratory assessments of aggression and prosocial behavior are frequently criticized based on construct validity (i.e., does it measure what it is supposed to measure in the laboratory context?) and external validity (i.e., does it generalize to other people, places, and times?).

It is important to note that despite their limitations pertaining to external validity, laboratory measures afford a number of highly important advantages such as (a) the ability to draw cause-and-effect conclusions, based on experimental manipulation of relevant independent variables; (b) the examination of the behavior of interest within a controlled and safe environment; (c) the opportunity to systematically and precisely test theories; and (d) the ability to measure actual behaviors rather than self-report endorsements of behaviors. The latter is especially important when the behaviors of interest are likely to be influenced by concerns related to social desirability and societal norms as is true in the case of helping and hurting others.

Overall, criticisms of laboratory measures of aggression fall within the dimensions of validity, methodological, and analytical concerns (but see responses to criticisms from Anderson & Bushman, 1997; Giancola & Chermack, 1998; Hyatt et al., 2019; Warburton & Bushman, 2019). Specifically, there are concerns that existing measures of aggression exclusively assess the least severe forms of behaviors on the aggression spectrum, thereby reducing (a) the generalizability of findings to real-world aggression and violence and (b) an understanding of the multidimensional aspect of aggression. Methodologically, the lack of nonaggressive options in tasks, failure to assess choice intention/motivation, cover stories introducing competitive or prosocial motives, demand characteristics and social desirability concerns, distance between aggressor and victim, and variations in the administration of the procedures have been questioned (Elson et al., 2014; Ritter & Eslea, 2005; Tedeschi & Quigley, 1996). Analytically, variations in the use and calculation of responses derived from these measures can yield inconsistent findings including

the size and direction of effects (Elson et al., 2014; c.f., Hyatt et al., 2019). Accordingly, recommendations have been made to standardize both the procedure and the analysis of responses from these tasks (Elson et al., 2014).

In the realm of prosocial behavior, laboratory paradigms, relative to scales and observational reports, are limited. Unlike laboratory measures of aggressive behavior, which are necessarily constrained by ethical guidelines, measures of prosocial behavior can allow for greater correspondence inside and outside the lab (e.g., donations to charity, helping another person). Although earlier attempts to assess prosocial behaviors occurred in more naturalistic settings (e.g., returning lost wallets: Hornstein et al., 1968 agreeing to donate blood: Anderson, 1983) or in laboratory settings (volunteering to take shock for another person: "Elaine paradigm": Batson et al., 1981), recent experimental studies assess prosocial behaviors using social dilemma games (Thielmann et al., 2020), willingness to help people inside and outside the lab (research assistant with another study, someone who dropped their pencils, and someone struggling to use crutches: e.g., Bushman & Anderson, 2009; Twenge et al., 2007; Van Baaren et al., 2004), and donating money or committing time to a charitable cause (Motsenok et al., 2022).

Similar to laboratory assessments of aggression, however, there are concerns about the extent to which there is correspondence between prosocial behaviors observed inside and outside the lab (Galizzi & Navarro-Martinez, 2019; Wispe, 1972). For instance, social preference games such as the prisoner's dilemma or the trust game are often used in the laboratory to assess prosocial and cooperative behaviors (Murnighan & Wang, 2016). However, a meta-analysis of 39 articles revealed that choices in social preference games are only weakly correlated with self-reported prosocial behaviors and helping behaviors in field settings (Galizzi & Navarro-Martinez, 2019). This work also revealed that there are significant differences in the way these games are framed, played, and evaluated across research studies, and these contextual differences can change the behaviors observed within the games. A more comprehensive meta-analysis involving 770 studies noted a small but positive relation between some prosocial personality traits and decisions in social preference games, although these relations were further moderated by the social affordances of the games studied (Thielmann et al., 2020).

The Tangram Help/Hurt Task[2]

The Tangram Help/Hurt Task (THHT) was designed to circumvent many of the concerns outlined for previous measures of aggressive and prosocial behaviors. The cover story states that the purpose of this experiment is to examine how people's personalities affect their performance on cognitive tasks, one being the THHT. In the THHT, participants are asked to assign 11 tangrams from a table containing 10 easy, 10 medium, and 10 hard tangrams to an ostensible participant who has an opportunity to win a gift certificate (e.g., US$10) if they complete 10 of the 11 tangrams assigned within the time limit (e.g., 10 min; see Saleem, Anderson, et al., 2015; Saleem, Prot, et al., 2015; for a detailed description). In the instructions, participants are told that they can assign any tangrams they wish to the other participant who will not see them or know who they are, thereby reducing concerns related to reciprocity and/or retaliation. Furthermore, participants are told that they do not have an option to win a gift certificate themselves because one of the objectives of this research is to examine whether prizes influence cognitive task performance. Participants could select mostly easy puzzles and improve the other participants' chances of winning the gift certificate, choose mostly hard puzzles and make it difficult for the other participant to win the gift certificate, or choose a range of puzzles from three level-of-difficulty categories. Consequently, hurting (helping) in this task constitutes the number of hard (easy) puzzles assigned to the other participant minus one, with negative scores converted to zero[3]. The value one is subtracted because the design of the task prohibits participants from selecting puzzles exclusively from one difficulty level (i.e., they choose 11 puzzles from among the 10 easy, 10 medium, and 10 difficult puzzles).

The medium difficulty puzzles were included in this task to provide a neutral response option that is neither helpful nor hurtful, but puzzles selected from this category are not scored and analyzed for three reasons. First, mathematically, if one calculates helping and hurting scores, they can automatically deduce the number of medium puzzles assigned. Second, the inclusion of medium puzzles within the scoring paradigm would further increase the interdependence of helping and hurting scores. Third, we chose to focus on responses (i.e., number of easy and hard puzzles) that unambiguously reflected intentions to help and harm another person as reflected by the motivation measures assessed post tangram selection and correlations with established trait measures of prosocial and aggressive outcomes. Indeed, the selection of medium puzzles (a) tends to correlate more strongly with motivations to provide a range of tangrams rather than motivations to help or harm the other person and (b) does not strongly or consistently correlate with traits associated with aggressive and prosocial behavior.

Further specifications can be made to operationalize hurting behavior on the THHT as an indirect (participants do not see each other), active (participants are solely responsible for the number of hard puzzles they assign), and relatively mild form of aggressive behavior. Similarly, the particular kind of prosocial behavior assessed in the THHT is consistent with the definition of helping provided by Dovidio (1984). Specifically, it is a voluntary act (participants can choose any 11 puzzles from the assignment table) performed to provide some benefit to another person from the perspective of the benefactor (participants' report selecting easy puzzles to help the other person), not necessarily the recipient. Finally, it is a relatively low-cost helping behavior compared with

high-cost behaviors such as donating money or volunteering time.

Evidence for the construct validity of the THHT has been provided in several articles (Gentile et al., 2009; Saleem et al., 2012; Saleem, Anderson, et al., 2015; Saleem, Prot, et al., 2015; Saleem et al., 2017). Briefly, help and hurt scores on the THHT are significantly correlated in the predicted direction with established trait measures of aggression, prosocialness, empathy, perspective taking, state hostility, narcissism, agreeableness, and control aggression schemas. In addition, short-term manipulations are known to increase aggressive (e.g., provocation), and prosocial (e.g., empathy) behaviors are significantly associated with helpful and hurtful scores on the THHT. Furthermore, helpful and hurtful scores are associated with existing laboratory measures of aggressive and prosocial outcomes (e.g., hot-sauce paradigm, impression evaluations, monetary incentives). Finally, scores on the THHT are not significantly associated with social desirability, achievement motivation, emotion regulation, and perception of THHT difficulty, providing evidence for discriminant validity.

Although there are several methodological advantages of the THHT that have been discussed in previous articles (Saleem, Anderson, et al., 2015; Saleem, Prot, et al., 2015; Saleem et al., 2017), we highlight three points that specifically address some of the concerns outlined by critics of laboratory measures of aggressive and prosocial behaviors. First, unlike some existing measures of aggressive and prosocial behaviors that have been criticized for not offering alternative responses, the THHT gives participants prosocial, aggressive, *and* nonaggressive options in the form of selecting tangrams from the easy, hard, and medium difficulty categories, respectively.

Second, concerns surrounding the cover story influencing motives and responses are addressed. Specifically, in the THHT participants have the opportunity to help or harm the other participant, but the cover story does not draw attention to, encourage, or deter participants from choosing any particular tangrams. Furthermore, because the participant cannot interact with the other participant and does not have the potential to earn a prize, and because the other participant cannot reciprocate the behavior, there is no potential for competitive or cooperative motives to be induced. Moreover, social desirability concerns have not been found to influence tangram selection choices (Saleem, Anderson, et al., 2015; Saleem, Prot, et al., 2015; Saleem et al., 2017).

Third, the concern about participants' motivations behind the behaviors has already been addressed in prior THHT studies. Indeed, an examination of these data reveals an underlying motivation to harm (help) the other person's chances of winning the prize as the basis of selecting mostly hard (easy) tangrams (Saleem, Anderson, et al., 2015; Saleem, Prot, et al., 2015; Saleem et al., 2017).

Nevertheless, a thorough review of studies using the THHT is needed to examine the effectiveness, strengths, and limitations of this task based on existing evidence. Classical measurement theory stipulates that multiple pieces of converging evidence from a large number of domains must be accumulated over time to demonstrate that the data produced by the instrument in question indeed reflects, as best as possible, the construct it was designed to assess (Crocker & Algina, 1986). Accordingly, we conducted a meta-analysis of existing research using the THHT. Importantly, guided by the recommendations made in previous criticisms of laboratory aggression and prosocial measures (e.g., Galizzi & Navarro-Martinez, 2019; McCarthy & Elson, 2018), we examined the role of motivations for helping and hurting on the THHT and preregistered our hypotheses and analysis plan in accordance with open-science practices.

Method

This meta-analytic review was preregistered https://osf.io/97wjr/

Literature Search

Seven computer databases were used to locate relevant studies published through June 28, 2022: (a) PsycINFO, (b) Medline, (c) Educational Resources Information Center (ERIC), (d) PubMed, (e) ProQuest, (f) Theses Global, and (g) BASE (Bielefeld Academic Search Engine). We searched the latter three databases for unpublished studies to address potential publication bias (i.e., the file drawer problem; Rosenthal, 1979). We used two search terms (i.e., *tangram** OR THHT) for the literature search because we did not want to miss any studies that used the Tangram Help/Hurt Task (THHT; Saleem, Anderson, et al., 2015; Saleem, Prot, et al., 2015; Saleem et al., 2017). The asterisk allows terms to have all possible endings (e.g., *tangram*, *tangrams*) and all phrases that include the word "tangram" (e.g., *tangram puzzles*, *Tangram Help/Hurt Task, tangram task*). Because we were specifically concerned with the validity of the THHT, we excluded any analog procedures, such as having participants complete difficult anagrams or impossible number sequences (e.g., Lutz, 2015).

We also searched Google Scholar and the Social Science Citation Index for any studies that cited either of the two articles that first described the tangram task (i.e., Saleem, Anderson, et al., 2015; Saleem, Prot, et al., 2015; Saleem et al., 2017) or the first study that used the task (Gentile et al., 2009).

To obtain studies that we might have missed, three additional steps were taken. First, we searched the reference sections of all retrieved studies. Second, we sent an announcement requesting unpublished and published THHT studies to seven Listservs: (a) European Association of Social Psychology, (b) International Society for Research on Aggression, (c) Society of Australasian Social Psychologists,

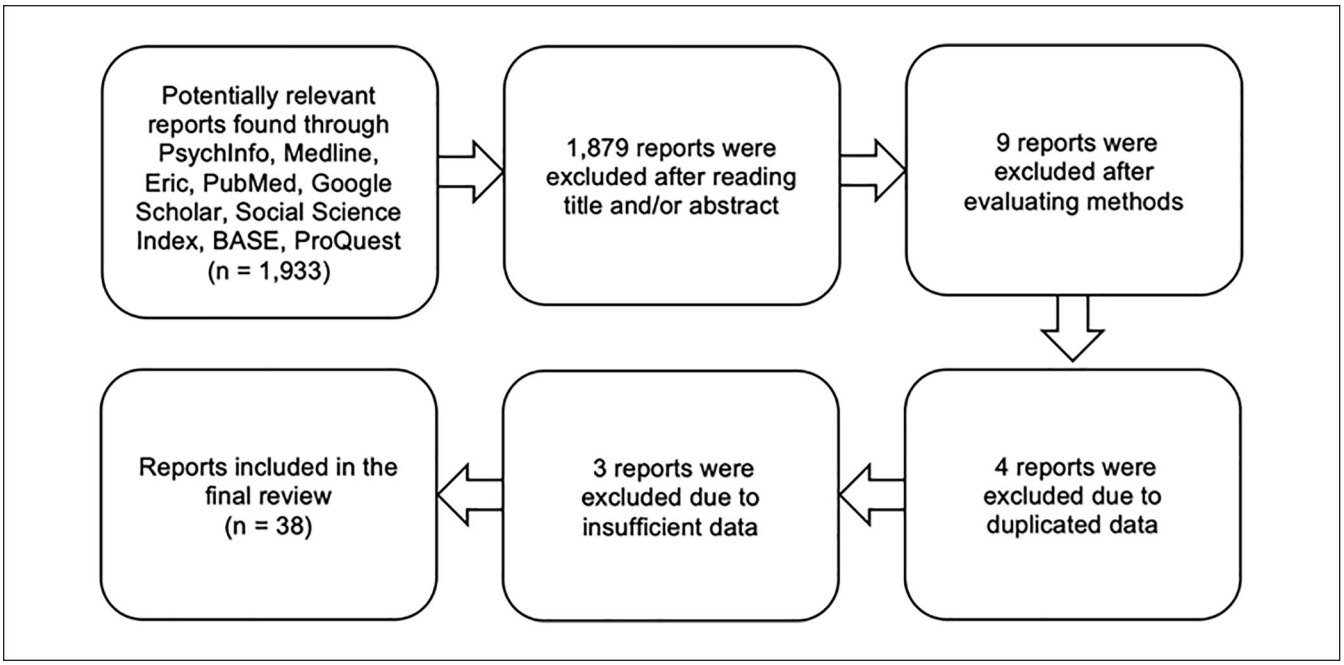

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart of literature search.

(d) Society of Experimental Social Psychology, (e) Society for Personality and Social Psychology (Division 8 of APA), (f) Society for the Psychological Study of Social Issues, and (g) Society for the Study of Peace, Conflict and Violence: Peace Psychology Division (Division 48 of APA). Third, we contacted all researchers who had conducted a study that used the THHT and requested from them any published and unpublished studies that used the task.

Preferred Reporting Items for Systematic Reviews and Meta-Analyses Flowchart

This search produced 1,933 research reports, but not all of them were relevant to this meta-analysis. To determine whether articles were relevant, we read their titles and abstracts. As shown in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart in Figure 1, the final sample included 38 articles (marked with an asterisk in the *References* section), which included 52 independent studies involving 11,060 participants. A Table containing characteristics of these studies is included in Supplemental Materials.

Type of Tangram Task

We coded whether the researchers used the THHT to measure helping (i.e., number of easy tangram puzzles assigned), hurting (i.e., number of hard tangram puzzles assigned), helping-hurting (i.e., number of easy puzzles minus number of hard puzzles), or hurting-helping (i.e., number of hard puzzles minus the number of easy puzzles). The focus of this article was to establish whether the THHT provides a valid measure of helping and hurting. The results for helping-hurting are in Supplemental Materials. The results for hurting-helping were too few to analyze, although we multiplied helping-hurting scores by −1 to obtain hurting-helping scores and combined these studies.

Outcome Variables

Helping/hurting independent variables. We coded whether authors manipulated independent variables predicted to increase helping behavior (e.g., playing a prosocial vs. neutral video game), which we expected to increase helping and decrease hurting on the tangram task. We also coded whether authors manipulated independent variables predicted to increase hurting behavior (e.g., playing a violent vs. neutral video game), which we expected to increase hurting and decrease helping on the tangram task.

Helping/hurting motivations. We coded whether authors measured helping motivations (e.g., "I wanted to help the other participant win the prize"), which we expected to be positively correlated with helping and negatively correlated with hurting on the tangram task. We also coded whether authors measured hurting motivations (e.g., "I wanted to hurt the other participant's chances of winning the gift certificate"), which we expected to be positively related to hurting and negatively related to helping on the tangram task.

Helping/hurting traits. Traits are enduring personal characteristics, individual differences, or patterns of behavior that are

relatively stable over time and across situations. We coded whether the authors measured traits related to helping (e.g., trait prosocialness, trait empathy, perspective taking), which we expected to be positively related to helping and negatively related to hurting on the tangram task. We also coded whether the authors measured traits related to hurting (e.g., trait aggressiveness, narcissism, control aggression schemas), which we expected to be positively related to hurting and negatively related to helping on the tangram task.

Helping/hurting states. States are temporary internal responses, such as thoughts and feelings. We coded whether authors measured states related to helping others (e.g., empathy, compassion, perspective taking), which we expected to be positively related to helping and negatively related to hurting on the tangram task. We also coded whether authors measured states related to hurting others (e.g., anger, hostility, aggressive cognition), which we expected to be positively related to hurting and negatively related to helping on the tangram task.

Other helping/hurting acts. We coded whether authors included other behavioral measures of helping others (e.g., picking up dropped pencils), which we expected to be positively related to helping and negatively related to hurting on the tangram task. We also coded whether authors included other behavioral measures of hurting others (e.g., giving a partner loud noise blasts, electric shocks, spicy food to eat, impossible number sequences to solve), which we expected to be positively related to hurting and negatively related to helping on the tangram task.

Moderators of Helping and Hurting on the THHT

We also examined several possible moderators of helping and hurting on the THHT (see preregistration). However, there were not enough studies to assess their impact on outcome variables. In addition, these moderator variables were often confounded. There were some moderator variables we could analyze (i.e., whether the study was conducted in our own labs or affiliated labs [i.e., labs of people we have published with] versus other labs; whether researchers probed for suspicion; whether participants practiced the tangram puzzles; and the amount of money offered as a reward for the partner solving the tangram puzzles), which we describe briefly in the Results section and report in detail in Supplemental Materials.

Intercoder Reliability

All studies were coded by two independent raters (the lead author and trained undergraduate research assistants), called "double coding" (Cooper, 2016). To assess intercoder reliability, the intraclass coefficient was used for continuous characteristics and the kappa coefficient was used for categorical characteristics (Vevea et al., 2019). The median reliability coefficient was .68 and the median percentage agreement was 83%. Coding disagreements were resolved via discussion between the first and last authors.

Analysis Strategy

The correlation coefficient was used as the effect-size index. Because the distribution of the correlation coefficient is not normally distributed unless the population correlation coefficient equals zero, Fisher's z-transformation was applied to each correlation coefficient before pooling them. Each z-transformed value was weighted by the inverse of its variance (i.e., N-3). Thus, larger studies received more weight when effect-size estimates were combined.

Data were analyzed using *R* (R Core Team, 2021). Random-effects models were used, which assume that effect sizes differ from population means by both participant-level sampling error and study-level variability (Borenstein et al., 2011). Although random-effects models are more conservative than fixed-effects models, they require fewer statistical assumptions and allow for generalizations to a broader set of studies than only the ones included in the meta-analytic review.

Conventional meta-analytic methods rely on the assumption that effect sizes are independent (Hedges et al., 2010). In this meta-analytic review, the independence assumption was often violated because several studies provided more than one correlation, resulting in clusters of correlations from these respective studies. For example, most studies used the tangram task to measure both helping and hurting. There were 777 correlations from the 52 studies. Thus, we calculated robust variance estimates to adjust standard errors of dependent effect size point estimates and confidence intervals (Hedges et al., 2010; Tipton, 2013, 2015) with the metafor v. 2.4-0 (Viechtbauer, 2010) and robumeta v. 2.0 (Fisher et al., 2017) packages in *R* (R Core Team, 2021).

For distributions with fewer than 10 effect sizes, we only provide descriptive statistics. Most meta-analysts urge caution when interpreting results from distributions with fewer than 10 effect sizes (e.g., Kepes et al., 2012; Sterne et al., 2011).

We also conducted a comprehensive battery of sensitivity analyses to determine whether the obtained results were robust to publication bias and outliers (e.g., Kepes et al., 2017). The sensitivity analyses are summarized after the main results, with the details provided in Supplemental Materials.

Results

This meta-analytic review included 52 independent studies involving 11,060 participants. First, we describe the overall results (with all outcome variables combined) for helping and hurting on the THHT. Second, we describe each outcome variable for helping and hurting on the THHT.

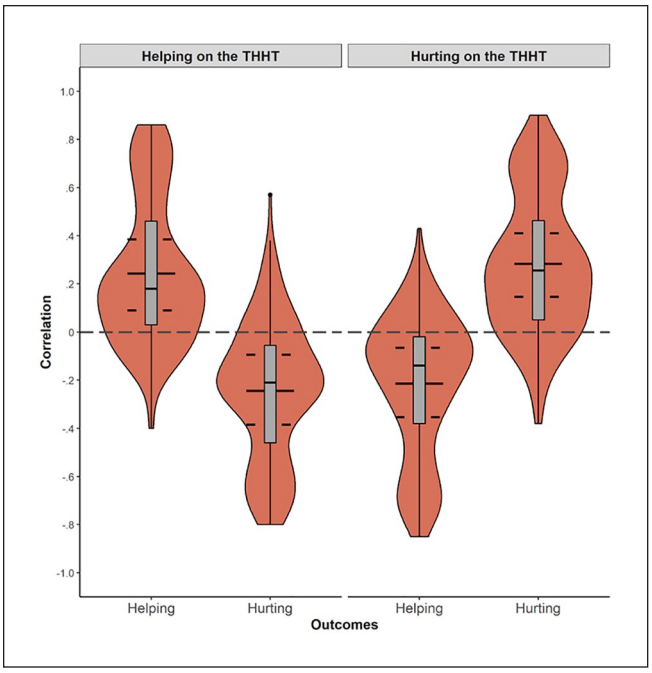

Figure 2. Violin plots.
Note. The density of the distribution at a given correlation is denoted by its width. The center of the violin plot contains a boxplot. The mean correlation is denoted by a solid line behind the box plot, and the 95% confidence interval bounds are denoted by dashed lines above and below the mean. A horizontal dashed line is drawn at zero.

Third, we describe the moderator variable analyses. Finally, we briefly describe the results from the sensitivity analyses.

Overall Results

As can be seen in the left-side violin plots in Figure 2, helping on the THHT was positively related to helping outcome variables and was negatively correlated with hurting outcome variables. As can be seen in the right-side violin plots in Figure 2, hurting on the THHT was positively related to hurting outcome variables and was negatively correlated with helping outcome variables. All four distributions are fairly symmetrical (i.e., the median and mean are similar, the median cuts the box in half, and the whiskers are about the same length).

Is the THHT a Valid Measure of Helping?

Other helping acts. As expected, other helping acts were positively related to helping ($r = .16$; 95% confidence interval [CI]: [.05, .27]; $k = 26$) and negatively related to hurting ($r = -.11$; 95% CI: [−.18, −.03]; $k = 28$) on the THHT. Both 95% CIs excluded the value zero and did not overlap.

Helping predictor variables. As expected, variables predicted to increase helping behavior increased helping ($r = .04$; 95% CI: [−.10, .17]; $k = 15$) and decreased hurting ($r = -.07$; 95% CI: [−.18, .04]; $k = 15$) on the THHT, although neither correlation differed significantly from zero and the 95% CIs overlapped. Interestingly, we ran additional analyses on the few studies that (a) *experimentally manipulated* an independent variable expected to increase helping behavior; and (b) used the original scoring method. These short-term experiments yielded a significant increase in helping, ($r = .14$; 95% CI: [.08, .20]; $k = 11$) and a significant decrease in hurting ($r = -.19$; 95% CI: [−.30, −.07]; $k = 10$).

Helping motivations. As expected, helping motivations were positively related to helping ($r = .91$; 95% CI: [.85, .98]; $k = 32$) and negatively related to hurting ($r = -.85$; 95% CI: [−.91, −.79]; $k = 32$) on the THHT. Both 95% CIs excluded the value zero and did not overlap.

Helping states. As expected, helping states were positively related to helping ($r = .15$; 95% CI: [.08, .21]; $k = 53$) and negatively related to hurting ($r = -.14$; 95% CI: [−.20, −.07], $k = 63$) on the THHT. Both 95% CIs excluded the value zero and did not overlap.

Helping traits. As expected, helping traits were positively related to helping ($r = .15$; 95% CI: [.10, .20]; $k = 31$) and negatively related to hurting ($r = -.13$; 95% CI: [−.18, −.08]; $k = 31$) on the THHT. Both 95% CIs excluded the value zero and did not overlap.

Is the THHT a Valid Measure of Hurting?

Other hurting acts. As expected, other hurting acts were positively related to hurting ($r = .26$; 95% CI: [.17, .35]; $k = 20$) and negatively related to helping ($r = -.28$; 95% CI: [−.38, −.18] $k = 18$) on the THHT. Both 95% CIs excluded the value zero and did not overlap.

Hurting predictor variables. As expected, independent variables predicted to increase hurting behavior increased hurting ($r = .12$; 95% CI: [.04, .20], $k = 24$) and decreased helping ($r = -.03$; 95% CI: [−.14, .08], $k = 24$) on the THHT, although the correlation was nonsignificant for helping. The 95% CIs also overlapped. As with "*Helping predictor variables*," an analysis of just the *experimental* studies yielded larger effects. Experimental manipulations expected to increase hurting behavior (using the original scoring method) yielded both a significant increase in hurting *(r* = .26; 95% CI: [.19, .34]; $k = 10$) and a significant decrease in helping ($r = -.17$; 95% CI: [−.22, −.10]; $k = 10$).

Hurting motivations. As expected, hurting motivations were positively related to hurting ($r = .79$; 95% CI: [.70, .88]; $k = 46$) and negatively related to helping ($r = -.76$; 95% CI: [−.83, −.69]; $k = 36$) on the THHT. Both 95% CIs excluded the value zero and did not overlap.

Hurting states. As expected, hurting states were positively related to hurting ($r = .17$; 95% CI: [.11, .23]; $k = 59$) and negatively related to helping ($r = -.13$; 95% CI: [−.19, −.07]; $k = 49$) on the THHT. Both 95% CIs excluded the value zero and did not overlap.

Hurting traits. As expected, hurting traits were positively related to hurting ($r = .19$; 95% CI: [.13, .24]; $k = 39$) and negatively related to helping ($r = -.14$; 95% CI: [−.19, −.09]; $k = 35$) on the THHT. Both 95% CIs excluded the value zero and did not overlap.

Moderator Variable Analyses

A detailed review of the moderator variable analyses is presented in the Supplementary Materials. One important moderator was whether the study was conducted in own labs or affiliated labs (i.e., labs of people we have published with) versus other labs. We suspect the larger effects pertaining to affiliated labs is a function of fidelity to the original paradigm and the use of best practice procedures (e.g., convincing cover story, funnel debriefing to assess and exclude suspicious participants, see details below in **Suggestions for Using the THHT**). Indeed, we conducted a post hoc 2 × 2 chi-square between lab affiliation and the use of best practice procedures. As expected, the percentage of studies that used best practice procedures was higher for affiliated labs (73%) than for unaffiliated labs (55%), which is an 18% difference. Although the chi-square test was not quite significant, $\chi^2(N = 96, df = 1) = 3.506$, $p = .061$, the phi coefficient was not trivial in size, $\phi = .191$. In terms of other moderators, we found larger effects when more valuable prizes were used, when suspicious participants were excluded, and when participants practiced tangrams before assigning them to another.

Sensitivity Analyses

We conducted a comprehensive sensitivity analysis to assess whether publication bias, outliers, or both affected our results. Both publication bias and outliers can adversely affect meta-analytic results and associated conclusions (Kepes et al., 2013). In fact, publication bias had been referred to as the potentially greater threat to the validity of meta-analytic results (Rothstein et al., 2005). All sensitivity analyses were conducted using *R* syntax (e.g., Kepes & McDaniel, 2015; Viechtbauer, 2017). The results from the sensitivity analyses are displayed in Table SM1 of Supplemental Materials.

Neither publication bias nor outliers appear to have a substantial influence on helping or hurting THHT effect sizes. However, some THHT helping distributions were affected by publication bias and/or outliers. The naive mean correlation for helping acts ($r = .16$) might be overestimated by an average of .01 to .08 (11%–50%) and the naive mean correlation for helping states ($r = .15$) might be overestimated by an average of .00 to .02 (0%–22%). Likewise, some THHT hurting distributions were affected by publication bias and/or outliers. The naïve mean correlation for hurting acts ($r = .25$) might be overestimated by an average of .01 to .03 (6–13%). The naive mean correlation for hurting independent variables ($r = .12$) might be overestimated by an average of .04 (33%).

Discussion

Main Findings

Overall, the results revealed that the THHT is a valid measure of both helping and hurting behavior. Helping scores on the THHT were significantly and positively associated with other helping acts (e.g., prosocial behaviors), helping states (e.g., positive mood), helping traits (e.g., empathy), and helping motivations (e.g., wanting the other participant to win money). Similarly, hurting scores on the THHT were significantly and positively associated with other hurting acts (e.g., aggressive behaviors), short-term independent variables (e.g., brief experimental manipulations known to influence hurting), hurting states (e.g., state hostility), longer term personality traits (e.g., trait aggressiveness), and motivations associated with hurting (e.g., wanting to prevent the other participant from getting any money).

Across all types of predictor variables (i.e., experimental, cross-sectional, and longitudinal) using various THHT administration and scoring methods hypothesized (by the original study authors) to yield positive associations with helping scores and negative associations with hurting scores, results were in the predicted direction but nonsignificant. However, the smaller subset of *experimental* studies did yield significant effects on both increasing help scores and decreasing hurt scores.

The results might appear to suggest that the effects associated with the aggression measure (hurt) from THHT were systematically larger than the prosocial behavior measure (help). However, a robust comparison of help versus hurt effects is not feasible at present; the studies differed on the types of primes used, sample sizes, the offline/online context, and several other factors, making such a comparison impossible. We prefer to stick with the more general conclusion that the THHT is an acceptable measure of both aggressive and prosocial behavior and leave consideration of whether it is a better measure of aggressive than of prosocial behavior to future research.

Sensitivity analyses revealed that publication bias did not significantly alter THHT effect sizes, although some of the distributions may be overestimated.

Important Advantages of the THHT

Several well-established laboratory paradigms assessing aggressive and prosocial behaviors have been validated in many ways (e.g., Anderson & Bushman, 1997; Bernstein et al., 1987; Carlson et al., 1989; Giancola & Chermack, 1998; Giancola & Zeichner, 1995; Hyatt et al., 2019; Parrott & Giancola, 2007), with most attention devoted to the Competitive Reaction Time Task (Warburton & Bushman, 2019). Nonetheless, the THHT is another useful tool that offers four important advantages and strengths. First, the THHT allows simultaneous assessment of helping and hurting behavior and therefore is useful for contexts where researchers are interested in studying these behaviors concurrently (e.g., Saleem et al., 2012).

Second, the THHT includes a medium category, allowing for a response that is neither aggressive nor prosocial. The ability for participants to assign tangrams from the medium category serves as an unambiguous neutral point, a choice that reflects neither an intention to help nor an intention to hurt the other person. The option of selecting a choice that neither reflects hurting nor helping is also useful in other research areas such as in intergroup relations where it is important to distinguish behaviors representing in-group favoritism from out-group derogation (see Greenwald & Pettigrew, 2014 for details).

Third, researchers have the flexibility to evaluate what kind of scoring method is appropriate for their particular research design. For some studies, it may be appropriate to use an overall difference score that forces helping and hurting to be interdependent and on opposite ends of a single continuum, such as examining the extent to which participants choose helping over hurting (or vice versa) in a given context. For other studies, it may be more appropriate to use only helping or hurting scores (e.g., Barlett & Anderson, 2011; Saleem, Anderson, et al., 2015; Saleem, Prot, et al., 2015). This is especially useful when the predicted outcome is specific to helping or hurting behavior but not necessarily both. For other researchers who are interested in helping and hurting behavior but are worried about the issue of interdependence and multicollinearity, it might make sense to use stronger adjustments of what constitutes helping and hurting behavior (e.g., helping is the number of easy puzzles −2 or −3). It is important to note that although the THHT allows flexibility for researchers to decide which scoring method best aligns with their theoretical questions of interest, this does not relax standards of scientific integrity regarding how and which results to report. On the contrary, we have recommended that if scholars are unsure about which scoring method they find most appropriate for their theoretical propositions, they should report all three (helping score, hurting score, and the difference score; Saleem, Anderson, et al., 2015; Saleem, Prot, et al., 2015; Saleem et al., 2017).

Fourth, the THHT is easily administered in a variety of different contexts (e.g., in the lab, online, in the field) and with different samples (e.g., children and non-English speakers) without much time, effort, or resource. Thus, it is a highly adaptable paradigm.

Limitations

There also are at least four limitations of the THHT. First, the design of the task (e.g., can only choose 11 tangrams and there are 10 per difficulty level) constraints scores such that higher scores on helping will yield lower scores on hurting, and vice versa. In other words, helping and hurting scores will be interdependent and negatively correlated, especially at extreme values. Several recommended strategies to ameliorate these concerns include: (a) taking the number of easy (hard) puzzles and subtracting zero, (b) setting negative scores to zero, and (c) including both scores in the models to examine their unique variance and account for multicollinearity. Despite these strategies, we continue to observe a moderate correlation between help and hurt scores on the THHT (Saleem, Anderson, et al., 2015; Saleem, Prot, et al., 2015; Saleem et al., 2017).

Second, the THHT assesses a relatively milder form of aggressive and prosocial behaviors compared with other established measures such as the Competitive Reaction Time Task for assessing physically aggressive behaviors (Warburton & Bushman, 2019) or assessing high-cost prosocial behaviors such as donating and volunteering behaviors (Motsenok et al., 2022), which have a higher correspondence to real world acts of prosocial behavior.

Third, as with virtually all measures of aggressive and prosocial behavior, the obtained effect sizes using the THHT tend to be in the small to moderate range. This is not surprising, given the plethora of variables—both situational and individual—that are known to influence aggressive and prosocial behaviors.

Fourth, as the THHT is relatively new, more research is needed using this task in different contexts and with different populations to better understand its generalizability.

Future Research Directions

New research studies using the THHT are also needed to understand (a) the specific kinds of helping and hurting outcomes best predicted by this task; (b) how sociodemographic, methodological, and contextual variables moderate the effects obtained with this task; and (c) the extent to which THHT scores reflect state- or trait-like characteristics depending on how much these scores vary within individuals and across situations. Although we have evidence for the intent behind helping and hurting choices on the THHT, future studies can also expand this work by examining (a) a range of intentions and motivations behind tangram assignment choices potentially using open-ended qualitative responses prior to or post tangram assignment, (b) how self-reported intentions to help and hurt on the THHT correspond

to indirect assessments of helping and hurting intentions (e.g., neuroimaging data), (c) whether assessments of intent or motivation to help or hurt on the THHT change pre and post tangram assignment, (d) participants' meta-perceptions of the receipt's expectations and intentions with respect to the prize, and (e) how perceptions of ability to complete tangram puzzles, both from the perspective of the participant and their expectations for the other participant, could influence THHT choices and intentions. Given the importance of intent in understanding prosocial and aggressive behaviors, it is worthwhile to continue to systematically study the role of intent in understanding choices on the THHT.

Suggestions for Using the THHT

Although there are likely to be some variations in the way scientific tools are adapted and used by scholars in the field, we offer five suggestions on how best to use the THHT. First, similar to other interaction paradigms, the underlying presumption of the THHT is that the participant believes there to be another participant to whom they are assigning the tangrams to. This assumption can be created and reinforced through a cover story and validated through a funnel debriefing process in which participant suspicion is assessed (Boynton et al., 2013). We recommend that participants who did not believe they were actually interacting with another person based on their debriefing responses should be excluded from the analysis of the THHT. We found larger effects when suspicious participants were excluded (see Supplemental Materials). Second, participants should solve a few practice tangram puzzles (at least one from each difficulty level) in front of the researcher to better gauge the difficulty of solving tangrams from easy, medium, and hard categories. Although there are tangram manipulation programs[4], this may be difficult to implement in online studies without having advanced programming skills. We found larger effects when participants practiced tangram puzzles (see Supplemental Materials). Third, participants should be asked if they understand how to use the THHT. In person, this provides the researcher another opportunity to review the THHT if the participant does not fully understand the task. Online, this question could serve as an attention check and would justify the exclusion of this participant from data analysis. Fourth, researchers should ask questions assessing participants' intentions and motivations for their tangram assignments (e.g., intent to help/hurt the other person, challenge them, etc.). Fourth, the ostensible partner should be offered a prize. We found larger effects for more valuable prizes (see Supplemental Materials). Fifth, helping and hurting scores should be calculated by subtracting 1 from the total number of easy and hard puzzles and restricting negative scores to 0. This scoring method not only accounts for the restrictions of the task (have to select 11 tangrams but there are only 10 per difficulty category) but also reduces the interdependence between help and hurt scores. If multiple scoring methods are analyzed, then we recommend reporting results pertaining to all of them.

Conclusion

Overall, the goal of the present research was to provide meta-analytic tests of the validity of the THHT as a laboratory assessment of helping and hurting behavior. Results revealed that scores on the THHT are associated with helping and hurting outcomes in theoretically predicted ways. Although there are well-validated laboratory assessments of prosocial and antisocial behavior, to our knowledge, none of these have the option to simultaneously assess both types of behaviors while allowing for a neutral option. Accordingly, the THHT serves as a valuable addition to our scientific tool kit to better understand prosocial and aggressive behaviors.

Acknowledgments

We would like to thank Kriti Chawla, Sarah C. Lightle, Tess Wells, and Yuwen Wan for their help coding studies, and Emily Davidson for her help preregistering this meta-analytic review.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Craig A. Anderson https://orcid.org/0000-0001-6353-0023
Brad J. Bushman https://orcid.org/0000-0002-1266-5101

Supplemental Material

Supplemental material is available online with this article.

Notes

1. For simplicity, by "independent variable" we mean the variable that is *hypothesized* to cause an effect on the "dependent variable," though of course correlational studies, especially cross-sectional correlational studies, do not lend as much strength to the causal claim as do true experimental studies because the "independent variable" is measured rather than manipulated.
2. Tangrams are two-dimensional jigsaw puzzles in which the goal is to create a specific shape using the available puzzle pieces.
3. Although it is mathematically possible to receive a negative score for helping and hurting if a participant chooses zero puzzles from the easy or hard categories, we recommend that negative scores be converted to zero to: (a) account for the design restrictions of the task and (b) reduce the correlation between helping and hurting scores.
4. Tangram Puzzles for Kids. ABCya: https://www.abcya.com/games/tangrams

References

Research reports included in the meta-analytic review are marked with an asterisk (*).

Anderson, C. A. (1983). Motivational and performance deficits in interpersonal settings: The effect of attributional style. *Journal of Personality and Social Psychology*, *45*(5), 1136–1147. https://doi.org/10.1037/0022-3514.45.5.1136

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. https://doi.org/10.1037//1089-2680.1.1.19

*Andrews, M. W. (2016). *Investigating prosocial gameplay and prosocial self-concept*. http://hdl.handle.net/11603/3100

Bader, F., Baumeister, B., Berger, R., & Keuschnigg, M. (2021). On the transportability of laboratory results. *Sociological Methods & Research*, *50*(3), 1452–1481. https://doi.org/10.1177/0049124119826151

*Barlett, C. P., & Anderson, C. A. (2011). Reappraising the situation and its impact on aggressive behavior. *Personality and Social Psychology Bulletin*, *37*(12), 1564–1573. https://doi.org/10.1177%2F0146167211423671

*Barlett, C. P., Helmstetter, K. M., Kowalewski, D. A., & Pezzillo, L. (2017). Piecing together the aggression puzzle: Testing the mediating variables linking early to later aggression. *Aggressive Behavior*, *43*(6), 523–530. https://doi.org/10.1002/ab.21710

*Barlett, C. P., Madison, C., DeWitt, C. C., & Bailey Heath, J. (2019). The moderating effect of dispositional fear of retaliation on the relationship between provocation and aggressive behavior. *Personality and Individual Differences*, *138*, 257–265. https://doi.org/10.1016/j.paid.2018.10.008

Baron, R., & Richardson, D. S. (1994). *Human aggression*. Plenum Press.

Bar-Tal, D., & Raviv, A. (1982). A cognitive-learning model of helping behavior development: Possible implications and applications. In N. Eisenberg (Ed.), *The development of prosocial behavior*. Academic Press.

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

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

*Behler, A. M. C., Wall, C. S. J., Bos, A., & Green, J. D. (2020). To help or to harm? Assessing the impact of envy on prosocial and antisocial behaviors. *Personality and Social Psychology Bulletin*, *46*(7), 1156–1168. https://doi.org/10.1177/0146167219897660

Bernstein, S., Richardson, D., & Hammock, G. (1987). Convergent and discriminant validity of the Taylor and Buss measures of physical aggression. *Aggressive Behavior*, *13*(1), 15–24. https://doi.org/10.1002/1098-2337(1987)13:1<15::AID-AB2480130104>3.0.CO;2-K

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. Wiley.

Boynton, M. H., Portnoy, D. B., & Johnson, B. T. (2013). Exploring the ethics and psychological impact of deception in psychological research. *IRB*, *35*(2), 7–13.

*Brown, D. (2021). [Unpublished raw data on description of data]. Brigham Young University.

Burnett Heyes, S. (2020). Just banter? Friendship, teasing and experimental aggression in adolescent peer networks. *Developmental Science*, *23*(3), e12926. https://doi.org/10.1111/desc.12926

Bushman, B. J., & Anderson, C. A. (2009). Comfortably numb: Desensitizing effects of violent media on helping others. *Psychological Science*, *20*, 273–277. https://doi.org/10.1111/j.1467-9280.2009.02287.x

Bushman, B. J., & Huesmann, L. R. (2010). Aggression. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Ch. 23, pp. 833–863). John Wiley & Sons.

Carlson, M., Marcus-Newhall, A., & Miller, N. (1989). Evidence for a general construct of aggression. *Personality and Social Psychology Bulletin*, *15*(3), 377–389.

*Chang, J. H. (2018). The killer: Moral choice in virtual environments [Master's thesis, Ohio State University]. *Ohiolink Electronic Theses and Dissertations Center*. http://rave.ohiolink.edu/etdc/view?acc_num=osu1532035898806428

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation* (pp. 37–94). Houghton Mifflin.

Cooper, H. (2016). *Research synthesis and meta-analysis: A step-by-step approach* (5th ed.). SAGE.

*Costello, T. H., Bowes, S. M., Stevens, S. T., Waldman, I. D., Tasimi, A., & Lilienfeld, S. O. (2022). Clarifying the structure and nature of left-wing authoritarianism. *Journal of Personality and Social Psychology*, *122*(1), 135–170. https://doi.org/10.1037/pspp0000341

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.

Denson, T. F., & Anderson, C. A. (in press). *Experimental methods. The Cambridge handbook of research methods and statistics for the social and behavioral sciences*. Cambridge University Press.

*Dopko, R. L., Capaldi, C. A., & Zelenski, J. M. (2019). The psychological and social benefits of a nature experience for children: A preliminary investigation. *Journal of Environmental Psychology*, *63*, 134–138. https://doi.org/10.1016/j.jenvp.2019.05.002

Dovidio, J. F. (1984). Helping behavior and altruism: An empirical and conceptual overview. *Advances in Experimental Social Psychology*, *17*, 361–427.

Elson, M., Mohseni, M. R., Breuer, J., Scharkow, M., & Quandt, T. (2014). Press CRTT to measure aggressive behavior: The unstandardized use of the competitive reaction time task in aggression research. *Psychological Assessment*, *26*(2), 419–432. https://doi.org/10.1037/a0035569

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). Robumeta: Robust variance meta-regression. https://CRAN.R-project.org/package=robumeta

Galizzi, M. M., & Navarro-Martinez, D. (2019). On the external validity of social preference games: A systematic lab-field study. *Management Science*, *65*(3), 976–1002. https://doi.org/10.1287/mnsc.2017.2908

*Gallucci, A., Riva, P., Romero Lauro, L. J., & Bushman, B. J. (2020). Stimulating the ventrolateral prefrontal cortex (VLPFC) modulates frustration-induced aggression: A tDCS experiment. *Brain Stimulation*, *13*(2), 302–309. https://doi.org/10.1016/j.brs.2019.10.015

*Gebhard, K. T. (2020). *Understanding how to prevent male violence: Investigating the role of threatened-masculinity shame and testing an intervention* (Order No. 13885838). Available from ProQuest Dissertations & Theses A&I; ProQuest

Dissertations & Theses Global. (2248633371). http://proxy.lib.ohio-state.edu/login?url=https://www-proquest-com.proxy.lib.ohio-state.edu/dissertations-theses/understanding-how-prevent-male-violence/docview/2248633371/se-2?accountid=9783

*Gentile, D. A., Anderson, C. A., Yukawa, S., Ihori, N., Saleem, M., Ming, L. K., . . .Sakamoto, A. (2009). The effects of prosocial video games on prosocial behaviors: International evidence from correlational, longitudinal, and experimental studies. *Personality and Social Psychology Bulletin*, *35*(6), 752–763. https://doi.org/10.1177%2F0146167209333045

Giancola, P. R., & Chermack, S. T. (1998). Construct validity of laboratory aggression paradigms: A response to Tedeschi and Quigley (1996). *Aggression and Violent Behavior*, *3*(3), 237–253. https://doi.org/10.1016/S1359-1789(97)00004-9

Giancola, P. R., & Zeichner, A. (1995). Construct validity of a competitive reaction-time aggression paradigm. *Aggressive Behavior*, *21*(3), 199–204. https://doi.org/10.1002/1098-2337(1995)21:3<199::AID-AB2480210303>3.0.CO;2-Q

*Gómez, Á., Bélanger, J. J., Chinchilla, J., Vazquez, A., Schumpe, B. M., Nisa, C. F., & Chiclana, S. (2021). Admiration for Islamist groups encourages self-sacrifice through identity fusion. *Humanities and Social Science Communications*, *8*(54). https://doi.org/10.1057/s41599-021-00734-9

*Gotowiec, S. (2019). When moral identity harms: The impact of perceived differences in core values on helping and hurting. *Personality and Individual Differences, 151*. https://doi.org/10.1016/j.paid.2019.06.032

Greenwald, A. G., & Pettigrew, T. F. (2014). With malice toward none and charity for some: Ingroup favoritism enables discrimination. *American Psychologist*, *69*(7), 669–684. https://doi.org/10.1037/a0036056

*Hawk, C. E., & Ridge, R. D. (2021). Is it only the violence? The effects of violent video game content, difficulty, and competition on aggressive behavior. *Journal of Media Psychology: Theories, Methods, and Applications*, *33*(3), 134–144. https://doi.org/10.1027/1864-1105/a000291

*Hawkins, I., Saleem, M., Gibson, B., & Bushman, B. J. (2021). Extensions of the proteus effect on intergroup aggression in the real world. *Psychology of Popular Media*, *10*(4), 478–487. https://doi.org/10.1037/ppm0000307

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

*Hodge, S., Taylor, J., & McAlaney, J. (2019). (A)morally demanding game? An exploration of moral decision-making in a purpose-made video game. *Media and Communication*, *7*(4), 213–225. https://doi.org/10.17645/mac.v7i4.2294

Hornstein, H. A., Fisch, E., & Holmes, M. (1968). Influence of a model's feeling about his behavior and his relevance as a comparison other on observers' helping behavior. *Journal of Personality and Social Psychology*, *10*(3), 222–226. https://doi.org/10.1037/h0026568

Hyatt, C. S., Chester, D. S., Zeichner, A., & Miller, J. D. (2019). Analytic flexibility in laboratory aggression paradigms: Relations with personality traits vary (slightly) by operationalization of aggression. *Aggressive Behavior*, *454*(4), 377–388. https://doi.org/10.1002/ab.21830

*Kearns, P. O., Tyler, J. M., & Graziano, W. G. (2021). More than fantasy: Prosocial daydreams relate to prosocial dispositions and behaviour. *International Journal of Psychology: Journal International De Psychologie*, *56*, 185–194. https://doi.org/10.1002/ijop.12673

Kepes, S., Banks, G. C., McDaniel, M. A., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, *15*, 624–662. https://doi.org/10.1177/1094428112452760

Kepes, S., Bushman, B. J., & Anderson, C. A. (2017). Violent video game effects remain a societal concern: Comment on Hilgard, Engelhardt, and Rouder (2017). *Psychological Bulletin*, *143*, 775–782. https://doi.org/10.1037/bul0000112

Kepes, S., & McDaniel, M. A. (2015). The validity of conscientiousness is overestimated in the prediction of job performance. *PLoS One*, *10*(10), e0141468. https://doi.org/10.1371/journal.pone.0141468

Kepes, S., McDaniel, M. A., Brannick, M. T., & Banks, G. C. (2013). Meta-analytic reviews in the organizational sciences: Two meta-analytic schools on the way to MARS (the Meta-analytic Reporting Standards). *Journal of Business and Psychology*, *28*(2), 123–143. https://doi.org/10.1007/s10869-013-9300-2

*Krumm, J. (2020). *A link to the future: Ostracism's effects on aggressive behavior* [Doctoral dissertation, The Ohio State University]. https://kb.osu.edu/bitstream/handle/1811/91610/1/Josiah_Krumm_thesis.pdf

Lin, H., Werner, K. M., & Inzlicht, M. (2021). Promises and perils of experimentation: The mutual-internal-validity problem. *Perspectives on Psychological Science*, *16*(4), 854–863. https://doi.org/10.1177/1745691620974773

Lutz, J. (2015). The validity of crowdsourcing data in studying anger and aggressive behavior: A comparison of online and laboratory data. *Social Psychology*, *47*(1), 38–51. https://doi.org/10.1027/1864-9335/a000256

*Mauersberger, H., Hess, U., & Hoppe, A. (2019). Measuring task conflicts as they occur: A real-time assessment of task conflicts and their immediate affective, cognitive, and social consequences. *Journal of Business and Psychology*, *35*(6), 813–830. https://doi.org/10.1007/s10869-019-09640-z

McCarthy, R. J., & Elson, M. (2018). A conceptual review of lab-based aggression paradigms. *Collabra: Psychology*, *4*(1), 1–12. https://doi.org/10.1525/collabra.104

Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science*, *7*(2), 109–117. https://doi.org/10.1177/1745691611432343

Motsenok, M., Kogut, T., & Ritov, I. (2022). Perceived physical vulnerability promotes prosocial behavior. *Personality and Social Psychology Bulletin*, *48*(2), 254–267. https://doi.org/10.1177/01461672211005879

Murnighan, J. K., & Wang, L. (2016). The social world as an experimental game. *Organizational Behavior and Human Decision Processes*, *136*, 80–94. https://doi.org/10.1016/j.obhdp.2016.02.003

*Ng, H. K. S., Hong, Y.-L., Chow, T. S., & Leung, A. N. M. (2019). Nature does not always give you a helping hand: Comparing the prosocial effects of nature at different resource and security

levels. *International Journal of Advanced Robotic Systems*, *45*(4). https://doi.org/10.1177/1729881417736895

Parrott, D. J., & Giancola, P. R. (2007). Addressing "the criterion problem" in the assessment of aggressive behavior: Development of a new taxonomic system. *Aggression and Violent Behavior*, *123*(3), 280–299. https://doi.org/10.1016/j.avb.2006.08.002

Penner, L. A., Dovidio, J. F., Pilavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annual Review of Psychology*, *56*, 365–392. https://doi.org/10.1146/annurev.psych.56.091103.070141

Pfattheicher, S., Nielsen, Y. A., & Thielmann, I. (2022). Prosocial behavior and altruism: A review of concepts and definitions. *Current Opinion in Psychology*, *44*, 124–129. https://doi.org/10.1016/j.copsyc.2021.08.021

Prot, S., & Anderson, C. A. (2013). Research methods, design, and statistics in media psychology. In K. Dill (Ed.), *The Oxford handbook of media psychology* (pp. 109–136). Oxford University Press.

*Rajchert, J. (2021a). [Unpublished raw data on description of data]. The Maria Grzegorzewska University.

*Rajchert, J. (2021b). [Unpublished raw data on description of data]. The Maria Grzegorzewska University.

*Rajchert, J., Konopka, K., & Boguszewski, P. (2018). Aggression and helping as responses to same-sex and opposite-sex rejection in men and women. *Evolutionary Psychology*, *16*(2). https://doi.org/10.1177/1474704918775253

*Rajchert, J., Żółtak, T., Szulawski, M., & Jasielska, D. (2019). Effects of rejection by a friend for someone else on emotions and behavior. *Frontiers in Psychology*, *10*, 764. https://doi.org/10.3389/fpsyg.2019.00764

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ritter, D., & Eslea, M. (2005). Hot sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *31*(5), 407–419. https://doi.org/10.1002/ab.20066

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analyses. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 1–7). Wiley. https://doi.org/10.1002/0470870168

*Saleem, M., Anderson, C. A., & Barlett, C. P. (2015). Assessing helping and hurting behaviors through the tangram help/hurt task. *Personality and Social Psychology Bulletin*, *41*(10), 1345–1362. https://doi.org/10.1177/0146167215594348

*Saleem, M., Anderson, C. A., & Gentile, D. A. (2012). Effects of prosocial, neutral, and violent video games on children's helpful and hurtful behaviors. *Aggressive Behavior*, *38*(4), 281–287. https://doi.org/10.1002/ab.21428

*Saleem, M., Barlett, C. P., Anderson, C. A., & Hawkins, I. (2017). Helping and hurting others: Person and situation effects on aggressive and prosocial behavior as assessed by the tangram task. *Aggressive Behavior*, *43*(2), 133–146. https://doi.org/10.1002/ab.21669

*Saleem, M., Prot, S., Cikara, M., Lam, B. C., Anderson, C. A., & Jelic, M. (2015). Cutting Gordian knots: Reducing prejudice through attachment security. *Personality and Social Psychology Bulletin*, *41*(11), 1560–1574. https://doi.org/10.1177/0146167215601829

*Schade, H. [Unpublished raw data on description of data].

Schroeder, D. A., & Graziano, W. G. (2015). The field of prosocial behavior: An introduction and overview. In D. A. Schroeder & W. G. Graziano (Eds.), *Oxford handbook of prosocial behavior* (pp. 3–34). Oxford University Press.

*Sedlar, A. (2019). The effects of experienced cyber-aggression on subsequent aggressive behavior among college students (Order No. 27539297). Available from ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global (2316417928). http://proxy.lib.ohio-state.edu/login?url=https://www-proquest-com.proxy.lib.ohio-state.edu/dissertations-theses/effects-experienced-cyber-aggression-on/docview/2316417928/se-2?accountid=9783

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., . . .Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta analyses of randomised controlled trials. *British Medical Journal*, *343*, 302–307. https://doi.org/10.1136/bmj.d4002

*Tear, M. J. (2015). Violent video games and social behaviour [PhD thesis, School of Psychology, The University of Queensland].

*Tear, M. J., & Nielsen, M. (2014). Video games and prosocial behavior: A study of the effects of non-violent, violent, and ultra-violent gameplay. *Computers in Human Behavior*, *41*, 8–13. https://doi.org/10.1016/j.chb.2014.09.002

Tedeschi, J. T., & Quigley, B. M. (1996). Limitations of laboratory paradigms for studying aggression. *Aggression and Violent Behavior: A Review Journal*, *1*, 163–177. https://doi.org/10.1016/1359-1789(95)00014-3

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, *146*(1), 30–90. https://doi.org/10.1037/bul0000217

Tipton, E. (2013). Robust variance estimation in meta-regression with binary dependent effects. *Research Synthesis Methods*, *4*(2), 169–187. https://doi.org/10.1002/jrsm.1070

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011.supp

Twenge, J. M., Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., & Bartels, J. M. (2007). Social exclusion decreases prosocial behavior. *Journal of Personality and Social Psychology*, *92*(1), 56–66. https://doi.org/10.1037/0022-3514.92.1.56

*Uskul, A. K., Cross, S. E., Günsoy, C., Gerçek-Swing, B., Alözkan, C., & Ataca, B. (2015). A price to pay: Turkish and Northern American retaliation for threats to personal and family honor. *Aggressive Behavior*, *41*(6), 594–607. https://doi.org/10.1002/ab.21598

*Van Damme, C., Deschrijver, E., Van, G. E., & Hoorens, V. (2017). When praising yourself insults others: Self-superiority claims provoke aggression. *Personality & Social Psychology Bulletin*, *43*(7), 1008–1019. https://doi.org/10.1177/0146167217703951

Van Baaren, R. B., Holland, R. W., Kawakami, K., & Van Knippenberg, A. (2004). Mimicry and prosocial behavior. *Psychological Science*, *15*(1), 71–74. https://doi.org/10.1111/j.0963-7214.2004.01501012.x

Vevea, J. L., Zelinsky, N. A. M., & Orwin, R. G. (2019). Evaluating coding decisions. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 173–204). Russell Sage Foundation.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

*Warburton, W. (2012). [Unpublished raw data on description of data]. Macquarie University.

Warburton, W. A., & Bushman, B. J. (2019). The competitive reaction time task: The development and scientific utility of a flexible laboratory aggression paradigm. *Aggressive Behavior*, *45*(4), 389–396. https://doi.org/10.1002/ab.21829

*Weber, R., Behr, K. M., Fisher, J. T., Lonergan, C., & Quebral, C. (2020). Video game violence and interactivity: Effect or equivalence? *Journal of Communication*, *70*(2), 219–244. https://doi.org/10.1093/joc/jqz048

Wispe, L. G. (1972). Positive forms of social behavior: An overview. *Journal of Social Issues*, *28*(3), 1–19. https://doi.org/10.1111/j.1540-4560.1972.tb00029.x

*Yang, Y., Yang, Z., Bao, T., Liu, Y., & Passmore, H. A. (2016). Elicited awe decreases aggression. *Journal of Pacific Rim Psychology*, *10*. https://doi.org/10.1017/prp.2016.8