

Assessing Helping and Hurting Behaviors Through the Tangram Help/Hurt Task

Muniba Saleem¹, Craig A. Anderson²,
and Christopher P. Barlett³

Personality and Social
Psychology Bulletin
2015, Vol. 41(10) 1345–1362
© 2015 by the Society for Personality
and Social Psychology, Inc
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0146167215594348
pspb.sagepub.com



Abstract

Across six studies, we validated a new measure of helpful and hurtful behaviors, the Tangram Help/Hurt Task. Studies 1 to 3 provided cross-sectional correlational convergent and discriminant validity evidence for the Tangram Task using college-based and adult online samples. Study 4 revealed that previously validated empathy primes increase helpful behaviors on the Tangram Task. Studies 5 and 6 revealed that previously validated provocation manipulations increase hurtful behaviors on the Tangram Task. The effects of various experimental manipulations on the Tangram Task were similar to or larger than on other established indices of helpful and hurtful behaviors. In addition, motivation items in all studies indicate that tangram choices are indeed associated with the intent of helping and hurting. We discuss the advantages and limitations of the Tangram Help/Hurt Task relative to established measures of helpful and hurtful behaviors.

Keywords

helpful, hurtful, prosocial, aggressive, validation, behavior

Received June 5, 2014; revision accepted June 11, 2015

The importance of measuring aggressive and prosocial behaviors in the laboratory is paramount for understanding such complex social behaviors. Anderson and Bushman (1997) discovered remarkable consistency in the size and direction of various effect sizes for aggression-based findings in the laboratory and in the “real-world.” This suggests that laboratory-based studies of aggressive behavior are generalizable to outside the laboratory and further exemplify why developing and validating high-quality measures of aggressive (and prosocial) behavior are important. The current research presents six studies that examined the validity of a laboratory-based tool designed to measure aggressive and prosocial behaviors simultaneously: the Tangram Help/Hurt Task.

Conceptually, prosocial and aggressive behaviors appear to be opposites, one involving behavior that helps another person, the other involving behavior that harms another person. Theoretically, they share the very important feature of being largely defined by *intent* rather than actual outcome (e.g., Anderson & Bushman, 2002; Batson, 2014; Malle, 2011). For this reason, perhaps, the psychological processes underlying both types of behavior may be very similar, involving both fairly automatic (impulsive) perception–decision–action sequences as well as more resource-intensive controlled sequences. Indeed, Hirsh, Galinsky, and Zhong (2011) suggested that the common factor behind seemingly

contradictory prosocial and antisocial outcomes is a disinhibited state characterized by reduced response conflict. This state of disinhibition influences the most salient response in a given context based on either trait tendencies or strong social cues (Hirsh et al., 2011). In addition, Graziano and Habashi (2010) provided an excellent theoretical integration of the similar processes underlying prejudice (which is strongly aligned to outgroup aggression) and prosocial behavior. Although testing various theoretical aspects of this integration is beyond the scope of the present article, and word limits preclude further theoretical discussion, it is important to note that our studies include tests of the underlying motivation or intent for both aggressive and prosocial behavior. Furthermore, we believe that the Tangram Task will aid researchers interested in testing and refining these recent theoretical advances.

¹University of Michigan, Ann Arbor, USA

²Iowa State University, Ames, USA

³Gettysburg College, PA, USA

Corresponding Author:

Muniba Saleem, Department of Communications & Research Center for Group Dynamics, Institute for Social Research, University of Michigan, 5373 North Quad, 105 South State Street, Ann Arbor, MI 48109, USA.
Email: saleemm@umich.edu

Assessing Aggressive and Prosocial Behavior

Aggressive Behavior

Myriad techniques exist to measure aggressive behavior. Popular methods include self-report measures of behavioral tendencies (e.g., Buss & Perry, 1992), and similar tendency reports by others (peer [Walder, Abelson, Eron, Banta, & Laulicht, 1961], teacher's [Huesmann, Eron, Guerra, & Crawshaw, 1994], and parent's [Achenbach & Edelbrock, 1983]). Although such measures are important to answering important research questions, they may be less suited to reliably measure laboratory-based state aggressive behaviors in certain situations.¹ Indeed, Anderson et al. (2010) argued that one criteria for best practice laboratory experimentation (in video game violence research specifically) includes not using trait-based measures to assess aggressive behavior after an experimental manipulation, because of their lack of sensitivity to short-term changes in goals and intentions.

Several laboratory measures have been used to validly and reliably assess aggressive behavior (see Bushman & Anderson, 1998, for review). Although these measures are not without criticism (Ritter & Eslea, 2005; Tedeschi & Quigley, 1996), a variety of empirical approaches have established their internal and external validity (e.g., Anderson & Bushman, 1997; Carlson, Marcus-Newhall, & Miller, 1989; Giancola & Chermack, 1998; Giancola & Parrott, 2008). Common examples include the Buss Aggression Machine (e.g., Buss, 1961), Competitive Reaction Time Task (e.g., Bushman, 1995; Taylor, Gammon, & Capasso, 1976), hot-sauce allocation (e.g., McGregor et al., 1998), point subtraction aggression paradigm (e.g., Cherek, 1981), ratings of an experimenter that have bearing on re-appointment (e.g., Pedersen, Gonzales, & Miller, 2000; Twenge, Baumeister, Tice, & Stucke, 2001), hand in ice-water (Pedersen et al., 2000), and others. Using such measures, researchers have uncovered and substantiated several important theoretical claims regarding aggression effects. Indeed, research reveals that aggressive behavior is affected by (a) violent media (e.g., Anderson et al., 2010), (b) provocation (e.g., Bettencourt & Miller, 1996), (c) reminders of our own death (e.g., McGregor et al., 1998), (d) presentation of aggressive primes (e.g., Berkowitz & LePage, 1967), and other situational factors (see Anderson & Bushman, 2002, for review).

Prosocial Behavior

Although several trait-based and observational methods exist for assessing prosocial behavior, fewer laboratory-based measures exist. Typical laboratory measures of prosocial behavior include amount of money donated to charity (e.g., van Baaren, Holland, Kawakami, & van Knippenberg, 2004), number of pencils (or other materials) picked up after they spill (e.g., McCrae & Johnston, 1998), monetary transfers in

an economic decision-making game (e.g., Piff, Kraus, Cote, Cheng, & Keltner, 2010), willingness to help a distressed person (e.g., Mikulincer, Shaver, Gillath, & Nitzberg, 2005), and others. Using such measures, research has shown that prosocial behaviors are affected by (a) presentation of pictures of religious symbols (e.g., Shariff & Norenzayan, 2007), (b) prosocial video games (e.g., Greitemeyer & Osswald, 2010), (c) being assisted in a difficult task by another (e.g., Bartlett & DeSteno, 2006), (d) secure attachment primes (e.g., Mikulincer et al., 2005), and others.

Aggressive and Prosocial Behavior

Despite the use of the aforementioned measures to assess aggressive behavior *or* prosocial behavior, these measures do not allow researchers to measure both aggressive *and* prosocial behavior simultaneously, which may be theoretically important in certain contexts. Indeed, researchers cannot equate a lack of aggressive (or prosocial) behavior using any of the aforementioned tasks as a prosocial (or aggressive) response. Although helpful and hurtful behaviors are conceptually distinct, they often are inversely related—especially in short-term real-world contexts. For example, when people engage in a hurtful behavior toward a target person, they seldom simultaneously engage in helpful behavior toward that same target. Of course, specific types of aggressive behavior (e.g., instrumental aggression) may involve harming another with an overarching prosocial goal. Consider, for example, an individual pushing an attacker away from a loved one. In this example, the act is consistent with our understanding of aggression (against the attacker), but the overall goal is likely primarily prosocial. Therefore, it may be important to measure aggressive and prosocial behaviors simultaneously to get more insights into such complex social behaviors. However, there is a paucity of research validating behavioral assessments of both types of social behavior simultaneously. Indeed, we are aware of only two laboratory-based tasks designed to measure both types of behavior. Each is discussed in detail below.

The first measure is called the Help/Hurt button (Liebert & Baron, 1972). Originally designed to assess how violent television images influence children's aggressive and prosocial behavior, this measure places children in front of a wooden box with two buttons labeled *Help* and *Hurt*. Participants are told that the buttons will make a task easier or harder for another ostensible child. For example, Gadow, Sprafkin, and Grayson (1990) told children that another (fictitious) child was steering a remote controlled boat. Pressing the *Help* button would allow the partner to steer the boat with ease, whereas pressing the *Hurt* button would make the steering wheel cold, making the "partner's" hand hurt, and steering the boat very difficult. Results showed a significant positive correlation between *Hurt* scores and trait aggression (assessed by parent ratings) for those exposed to an aggressive cartoon.

The Help/Hurt measure has several clear advantages. First, this measure can assess both helping and hurting simultaneously. Second, the instructions and apparatus are amenable to change. For instance, Liebert and Baron (1972) used a similar measure; however, pressing the Hurt button made it difficult for another child to turn a handle, whereas pressing the Help button made this task easier. Third, the cover story is easily understandable for children, and research has used this measure in both emotionally disturbed and learning disabled children (e.g., Sprafkin & Gadow, 2001). However, several limitations dampen the enthusiasm for using the Help/Hurt button measure. First, this measure has only been validated with children and in lab settings, making it potentially inappropriate to use with college-aged or online samples. Second, the primary dependent measure is the length of time the Help or Hurt button is pressed, rather than the decision to choose helping or hurting behavior. Finally, and perhaps most importantly, there is no third non-aggressive, non-prosocial option. In other words, participants either help or hurt the fictitious partner without substantial alternatives.

Another laboratory-based measure that allows researchers to assess aggression and prosocial behavior is the selection of aversive versus pleasant pictures for another hypothetical participant to view (Mussweiler & Förster, 2000). In this task, participants assign 10 of 30 pictures (10 pleasant, 10 neutral, 10 unpleasant) to another participant to view. The advantage of this task is that it allows for a neutral option. The disadvantage is that it cannot be administered with children as the aversive images are quite unpleasant (e.g., picture of a rotting animal corpse). Furthermore, it is unclear to what extent responses on this task are associated with trait aggression and prosocialness as no published studies have documented this.

To maximize the advantages of many of these paradigms while minimizing the disadvantages, we created a laboratory-based measure of aggressive and prosocial behavior called the Tangram Help/Hurt Task. Participants are told that they will interact with another participant on a puzzle task. Furthermore, participants are instructed that (a) they are to select 11 (out of a possible 30) puzzles of varying difficulty for the ostensible participant to complete, and (b) if the other participant completes 10 of these puzzles in 10 min, then this other participant wins a prize (e.g., gift certificate). The 30 puzzles are pre-classified into easy, medium, and hard, 10 puzzles per difficulty level, based on prior studies. Helpful and hurtful scores are calculated based on the number of easy and hard puzzles assigned to the other participant. A detailed discussion of the advantages of this Tangram Task is presented after our empirical studies of its validity.

Overview of the Current Research

This article reports six studies that tested the convergent and discriminant validity of the Tangram Help/Hurt Task. Studies 1 to 3 used a cross-sectional correlational design to test the

relations between scores on the Tangram Help/Hurt Task and other validated self-report measures of aggressive and prosocial personality in college-based and adult online samples. Study 4 primed participants with empathy (or not), an experimental procedure known to influence prosocial behavior. Studies 5 and 6 experimentally manipulated provocation (or not), a factor known to influence aggression. Additional measures of helpful and hurtful behaviors were included in Studies 4 through 6 for comparison with the Tangram Task.

Studies 1 to 3: Cross-Sectional Tests of the Help/Hurt Tangram Task

Studies 1 to 3 tested the convergent and discriminant validity of the Tangram Help/Hurt Task. Specifically, Study 1 tested the association between trait measures of aggression and prosocialness and tangram choices. Study 2 tested the association between trait empathy, perspective taking, state hostility, prosocialness, aggression, and tangram choices. Study 3 tested the association between tangram choices, trait neuroticism, sensation seeking, agreeableness, and control aggression schemas. Study 3 also provided evidence for discriminant validity by testing the association between tangram choices, social desirability, emotional regulation, achievement motivation, and participants' perceptions of tangram difficulty. These constructs allowed us to test alternative explanations of participants' tangram choices. Motivations for tangram assignment were assessed in all three studies.

Participants

Study 1. A total of 166 participants from a Midwestern University participated for course credit. Four participants indicated they were suspicious about the presence of another participant during the debriefing and were excluded from the analyses, leaving 87 females and 75 males ($M_{\text{age}} = 19.59$ years; $SD = 1.96$).

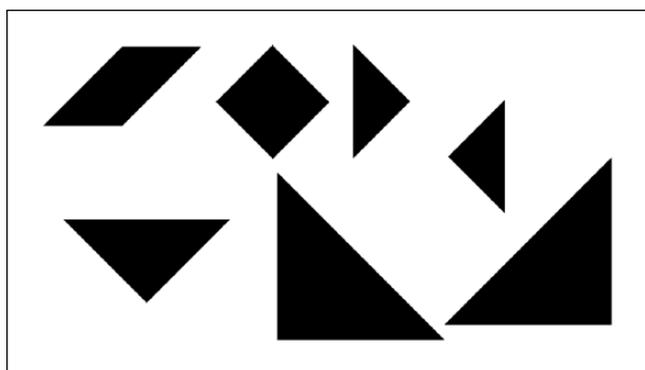
Study 2. A total of 200 participants completed the survey on Amazon MTurk for monetary compensation. Twenty-one participants were unable to view the tangram instruction video due to problems with their video or audio player. Two participants indicated that they did not understand the Tangram Help/Hurt Task after watching the online instructional video. Of the remaining 177 participants, 73 were female, 101 were male, and 3 unidentified ($M_{\text{age}} = 33.40$ years, $SD = 10.84$).

Study 3. A total of 160 participants completed the survey online through Amazon MTurk for monetary compensation. Fourteen participants indicated that they did not believe that the tangrams were assigned to another participant and were excluded from the analysis. Of the remaining 146 participants, 71 were female and 75 were male ($M_{\text{age}} = 36.05$ years; $SD = 12.84$).

Table 1. Means, Standard Deviations, and Alphas for All Measures Used in Studies 1 to 3.

Measures	Study 1, N = 162			Study 2, N = 177			Study 3, N = 146		
	M	SD	α	M	SD	α	M	SD	α
Easy	4.33	2.70		6.69	3.32		6.49	3.32	
Medium	4.10	1.63		2.67	1.85		2.79	1.87	
Hard	2.58	1.88		1.64	2.14		1.73	2.42	
BP physical	2.58	0.96	.81	3.02	1.59	.80			
BP verbal	3.43	1.10	.70	3.88	1.27	.65			
BP anger	2.88	0.88	.76	2.58	1.33	.82			
BP hostility	2.82	1.11	.79	3.46	1.40	.72			
BP composite	2.87	0.75	.89	3.23	1.01	.83			
Prosocialness	3.33	0.65	.80	4.85	1.07	.88			
Empathy				3.67	0.86	.90			
Perspective taking				3.68	0.75	.86			
State hostility				2.15	0.63	.93			
Narcissism							0.23	0.21	.80
Control aggression schema							3.10	0.67	.90
sensation seeking							2.70	0.80	.83
Agreeableness							3.81	0.69	.84
Social desirability							0.46	3.06	.80
Achievement motivation							3.32	0.69	.86
Emotion regulation							3.34	0.49	.66
Tangram Task difficulty							3.02	0.99	.94
Motivation to help	3.43	1.15	.88	3.94	1.12	.88	3.76	1.30	.94
Motivation to hurt	1.75	0.95	.79	1.87	1.06	.91	1.60	0.92	.81

Note. BP = Buss–Perry.

**Figure 1.** Tangram pieces.

Materials

Table 1 lists the descriptive statistics for all measures used in Studies 1 to 3. It also shows which measures were included in each study.

Tangram Help/Hurt Task. Tangrams are based on seven differently shaped plastic pieces (e.g., small square, large triangle) used to form a specified outlined shape. Figure 1 displays the standard seven pieces. Participants first view the outline shape and then use the Tangram pieces to create the target

shape. We selected 30 tangrams and assigned them to easy, medium, and hard difficulty levels, based on the number of pieces required to complete a particular tangram shape and on pilot testing. Easy tangrams can be completed using one to three tangram pieces, medium shapes can be completed using four to six pieces, and hard shapes require all seven pieces.² This tangram assignment table (see the appendix) was used by participants to assign 11 tangrams to the “other participant,” and was supposedly used by the “other participant” to assign 11 puzzles for the participant to complete within the 10-min time limit.

Tangram assignment motivation. Participants indicated their agreement with two statements assessing their motivation to help (e.g., I wanted to help the other participant win the prize) and two assessing motivation to hurt (e.g., I wanted to make it difficult for the other participant to win the prize) on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale.

Measures Assessing Convergent Validity

Trait aggression, state hostility, sensation seeking, narcissism, and control aggression schemas have been associated with aggressive behaviors in the previous literature (Anderson & Bushman, 2002; Joireman, Anderson, & Strathman, 2003; Warburton, 2007). Similarly, trait prosocialness, empathy,

perspective taking, and agreeableness have been associated with prosocial behaviors (Dovidio, Piliavin, Schroeder, & Penner, 2006; Graziano, Habashi, Sheese, & Tobin, 2007). Thus, we expected these trait measures to correlate with helping and hurting scores on the Tangram Task.

Trait aggression. Trait aggression was assessed through the 29-item Buss–Perry Aggression Questionnaire (BPAQ; Buss & Perry, 1992) in Study 1 and the 12-item brief measure (Webster et al., 2014) in Study 2. Participants indicated agreement with statements (e.g., “If somebody hits me, I hit back”) on a 1 (*extremely uncharacteristic of me*) to 5 (*extremely characteristic of me*) scale.

Trait prosocialness. Trait prosocialness was assessed through the 14-item Prosocial Personality Battery (Penner, Fritzsche, Craiger, & Freifeld, 1995) in Study 1 and a 10-item brief measure (Prot et al., 2014) in Study 2. Participants indicated how often they engaged in each of the behaviors specified (e.g., “I have done volunteer work for a charity”) on a 1 (*never*) to 5 (*very often*) scale.

Trait empathy and perspective taking. Trait empathy and perspective taking were assessed using the Interpersonal Reactivity Index (Davis, 1983). Participants indicated their agreement with statements on a 1 (*does not describe me well*) to 5 (*describes me very well*) rating scale. Examples of perspective-taking items included, “I sometimes try to understand my friends better by imagining how things look from their perspective,” and “I believe that there are two sides to every question and try to look at them both.” Examples of empathy items included, “I often have tender, concerned feelings for people less fortunate than me,” and “Sometimes I don’t feel very sorry for other people when they are having problems” (reverse-scored).

State hostility. We used a shortened version of the State Hostility Scale (Anderson, Deuser, & DeNeve, 1995). Participants indicated the extent to which they currently feel various antisocial (12-items) and prosocial (10-items) emotions (e.g., “I feel furious,” “I feel happy”) using a 1 (*strongly disagree*) to 5 (*strongly agree*) scale.³

Sensation seeking. Participants indicated their agreement with eight statements (e.g., “I would like to try bungee jumping”) on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale (Hoyle, Stephenson, Palmgreen, Lorch, & Donohew, 2002).

Narcissism. Participants read 16 pairs of statements and selected the option that best represents them within each pair (e.g., “I like to be the center of attention”). Narcissistic options were coded as 1 and non-narcissistic options were coded as 0 (Ames, Rose, & Anderson, 2006).

Agreeableness. Agreeableness was assessed with nine statements (e.g., “I see myself as someone who likes to cooperate

with others”) on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale (John & Srivastava, 1999).

Control aggression schemas. Control aggression schemas were assessed using Warburton’s (2007) scale. Participants indicated their agreement with 35 statements (e.g., “The world belongs to those who can dominate others”) on a 1 (*completely untrue*) to 6 (*completely true*) scale. Three items negatively correlated with other items and were removed.

Measures Assessing Discriminant Validity

It is possible that tangram choices are influenced by participants’ perception of Tangram Task difficulty, social desirability concerns, achievement motivation, and/or emotion regulation.

Perception of Tangram Task difficulty. Four statements assessed the extent to which participants’ perceived the Tangram Task to be difficult (e.g., “The Tangram Task seems hard”) on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale in Study 3.

Social desirability. Participants indicated whether each of the 11 statements (e.g., “I’m always willing to admit it when I make a mistake”) was true/false in describing them (Reynolds, 1982). Socially desirable responses were coded as 1 and other responses were coded as 0.

Achievement motivation. Participants rated their agreement with 10 statements (e.g., “I am appealed by situations allowing me to test my abilities”) on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale (Lang & Fries, 2006).

Emotional regulation. Participants rated their agreement with 10 statements (e.g., “I control my emotions by not expressing them”) on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale (Gross & John, 2003).

Overall Procedure

All participants completed an informed consent and were told that they would be completing a puzzle task with another participant. Participants received standardized tangram instructions in person in Study 1 and online through a video in Studies 2 to 3. Participants practiced solving tangrams with a practice packet in Study 1 and saw examples of tangrams being solved by an experimenter through a video in Studies 2 and 3. Next, participants answered trait measures specific to each study (see Table 1). Next, they chose 11 tangrams to assign the other participant either by circling with a pen/pencil (Study 1) or clicking (Studies 2 and 3) directly on the tangram assignment table (see the appendix). Then, participants completed questions assessing their motivation for tangram assignment and demographic information. Next, participants were asked questions about the study using

open-ended questions designed to assess suspicion. Finally, participants were debriefed, thanked, and dismissed.

Results

Preliminary analyses. We propose two ways of scoring the Tangram Task depending on the researcher's goals: (a) separate helping and hurting scores, and (b) an overall tangram difference score. Importantly, all studies included analyses using both scoring techniques.

Separate helping and hurting scores. Researchers interested in assessing helpful and hurtful behaviors can compute two separate, but inversely related, scores using the Tangram Task. Because the task includes 10 puzzles per difficulty level and required 11 choices, participants have to pick from at least two categories. It is possible for someone to pick 10 medium tangrams and 1 easy (or hard) tangram to complete the 11 required. However, this individual is not necessarily intending to help (or harm) the other participant, because the other participant needed to complete only 10 tangrams to win the gift certificate. Thus, "helping" was operationally defined as the number of easy puzzles greater than 1. Similarly, "hurting" was defined as the number of hard puzzles greater than 1.⁴

Note that although helping and hurting scores are inversely related using this scoring method; they are not necessarily incompatible with each other. For example, an individual may decide to assign two easy, seven medium, and two hard puzzles in which case their helping and hurting scores would be 1 and 1, respectively. Another individual may decide to assign two easy, two medium, and seven hard puzzles in which case their helping and hurting scores would be 1 and 6, respectively. Although both of these individuals have the same helping score, the latter is clearly intending to hurt the other participant's chances of winning by assigning them a greater number of harder puzzles, whereas, the former is not necessarily intending to hurt or help the other participant.

Difference score. Researchers interested in assessing to what extent an individual is more likely to be helpful versus hurtful can compute a difference score using the Tangram Task. The difference score was computed by subtracting the hurting score from the helping score (defined earlier). Thus, higher numbers represent a tendency to be helpful whereas lower numbers represent a tendency to be hurtful.

Studies 1 to 3 General Findings

Table 2 reports the zero-order correlations for all measures. Several general findings emerge (see Table 2). First, the helping, hurting, and the overall tangram scores were (necessarily) highly intercorrelated. Second, participant sex was unrelated to tangram scores. This obviates the need to control for sex in other more substantive analyses. Third, motivations

to help and hurt the "other participant" were strongly and consistently related to the tangram scores in exactly the expected pattern. Fourth, among the Buss–Perry trait aggression subscales, the tangram scores were consistently associated with physical aggression, anger, and hostility, and were least associated with verbal aggression. Fifth, the tangram scores correlated with other individual difference variables with which they should correlate (convergent validity) and yielded weak and/or essentially zero correlations with variables that one would hope that they would not be related (discriminant validity). Specifically, trait aggressiveness was consistently and modestly positively correlated with the hurting scores, across all three relevant samples, $r_s = .27$ to $.30$, using the Buss–Perry composite score. Helping score consistently correlated positively with trait prosocialness, $r_s = .18$ to $.24$. Similarly, both empathy and perspective taking correlated positively with helping and negatively with hurting.

Studies 1 to 3 Specific Findings

Beyond these general findings, there were several important distinctions. First, Study 1 used a college sample, whereas Studies 2 and 3 used an online MTurk sample. Second, though helping was positively associated with prosocialness in Studies 1 and 2, hurting was significantly negatively associated with prosocialness only in Study 2. Whether this is the result of differences in the prosocialness measures or the samples is unclear, but also is not terribly important for present purposes. Third, state hostility (Study 2) was negatively associated with helping and positively associated with hurting scores. Fourth, Study 3 yielded very small nonsignificant associations between the tangram scores and social desirability, achievement motivation, emotion regulation, and perception of Tangram Task difficulty, thereby providing evidence of discriminant validity. Study 3 also yielded theoretically expected correlations between tangram scores and narcissism, control aggression schema, agreeableness, and sensation seeking.

Regression analyses were conducted in Studies 1 to 2 to test the effects of trait aggression (prosocialness) on hurting (helping) after controlling for helping (hurting). In Study 1, trait aggression was significantly related to hurting even after controlling for helping, $t(1, 158) = 2.91$, $p < .01$, $b = 0.27$, 95% confidence interval, $CI = [0.09, 0.45]$. Similarly, trait prosocial behavior significantly predicted helping while controlling for hurting, $t(1, 158) = 4.39$, $p < .001$, $b = 0.60$, 95% $CI = [0.33, 0.87]$. In Study 2, trait aggression (prosocialness) was positively associated with hurting, $b = 0.11$ (helping, $b = 0.21$) while controlling for hurting (helping) but not significantly.

Summary of the Correlational Studies

In sum, the three correlational studies provide strong convergent and discriminantly validity evidence for the Tangram

Table 2. Zero-Order Correlations of Tangram Scores With Other Measures Used in Studies 1 to 3.

Zero-order <i>r</i> s	Study 1, <i>N</i> = 162			Study 2, <i>N</i> = 177			Study 3, <i>N</i> = 146		
	Helping	Hurting	Help-hurt	Helping	Hurting	Help-hurt	Helping	Hurting	Help-hurt
Helping									
Hurting	-.73***			-.77***			-.73***		
Help-hurt	.96***	-.89***		.97***	-.90***		.96***	-.90***	
BP physical	-.15 [†]	.25**	-.21*	-.22**	.22**	-.23**			
BP verbal	-.14	.12	-.14	-.11	.13	-.12			
BP anger	-.20*	.30***	-.26**	-.27**	.27**	-.28**			
BP hostility	-.18*	.22**	-.21*	-.19*	.17*	-.19*			
BP composite	-.22*	.30***	-.27**	-.27**	.27**	-.29**			
Prosocialness	.18*	.02	.10	.24**	-.22**	.24**			
Empathy				.17*	-.22**	.20*			
Perspective taking				.24**	-.25**	.26**			
State hostility				-.14 [†]	.16*	-.16*			
Narcissism							-.26**	.32***	-.30***
Control aggression schema							-.20*	.18*	-.21***
Sensation seeking							-.09	.19*	-.14*
Agreeableness							.25**	-.25**	.27**
Social desirability							.11	-.02	.08
Achievement motivation							-.03	.05	-.04
Emotion regulation							.05	-.02	.04
Tangram Task difficulty							.09	-.02	.07
Motivation to help	.64***	-.60***	.67***	.83***	-.77***	.85***	.80***	-.74***	.83***
Motivation to hurt	-.50***	.46***	-.52***	-.76***	.74***	-.80***	-.65***	.69***	-.72***
Sex: 0 = female; 1 = male	.05	-.06	.06	.03	-.02	.03	-.05	.06	-.06
<i>M</i> s	3.47	1.70	1.78	5.72	1.13	4.59	5.52	1.20	4.32
<i>SD</i> s	2.66	1.68	4.06	3.26	1.79	4.75	3.26	2.12	5.03

Note. BP = Buss-Perry.

[†]*p* = .06. **p* < .05. ***p* < .01. ****p* < .001.

Task. These results suggest that the Tangram Task is a valid assessment of both, prosocial and aggressive behaviors.

Study 4: Empathy Prime Effect on the Tangram Task

Study 4 tested the effects of a validated empathy manipulation on the Tangram Task. Pre-experimental measures of trait aggression and trait prosocialness were assessed. In addition, we explored whether the effect of the empathy-inducing prime on tangram choices is influenced by participant mood by including an assessment of state hostility. Finally, to compare the effects of the experimental manipulation on the Tangram Task as well as an established measure of helpfulness, participants made a recommendation for additional monetary compensation for the other participant.

Method

Participants. A total of 312 participants completed the study online through Amazon MTurk. Forty participants indicated

that they did not believe that there was another participant in this experiment.⁵ Suspicion rate did not vary significantly by condition, *p* > .20. These participants were excluded from the main analyses. Of the remaining 272 participants, 154 were female, 113 were male, and 5 unidentified (*M*_{age} = 35.58 years, *SD* = 12.21).

Materials. Helping and hurting scores were assessed by the Tangram Help/Hurt Task used in previous studies. In addition, we included measures of trait prosocial behavior from Study 1, tangram assignment motivations, trait aggression, state hostility, and demographics from Study 2.

Priming manipulation. Participants read an essay written by another ostensible participant describing something important that happened to them recently. Participants in the empathy-inducing condition read about a same-sex participant describing their recent break-up from a serious relationship. Participants in the neutral condition read about a same-sex participant describing their recent weekend in which they hung out with friends. The essays were of approximately the same length. This manipulation has

Table 3. Means, Standard Deviations, and Alphas for All Measures Used in Studies 4 to 6.

Measures	Study 4, N = 272			Study 5, N = 152			Study 6, N = 99		
	M	SD	α	M	SD	α	M	SD	α
Trait aggression	3.01	1.06	.85	2.05	0.72	.82	2.91	0.91	.81
Prosocialness	2.98	0.66	.86				5.78	0.73	.73
State hostility	2.02	0.58	.92						
Extraversion							4.77	1.63	.84
Agreeableness							4.95	1.09	.37
Conscientiousness							5.76	0.99	.36
Emotional stability							4.58	1.23	.53
Exploration							5.33	1.11	.32
Easy	7.33	2.99	—	3.06	2.50	—	4.28	3.37	—
Medium	2.46	1.70	—	4.21	1.64	—	3.57	1.96	—
Hard	1.21	2.01	—	3.73	2.52	—	3.15	2.65	—
Monetary reward	5.08	3.79	—				8.64	6.10	—
Positive evaluation				5.42	1.38	.84	4.83	0.97	.88
Weight (g)							9.14	6.61	—
Hot-sauce \times Weight							19.80	21.81	—
Motivation to help	4.16	1.07	.90	2.89	1.24	—	3.49	1.06	.91
Motivation to hurt	1.65	1.03	.89	2.58	1.25	—	1.96	0.87	.78

been successfully used in the past to prime prosocial behavior (Batson, Klein, Highberger, & Shaw, 1995; DeWall & Baumeister, 2006).

Monetary reward for the other participant. As an additional assessment of helpful behavior, we asked participants how much to pay the other participant *in addition* to the amount specified for this research study. Participants moved a slider scale that ranged from \$0 to \$10.00.

Procedure. After completing an online informed consent, participants were told that the goal of the study is to understand how people interact with each other based on writing styles. Participants received standardized tangram instructions through a video, and indicated whether they understood the task. Next, participants wrote a brief essay about something important that happened to them recently and were told that they would read a similar essay written by “the other participant.” Participants then completed demographic, trait aggression, and trait prosocial behavior items. Next, participants were randomly assigned to read either an empathy-inducing or neutral essay from “the other participant.” After reading the essay, participants assigned 11 tangrams to that “other” participant. Then, participants completed the post-experimental questionnaires assessing their motivation for tangram assignment, state hostility, and recommendation for additional monetary compensation for the other participant. Finally, participants answered an open-ended question about any thoughts they had about the other participant, were debriefed, and thanked.

Results

Preliminary analyses. Means, standard deviations, and correlations among the six post-experimental measures are provided in Tables 3 and 4. As seen in Table 4, the usual pattern of correlations occurred for the three tangram and two motivation measures. Furthermore, monetary reward correlated positively with helping behavior and motivation, and negatively with hurting motivation.

For each pre-experimental measure and demographic variable, we tested the main effect and two-way interaction with condition in separate ANCOVAs for helping, hurting, and the difference score. Trait aggression, participant age, and sex yielded significant effects and thus were included as covariates in the main analyses; all other measures were dropped.⁶

A one-way (condition: empathy-inducing/neutral) ANOVA revealed that condition did not significantly influence state hostility, $F < 1.00$, $p > .10$, and thus it was dropped from the main analyses.

Main analyses

Difference score. A one-way (condition: empathy-inducing/neutral) ANCOVA with trait aggression, participant sex, and age as covariates revealed a significant effect of condition, $F(1, 262) = 11.57$, $p < .01$, $d = 0.42$, $\eta_p^2 = .04$ (see Figure 2). Participants who read the empathy-inducing essay were more helpful–less hurtful, $M = 6.59$, 95% CI = [5.87, 7.31], compared with participants who read the neutral essay, $M = 4.85$, 95% CI = [4.13, 5.56]. In addition, participant sex yielded a significant effect, $F(1, 262) = 5.90$,

Table 4. Zero-Order Correlations Between Tangram Scores and Other Outcome Measures Used in Studies 4 to 6.

Zero-order <i>r</i> s	Study 4, N = 272			Study 5, N = 152			Study 6, N = 99		
	Helping	Hurting	Help-hurt	Helping	Hurting	Help-hurt	Helping	Hurting	Help-hurt
Helping									
Hurting	.74***			-.69***			-.73***		
Help-hurt	.96***	-.89***		.92***	.92***		.95***	-.91***	
Monetary reward	.21***	-.09	.18***				.50***	-.38***	.48***
Positive evaluation				.31**	-.50***	.44***	.34**	-.37***	.38***
Weight (g)							-.29**	.28**	-.31**
Hot-sauce × Weight							-.24*	.28*	-.27*
Motivation to help	.71***	-.68***	.75***	.52***	-.51***	.56***	.82***	-.74***	.84***
Motivation to hurt	-.70***	.73***	-.76***	-.54***	.65***	-.64***	-.70***	.75***	-.77***
<i>M</i>	6.35	0.78	5.58	2.20	2.83	-0.63	3.36	2.36	1.00
<i>SD</i>	2.95	1.71	4.36	2.34	2.38	4.35	3.27	2.42	5.31

p* < .01. *p* < .001.

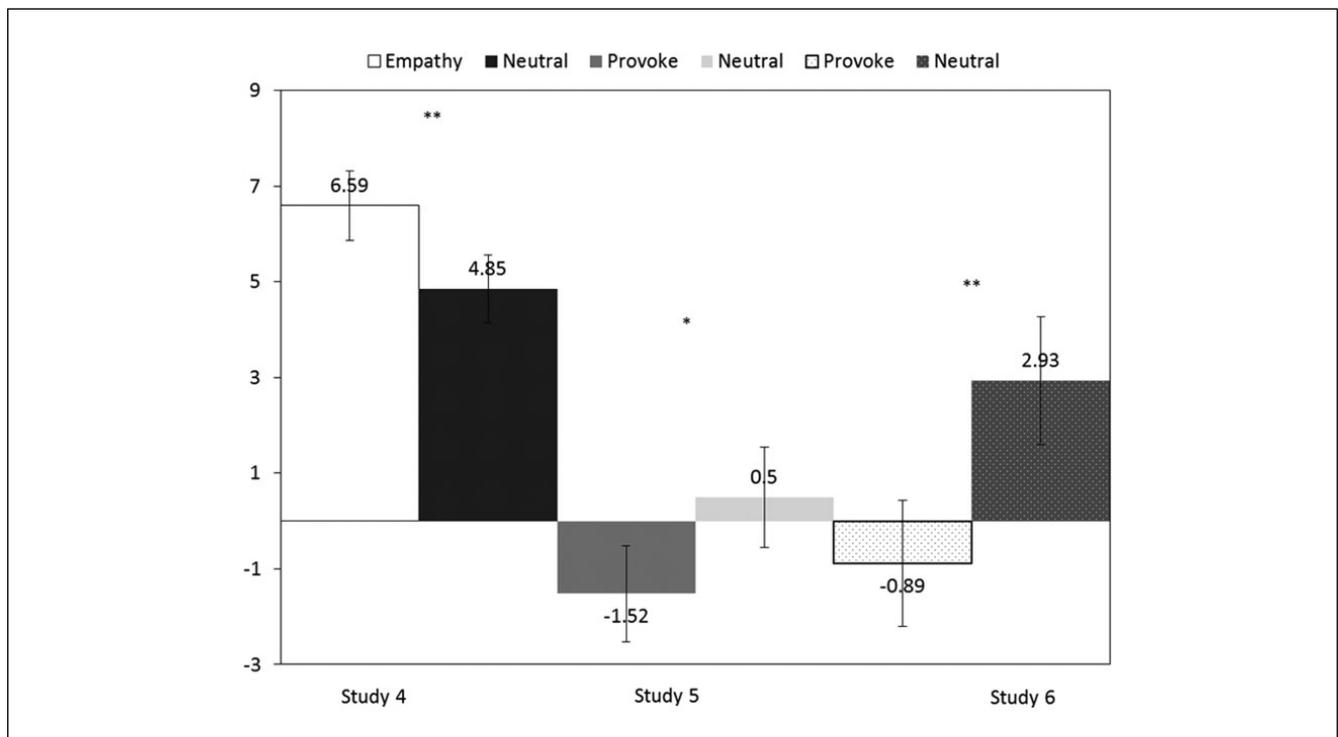


Figure 2. Difference score as a function of experimental manipulations in Studies 4 to 6.

Note. Bars indicate 95% CIs. CI = confidence interval.

p* < .05. *p* < .01.

p < .05, *d* = 0.30, $\eta_p^2 = .02$. Females scored higher than males, *M*s = 6.36, 5.08, respectively. Older participants scored higher than younger ones, $F(1, 262) = 6.71, p < .05, b = 0.68, d = 0.32, \eta_p^2 = 0.03$. Trait aggression was nonsignificant in these analyses, $F(1, 262) = 2.82, p > .05, b = -0.45$.

Separate helping and hurting scores. As shown in Figure 3, participants in the empathy condition were more helpful

than participants in the neutral condition (*M*s = 7.03 and 5.88, 95% CIs = [6.54, 7.51] and [5.40, 6.37], respectively), $F(1, 262) = 10.98, p < .01, d = 0.41, \eta_p^2 = .04$. Conversely, participants in the neutral condition were significantly more hurtful than participants in the empathy condition (*M*s = 1.03 and 0.44, 95% CIs = [0.75, 1.32] and [0.15, 0.72], respectively), $F(1, 262) = 8.51, p < .01, d = 0.36, \eta_p^2 = .03$.

In addition, participant sex influenced helping, $F(1, 262) = 5.67, p < .05, d = 0.29, \eta_p^2 = .02$, and hurting, $F(1, 262) = 4.24,$

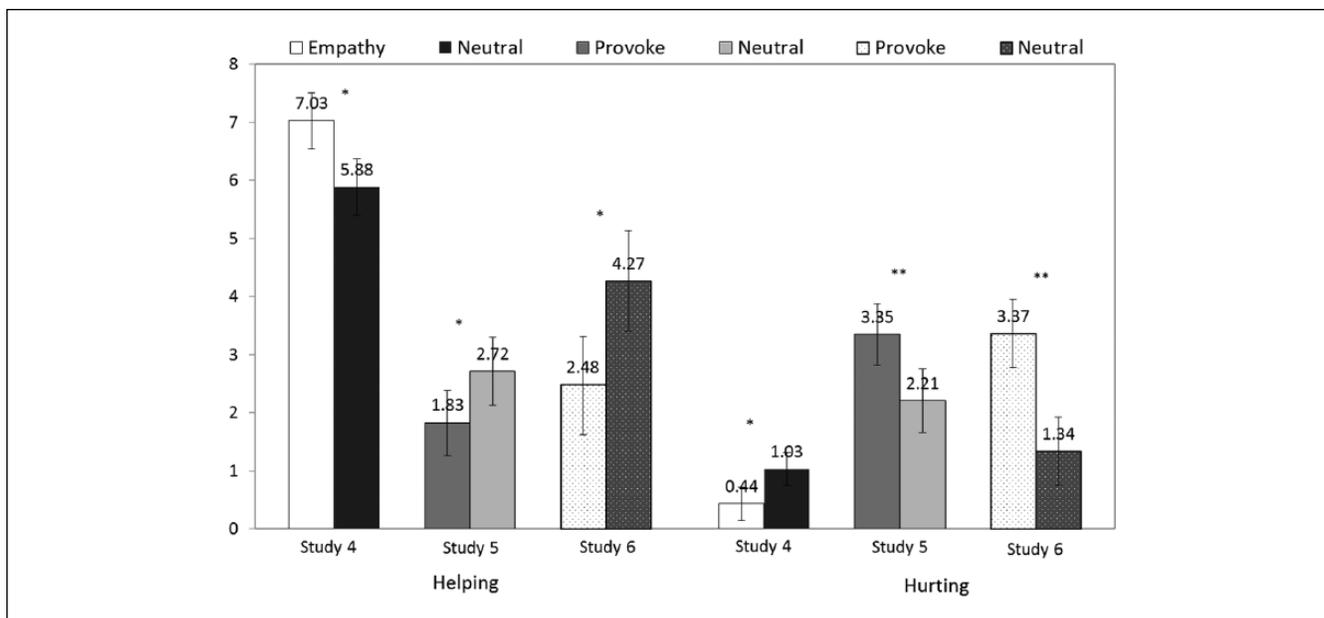


Figure 3. Helpful and hurtful behavior as a function of experimental manipulations in Studies 4 to 6.

Note. Bars indicate 95% CIs. CI = confidence interval.

* $p < .05$. ** $p < .01$.

$p < .05$, $d = 0.25$, $\eta_p^2 = .02$. Females, relative to males, were more helpful ($M_s = 6.88$ and 6.03 , respectively) and less hurtful ($M_s = 0.52$ and 0.95 , respectively). Finally, participant age was positively associated with helping, $F(1, 262) = 6.67$, $p < .05$, $b = 0.46$, and negatively associated with hurting, $F(1, 262) = 4.50$, $p < .05$, $b = -0.22$. Trait aggression was positively associated with hurting ($b = 0.19$) and negatively associated with helping ($b = -0.26$), but not significantly, $F_s < 3.5$, $p_s > .05$.

Monetary reward recommendation. A one-way (condition: empathy-inducing/neutral) ANOVA revealed a significant effect of condition, $F(1, 270) = 11.42$, $p < .01$, $d = 0.41$, $\eta_p^2 = .04$. The empathy-inducing essay yielded a higher monetary reward for the other participant ($M = 5.84$) than did the neutral essay ($M = 4.31$). Furthermore, helpful scores and the help–hurt difference scores on the Tangram Task were significantly positively associated with monetary rewards for the other participant.⁷

In sum, an empathy-inducing prime known to increase prosocial behavior increased helping as assessed by the Tangram Task. Importantly, these effects were found using both scoring methods. These effects cannot be attributed to changes in mood as the prime did not influence state hostility.

Study 5: Provocation Effect on the Tangram Task

Study 5 tested whether a provocation manipulation that is known to increase aggressive behavior would influence

hurting behavior as assessed by the Tangram Task. Participants also evaluated the other participant on several theoretically relevant dimensions.

Method

Participants. One hundred fifty-two students from a large Midwestern University participated in this experiment for course credit. Fifteen were rated as suspicious based on debriefing questions, so their data were deleted. Suspicion did not vary significantly by condition, $p > .20$. Of the remaining 137 participants, 59 were male, 72 female, and 6 unidentified. The mean age was 19.60 ($SD = 2.15$) years.

Materials. Helping and hurting were assessed through the Tangram Help/Hurt Task. Demographics were assessed using the same items as in previous studies. Due to time constraints, only the physical aggression subscale of the BPAQ scale was measured pre-experimentally. Helpful and hurtful intention/motivation for tangram assignment were assessed using one item each (e.g., “I wanted to help [make it difficult for] the other participant to win the prize”).

Provocation. Participants wrote an essay on the topic of abortion. They were led to believe that their essay would be evaluated by another participant through written feedback. Similarly, they were told that they would evaluate the other participant’s essay and give them feedback. Participants in the provocation condition received negative feedback on all dimensions, whereas those in the neutral condition received

average feedback on all dimensions. This manipulation is known to increase aggression (Bushman & Baumeister, 1998).

Evaluation. Participants rated their agreement with 11 statements about the other participant (e.g., “The other participant is competent,” “The other participant is fair and reasonable”) using a 1 (*strongly disagree*) to 10 (*strongly agree*) scale. Similar items have been used in previous research (Pedersen et al., 2000; Twenge et al., 2001).

Procedure. After consenting, participants were told that we are interested in understanding the relationship between participants’ writing style, impressions of others, and performance on a puzzle task. Participants received standard tangram instructions and then completed the physical trait aggression scale. Next, participants wrote an essay either supporting or opposing abortion (their choice). When participants finished writing, they were given the essay ostensibly written by the other participant and were asked to evaluate it. Next, they received the other participants’ supposed essay evaluation—the experimental manipulation of provocation. Next, they chose 11 tangrams to assign the other participant, and then evaluated that person on 11 dimensions. Then, participants completed demographics items and were probed for suspicion before being thanked and fully debriefed.

Results

Preliminary analyses. Means and standard deviations for the tangram scores are provided in Tables 3 and 4. As displayed in Table 4, all three tangram scores were significantly related to motivation items and to evaluations of the other participant. The same strategy as Study 4 was used to determine which covariates to include in the main analyses. Trait aggression yielded significant effects on helping, hurting, and the difference score and thus was included in the main analyses.

Main analyses

Difference score. A one-way (condition: provocation/neutral) ANCOVA with physical aggression as a covariate revealed a significant effect of condition, $F(1, 127) = 7.72$, $p < .05$, $d = 0.49$, $\eta_p^2 = .06$ (see Figure 2). Provoked participants were less helpful/more hurtful compared with participants in the neutral condition, $M_s = 0.50, -1.52$, 95% CIs = $[-0.54, 1.55]$ and $[-2.51, -0.52]$, respectively. In addition, trait physical aggression was negatively associated with a tendency to help over hurt the other participant, $F(1, 127) = 7.77$, $p < .05$, $b = -1.01$; $d = 0.49$, $\eta_p^2 = .06$.

Separate helping and hurting scores. Provoked participants scored higher on hurting than participants in the neutral condition ($M_s = 3.35$ and 2.21 , 95% CIs = $[2.82, 3.88]$ and $[1.66, 2.76]$, respectively), $F(1, 127) = 8.73$, $p < .01$, $d = 0.52$,

$\eta_p^2 = .06$ (see Figure 3). Conversely, participants in the neutral condition scored higher on helping than participants in the provocation condition ($M_s = 2.72$ and 1.83 , 95% CIs = $[2.13, 3.30]$ and $[1.27, 2.39]$, respectively), $F(1, 127) = 4.71$, $p < .05$, $d = 0.39$, $\eta_p^2 = .04$. In addition, trait aggression was positively related to hurting, $F(1, 127) = 12.76$, $p < .01$, $b = 0.69$, and negatively (but nonsignificantly) related to helping, $F(1, 127) = 2.56$, $p = .11$, $b = -0.33$.

Evaluation of the other participant. As expected, participants in the neutral, relative to provoked, condition gave the other participant more positive evaluations, $M_s = 5.99; 4.89$, $F(1, 129) = 24.48$, $p < .001$, $d = 0.87$, $\eta_p^2 = .16$. Furthermore, as shown in Table 4, evaluation was positively associated with helpful and negatively associated with hurtful scores on the Tangram Task. Thus, provocation affected tangram performance and evaluation in the same way.

Comparison of hurting indices. To test whether the experimental manipulation influenced the hurting scores from the Tangram Task differently than the evaluation for the other participant, a 2 (condition: provocation/neutral) \times 2 (outcome: hurting score from the Tangram Task/evaluation) mixed ANOVA was conducted with outcome as the within-subjects factor. In these analyses, we reverse-scored the positive evaluation items, so higher scores represent negative evaluations of the other participant. In addition, both outcomes were standardized. As expected, the condition effect was significant, $F(1, 129) = 19.51$, $p < .01$. In addition, the Condition \times Outcome interaction was significant, $F(1, 129) = 4.49$, $p < .05$. Simple effects analyses revealed that the condition effect on negative evaluations was significantly stronger than on hurting scores assessed through the Tangram Task, $F_s(1, 129) = 24.58, 6.90$, $p_s < .05$, $d_s = .87; .46$, $\eta_p^2 = .16; .05$. We suspect that this may be because the Tangram Task represents a behavior and the evaluation of the other participant represents an attitude, which is generally more sensitive to priming manipulations.

Study 6: Provocation Effect on the Hot-Sauce and Tangram Tasks

Study 6 tested and compared the effects of a provocation manipulation on the Tangram, other participant evaluation, and Hot-Sauce Tasks. Pre-experimental measures of trait aggression, trait prosocialness, and the 10-item personality inventory were included to provide additional convergent and discriminant validity evidence for the Tangram Task.

Method

Participants. A total of 104 students from a large Midwestern University participated in this experiment for course credit. Five were rated as suspicious based on debriefing questions;

their data were deleted. Suspicion did not vary significantly by condition, $p > .10$. Of the remaining 99 participants, 9 were male, 90 female.⁸ The mean age was 19.10 ($SD = 0.61$) years.

Materials. Table 2 includes the means, standard deviations, and alphas for all measures.

Pre-experimental measures. Trait aggression, trait prosocialness, and demographics were assessed as in Study 3. Personality was assessed using the Ten-Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003). The TIPI contains two items designed to assess each of the Big Five personality traits (i.e., extraversion, agreeableness, conscientiousness, emotional stability, and exploration).

Experimental manipulation

Provocation. The provocation method was the same as in Study 5, except the essay topic was affirmative action instead of abortion.

Outcome measures

Tangram Task. Helping and hurting scores were assessed through the Tangram Help/Hurt Task used in previous studies.

Hot-Sauce Task. The Hot-Sauce Task was administered based on the procedure described by McGregor et al. (1998). We assessed two indices, the weight in grams of the hot-sauce assigned to the other participant (McGregor et al., 1998), and the product of the hotness of the sauce selected and the weight in grams (C. Barlett, Branch, Rodeheffer, & Harris, 2009).

Evaluation and monetary reward. Participants rated their agreement with statements describing the other participant as intelligent, skillful, competent, kind, and warm on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale (Pedersen et al., 2000; Twenge et al., 2001). Additional items assessed the extent to which they liked the other participant and would want to work with them again. Finally, participants were told that at times, we reward research participants with a monetary reward for their outstanding performance. Participants made a recommendation of how much money the other participant should get in addition to the credit they were receiving on a continuous scale (\$0-\$20.00).

Procedure. After completing informed consent, participants were told that we were interested in understanding how people interact with someone they do not know. Participants were told that they will complete writing, puzzle, and food tasks. Regarding the food task, participants were told that today all participants are randomly assigned to taste a food sample from the dry or spicy category. Participants were told

that both participants will assign a food sample from one of these categories to the other participant and will be informed of the other participants' food preferences. Next, they rated their food preferences. After that, participants received standard Tangram Task instructions and completed pre-experimental measures. Next, participants wrote an essay either supporting or opposing affirmative action (their choice). After submitting their essay, participants were told that the other participant has not yet completed their essay and thus were asked to answer demographic questions meanwhile. Next, they were given the essay ostensibly written by the other participant and were asked to evaluate it. Next, they received the other participants' supposed essay evaluation—the experimental manipulation of provocation. Next, they completed the Tangram and Hot-Sauce Tasks. The order of these two tasks was counterbalanced across participants. Next, they tasted the food sample assigned to them, evaluated the food sample, and evaluated the other participant on seven dimensions. Finally, participants were asked questions assessing their suspicion of the experiment, thanked, and fully debriefed.

Results

Preliminary analyses. Tables 3 and 4 present the means, standard deviations, and correlations among the outcome measures. As can be seen, the usual pattern of correlations occurred for the three tangram and two motivation measures. In addition, the weight of hot-sauce, the product of hot-sauce number and weight, the evaluation, and the monetary reward scores correlated with Tangram Task scores and the motivation measures as they theoretically should, providing further convergent validity evidence.

The same strategy as in Studies 4 to 6 was used to determine which covariates to include in the main analyses. Only trait prosocialness yielded significant effects, so others were dropped from main analyses.

Main analyses

Difference score. A one-way (condition: provocation/neutral) ANCOVA with trait prosocialness as a covariate revealed a significant effect of essay condition, $F(1, 96) = 16.35, p < .01, d = 0.83, \eta_p^2 = .15$ (see Figure 2). Provoked participants were less helpful/more hurtful compared with participants in the neutral condition, $M_s = -0.89, 2.93$; 95% CIs = $[-2.22, 0.43]$ and $[1.60, 4.27]$, respectively. Trait prosocialness was positively associated with the difference score, $F(1, 96) = 12.56, b = 1.69, p < .01$.

Separate helping and hurting scores. As shown in Figure 3, provoked participants scored higher on hurting than participants in the neutral condition ($M_s = 3.37$ and 1.34 , 95% CIs = $[2.78, 3.95]$ and $[0.75, 1.93]$, respectively), $F(1, 96) = 23.12, p < .001, d = 0.98, \eta_p^2 = .19$. Conversely,

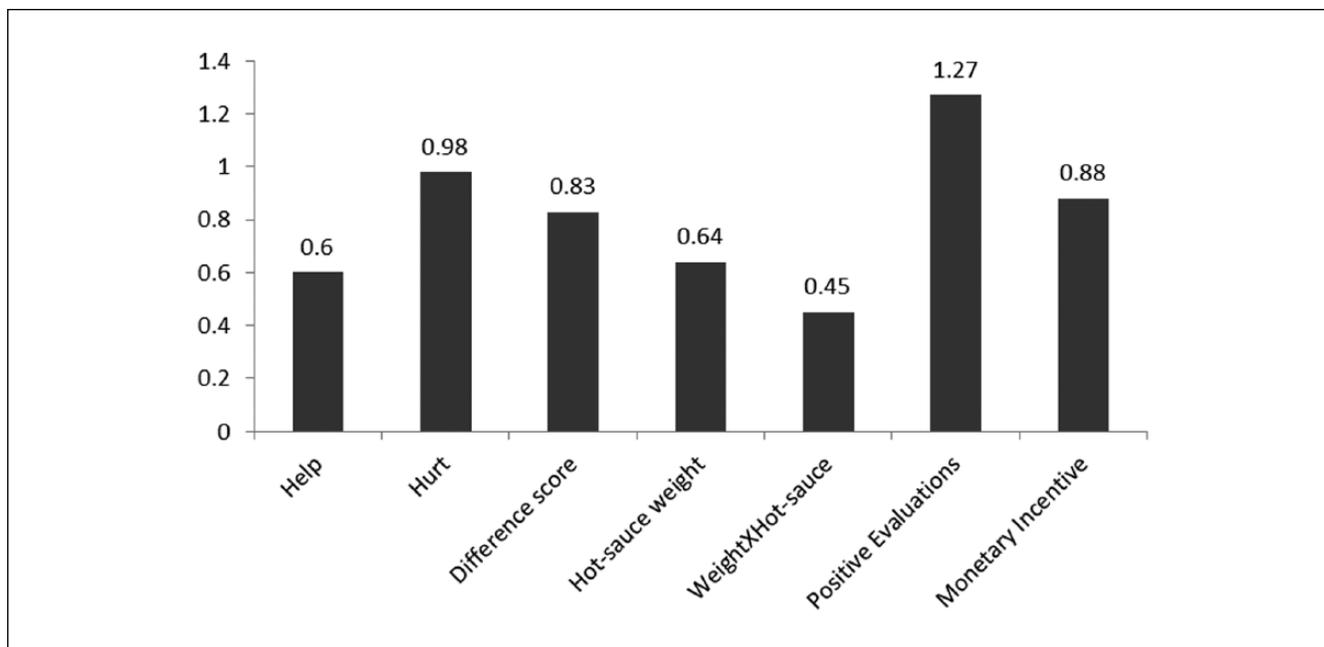


Figure 4. Cohen's *d* effect sizes reflecting the effect of provocation on all outcomes in Study 6.

participants in the neutral condition scored higher on helping than those who were provoked ($M_s = 4.27$ and 2.48 , 95% CIs = $[3.41, 5.13]$ and $[1.62, 3.32]$, respectively), $F(1, 96) = 8.72$, $p < .01$, $d = 0.60$, $\eta_p^2 = .08$. Trait prosocialness was positively associated with helping and negatively associated with hurting, $F_s(1, 96) = 9.86, 10.33$, $b_s = 0.98, -0.69$, $p_s < .01$.

Hot-Sauce Task. Recall that there were two indices, the weight in grams of the hot-sauce assigned to the other participant and the product of the hot-sauce selected and the weight in grams. Two one-way (condition: provocation/neutral) ANCOVAs with trait prosocialness as a covariate revealed a significant effect of condition on both the weight and the Weight \times Hot-sauce product, $F_s(1, 96) = 9.71, 4.84$, $p_s < .05$, $d_s = .64, 0.45$, $\eta_p^2_s = .09, 0.05$. Provoked participants had higher scores on the weight and Weight \times Hot-sauce product indices, $M_s = 11.11; 24.41$, 95% CIs = $[9.33, 12.90]$ and $[18.50, 30.33]$, than participants in the neutral condition, $M_s = 7.13; 15.10$, 95% CIs = $[5.33, 8.93]$ and $[9.34, 12.90]$. Trait prosocialness was negatively associated with Weight \times Hot-sauce product, $F(1, 96) = 4.31$, $b = -4.42$, $p < .05$, but not with weight.

Evaluation. A one-way (condition: provocation/neutral) ANCOVA with trait prosocialness as a covariate revealed a significant effect of condition, $F(1, 95) = 38.21$, $p < .001$, $d = 1.27$, $\eta_p^2 = .29$. Participants in the neutral, relative to provoked, condition gave the other participant more positive evaluations, $M_s = 5.35, 4.32$, 95% CIs = $[5.11, 5.58]$ and $[4.09, 4.55]$, respectively. Trait prosocialness was non-significant.

Monetary reward. A one-way (condition: provocation/neutral) ANCOVA with trait prosocialness as a covariate revealed a significant effect of condition, $F(1, 95) = 18.54$, $p < .001$, $d = 0.88$, $\eta_p^2 = .16$. Participants in the neutral, relative to provoked, condition recommended a higher monetary reward for the other participant, $M_s = 11.05, 6.23$, 95% CIs = $[9.48, 12.63]$ and $[4.66, 7.80]$, respectively. Trait prosocialness was positively associated with recommending a higher monetary reward, $F(1, 95) = 4.32$, $b = 1.16$, $p < .05$.

Correlations across outcomes. As shown in Table 4, the tangram measures correlated appropriately with the monetary reward, evaluation, and hot-sauce measures, providing further validation.

Effect sizes across outcomes. Figure 4 displays the effect sizes of the essay manipulation on the multiple outcomes assessed in this study. Indices of hurtfulness from the Tangram Task (hurt score and difference score) were more sensitive to the provocation manipulation than the indices of hurtfulness from the Hot-Sauce Task (weight in grams and Weight \times Hot-sauce product). However, the positive evaluation and monetary reward measures were more sensitive to the provocation manipulation than the Tangram Task and the hot-sauce indices. We suspect that the latter two measures reflect more accurate assessments of behaviors, whereas, the former two measures represent cognitive and affective constructs that are generally more sensitive to priming methods.

Comparison of hurting indices. We tested the effect of essay condition on aggression as assessed by the Tangram

(hurting score) and Hot-Sauce (weight) tasks. Both outcomes were standardized in this analysis. A 2 (condition: provocation/neutral) \times 2 (task: tangram/hot-sauce) mixed ANOVA was conducted with task as the within-subjects factor. As expected, the condition effect was significant, $F(1, 97) = 24.51, p < .001, d = 1.01, \eta_p^2 = .20$. More importantly, none of the within-subjects effects were significant suggesting that the effect of condition on hurting behavior is not significantly different across these two tasks.

In sum, provocation increased hurting and decreased helping behavior as assessed by the Tangram Task. Once again, this effect was reliable for both scoring methods. The Tangram Task was more sensitive to the provocation manipulation than the Hot-Sauce Task but not as sensitive as the evaluation and monetary reward outcomes. Finally, the correlations between the pre-experimental measures and the helping and hurting scores assessed through the Tangram Task were comparable in size and direction with those observed with other established assessments of helpfulness and hurtfulness.

Discussion

The Present Studies

The primary goal of this research was to introduce and validate the Tangram Task as a measure of helping and hurting behaviors. Studies 1 to 3 provide convergent and discriminant validity evidence for the task using college-aged and adult samples through correlational designs. Across these three studies, helpful and hurtful scores were significantly correlated with established trait assessments of aggression, prosocialness, empathy, perspective taking, state hostility, narcissism, agreeableness, and control aggression schemas. Specific to trait aggression, it is important to note that hurting behavior on the Tangram Task was more strongly related with the physical, anger, and hostile subscales of the BPAQ, relative to the verbal subscale. In addition, tangram choices were not influenced by social desirability, achievement motivation, emotion regulation, sex of participant, or perception of the Tangram Task difficulty.

Experimental evidence from Study 4 validated the Tangram Task for use in studies priming prosocial behavior. Studies 5 and 6 validated the Tangram Task for studies of provocation effects on aggression. Additional indices of helpfulness and hurtfulness included in Studies 4 (monetary reward), 5 (evaluation of other participant), and 6 (Hot-Sauce Task, evaluation of other participant, and monetary reward) significantly correlated with scores on the Tangram Task. The effects of the manipulations on the Tangram Task were comparable and in some cases more sensitive than established indices of helpful and hurtful behaviors (i.e., monetary reward and Hot-Sauce Tasks). Conversely, evaluations of the other participant (in Studies 5 and 6) were more sensitive to the experimental manipulations than the Tangram Task. This

may be due to the former representing an attitude, which is generally more sensitive to priming manipulations, and the latter representing a behavior. Note that the correlations observed between the pre-experimental trait measures and the Tangram Task were also similar in direction and size to those observed for established measures of prosocial and aggressive behaviors (see Supplemental Material analyses for these tables).

Finally, the intention/motivation assessments in all studies provided evidence that the assignment of harder puzzles was motivated by a desire to hurt (and not help) the other participant, whereas, the assignment of easier puzzles was motivated by a desire to help (and not hurt) the other participant. These consistent findings across multiple designs and methods provide further evidence of the key role played by intentions in both prosocial and antisocial domains (e.g., Graziano & Habashi, 2010).

Related Studies

In addition to the present six studies, we know of at least five additional experiments that have used the Tangram Task. In an intergroup context, Saleem et al. (under review: Study 3) found that participants primed with secure attachment chose more easy tangrams for outgroup members than did those in the neutral condition. Another experimental study (C. P. Barlett & Anderson, 2011, Study 1) revealed that the number of hard tangrams selected for another participant was highest for participants who were both (a) previously provoked, and (b) who had not been given any mitigating information concerning the provocation, relative to participants who either did receive mitigating information after the provocation, or who were unprovoked. Thus, in addition to provocation, the Tangram Task is sensitive to re-appraisal manipulations.

Experimental evidence from Gentile et al. (2009, Study 3) found that college-aged participants randomly assigned to play a prosocial video game selected significantly more easy tangram puzzles for the "other" person to solve, relative to participants who played a neutral video game. Conversely, participants in the violent video game condition selected significantly more hard tangrams, relative to those who had played a neutral video game. These video game results were further replicated using the Tangram Task with children 9 to 14 years of age (Saleem, Anderson, & Gentile, 2012).

The Tangram Help/Hurt Task has several advantages over other established measures of helpful and hurtful behaviors. First, the Tangram Task allows simultaneous assessment of helping and hurting behavior. Although these two scores are negatively correlated, by assessing them in the same paradigm at the same time, researchers can use regression techniques to assess the unique and shared variances with correlational and experimental variables, and could even assess changes in helping/hurting choices over time to

examine the roles of automatic and controlled processes (cf. Graziano & Habashi, 2010). Second, 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 a simple overall difference score. For other studies, it may be more appropriate to have separate help and hurt scores. Indeed, in some studies it might make sense to use even stronger adjustments to what constitutes clear evidence of helpful and hurtful intent. For example, counting only the number of hard tangram selections greater than 2 or 3, and assigning zeros to participants who chose 0, 1, or 2 hard tangrams, would further reduce the correlation between the “help” and “hurt” scores and provide evidence for extreme forms of aggression. Third, the Tangram Task includes a medium category, allowing for a response that is neither aggressive nor prosocial response (a common criticism levied against other validated aggression measures; see Ritter & Eslea, 2005). Fourth, tangrams can easily be used with adults and children. One can adjust the difficulty of the task for different populations by selecting a different set of tangrams or by setting different time limits. Fifth, the cover story and various materials used are easily amenable to change, therefore allowing use in a diverse array of studies. Sixth, the Tangram Help/Hurt Task relies on simple count data compatible with paper or computer-based administration, and does not require fancy timing equipment. Indeed, it is easily administered in online studies. Seventh, because the Tangram Task inherently involves puzzle completion and assignment, task instructions should easily translate into other languages allowing this measure to be used cross-culturally. Finally, the Tangram Task could easily be used as a provocation stimulus or a success/failure manipulation by assigning mostly easy/hard puzzles from another ostensible participant.

Limitations

It also is important to note several limitations of the Tangram Help/Hurt Task. First, all the studies reported in this article reveal a strong negative correlation between the helping and hurting scores, when used separately. Individuals who score high on helpfulness by selecting a greater number of easy puzzles will score low on hurtfulness, and vice versa. Indeed, even after using our “greater than 1” scoring procedure, the correlation between the helpful and hurtful scores remained high. This concern can be addressed in several ways: (a) ignoring the medium category for the analyses, thus reducing interdependence; (b) using the number of easy and difficult puzzles greater than 1 instead of raw scores so that participants can obtain a score of 0 on both helpfulness and hurtfulness; (c) entering both helpful and hurtful scores as a within-subjects factor in analyses; (d) using a difference score (helpful score-hurtful score); and (e) using regression procedures to examine the effects of an independent variable of either helping or hurting, while statistically controlling for

the other tangram score. Finally, one can further reduce the correlation between helpful and hurtful scores by setting more extreme rules for what counts as helpful or hurtful behavior.

A second limitation identified by the present studies is that the correlations between relevant trait measures and tangram choices are small to moderate ($r_s = .17-.32$). This is common throughout social psychology, especially when using college student samples that have relatively restricted range, relative to other samples (e.g., Kalmoe, 2015). For online MTurk samples, the online setting may have induced greater suspicion regarding the presence of another participant who was involved in the experiment at the same time as the participant. Two other possible contributors to the small magnitude of the correlations are (a) the fact that general trait measures usually do not predict specific behaviors very strongly, and (b) the reported correlations did not adjust for unreliability of the trait measures, or of the Tangram Task itself. In short, obtained correlations support the convergent validity predictions for this task, and are likely as good as convergent validity correlations for most other brief laboratory style measures of aggressive behavior.

Third, though we established discriminant validity for the Tangram Task using assessments of social desirability, achievement motivation, emotion regulation, and perception of Tangram Task difficulty, other constructs may relate to the Tangram Task. Future research can address these concerns.

Future research also could compare the convergent validity of the Tangram Help/Hurt Task with additional aggressive and prosocial behavioral measures such as competitive reaction time task (Bushman, 1995), prisoner’s dilemma task (Rapoport & Chammah, 1965), and intention to volunteer or donate to charities (Twenge, Baumeister, DeWall, Ciarocco, & Bartels, 2007). In addition, it is important to assess whether experimental manipulations of aggressive and prosocial cognitions influence choices on the Tangram Help/Hurt Task. Finally, the Tangram Help/Hurt Task should be used with different samples that have diverse demographic characteristics to better understand its generalizability.

Artificial, laboratory-based aggression paradigms like the Tangram Task are not necessarily designed to capture the full essence, nuances, and contextual complexity of “real-world” aggression. Rather,

the primary goal of most laboratory research is the development of theories designed to explain underlying processes and mechanisms, it is these theoretical principles that one wishes to generalize, not the specific characteristics of the sample, setting, manipulation, or measure. (Anderson & Bushman, 1997, p. 22)

We encourage other researchers to use the Tangram Help/Hurt Task to explore its conceptual and methodological advantages and limitations.

Appendix

Tangram Assignment Table.

Easy				
Medium				
Hard				

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.

Notes

1. It may be appropriate to assess state-based aggression using observational ratings after an experimental manipulation; however, such situations are not common.
2. The categorization of tangrams into difficulty levels was also based on visual perception and completion times assessed in a pilot study. The methods and analyses are available in the Supplemental Material (available online).
3. Because this scale includes both positive and negative mood items, it can be seen as either a state hostility or a state happiness measure, depending on which subset of items is reversed.
4. For example, participants who chose three easy tangrams were assigned a “helping” score of 2. As in Gentile et al.’s (2009) work, the correlation between the raw number of easy and hard tangrams was large, $r_s > -.80$, $p_s < .001$. Using the “greater than 1” scoring procedure (Gentile et al., 2009) reduced this correlation, $r_s = -.73$ to $.77$, $p_s < .001$. Note that using the raw scores in the main analyses yielded results that were essentially the same as those reported here.
5. The high suspicion rate is likely due to conducting this study online where participants may question the presence of another participant taking part in the experiment at the same time as them and simultaneously completing the essay task.
6. This strategy was adopted for all subsequent studies. The Supplemental Material analyses describe the main analyses of Studies 4 to 6 without control variables.
7. See the Supplemental Material analyses for a specific test comparing the effects of the prime on both outcomes. Similar analyses are included for subsequent studies with multiple outcomes.
8. Due to the low numbers of males in this study, sex was not included in any subsequent analyses.

Supplemental Material

The online supplemental material is available at <http://pspb.sagepub.com/supplemental>.

References

- Achenbach, T. M., & Edelbrock, C. (1983). *Manual for the Child Behavior Checklist and revised child behavior profile*. Burlington: Department of Psychiatry, University of Vermont.
- Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality, 40*, 440-450.
- Anderson, C. A., & Bushman, B. J. (1997). External validity of "trivial" experiments: The case of laboratory aggression. *Review of General Psychology, 1*, 19-41.
- Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology, 53*, 27-51.
- Anderson, C. A., Deuser, W. E., & DeNeve, K. M. (1995). Hot temperatures, hostile affect, hostile cognition, and arousal: Tests of a general model of affective aggression. *Personality and Social Psychology Bulletin, 21*(5), 434-448.
- Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., . . . Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries. *Psychological Bulletin, 136*, 151-173.
- Barlett, C., Branch, O., Rodeheffer, C., & Harris, R. (2009). How long do the short-term violent video game effects last? *Aggressive Behavior, 35*, 225-236.
- Barlett, C. P., & Anderson, C. A. (2011). Re-Appraising the situation and its impact on aggressive behavior. *Personality and Social Psychology Bulletin, 37*, 1564-1573.
- Bartlett, M. Y., & DeSteno, D. (2006). Gratitude and prosocial behavior: Helping when it costs you. *Psychological Science, 17*, 319-325.
- Batson, C. D. (2014). *The altruism question: Toward a social-psychological answer*. New York, NY: Psychology Press.
- Batson, C. D., Klein, T. R., Highberger, L., & Shaw, L. L. (1995). Immorality from empathy induced altruism: When compassion and justice conflict. *Journal of Personality and Social Psychology, 68*, 1042-1054.
- Berkowitz, L., & LePage, A. (1967). Weapons as aggression-eliciting stimuli. *Journal of Personality and Social Psychology, 7*, 202-207.
- Bettencourt, B. A., & Miller, N. (1996). Gender differences in aggression as a function of provocation: A meta-analysis. *Psychological Bulletin, 119*, 422-447.
- Bushman, B. J. (1995). Moderating role of trait aggressiveness in the effects of violent media on aggression. *Journal of Personality and Social Psychology, 69*, 950-960.
- Bushman, B. J., & Anderson, C. A. (1998). Methodology in the study of aggression: Integrating experimental and nonexperimental findings. In R. Geen & E. Donnerstein (Eds.), *Human aggression: Theories, research, and implications for social policy* (pp. 23-48). San Diego, CA: Academic Press.
- Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology, 75*(1), 219-229.
- Buss, A. H. (1961). *The psychology of aggression*. New York, NY: John Wiley.
- Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology, 63*, 452-459.
- Carlson, M., Marcus-Newhall, A., & Miller, N. (1989). Evidence for a general construct of aggression. *Personality and Social Psychology Bulletin, 15*, 377-389.
- Cherek, D. R. (1981). Effects of smoking different doses of nicotine on human aggressive behavior. *Psychopharmacology, 75*(4), 339-345.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126.
- DeWall, C. N., & Baumeister, R. F. (2006). Alone but feeling no pain: Effects of social exclusion on physical pain tolerance and pain threshold, affective forecasting, and interpersonal empathy. *Journal of Personality and Social Psychology, 91*, 1-15.
- Dovidio, J. F., Piliavin, J. A., Schroeder, D. A., & Penner, L. (2006). *The social psychology of prosocial behavior*. Mahwah, NJ: Lawrence Erlbaum.
- Gadow, K. D., Sprafkin, J., & Grayson, P. (1990). The Help-Hurt Game as a measure of aggression in children with learning and behavior disorders. *Learning and Individual Differences, 2*, 337-351.
- 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, experimental, and longitudinal studies. *Personality and Social Psychology Bulletin, 35*, 752-763.
- 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.
- Giancola, P. R., & Parrott, D. J. (2008). Further evidence for the validity of the Taylor aggression paradigm. *Aggressive Behavior, 34*(2), 214-229.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504-528.
- Graziano, W. G., & Habashi, M. M. (2010). Motivational processes underlying both prejudice and helping. *Personality and Social Psychology Review, 14*, 313-331.
- Graziano, W. G., Habashi, M. M., Sheese, B. E., & Tobin, R. M. (2007). Agreeableness, empathy, and helping: A person \times situation perspective. *Journal of Personality and Social Psychology, 93*, 583-599.
- Greitemeyer, T., & Osswald, S. (2010). Effects of prosocial video games on prosocial behavior. *Journal of Personality and Social Psychology, 98*, 211-221.
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*, 348-362.
- Hirsh, J. B., Galinsky, A. D., & Zhong, C. B. (2011). Drunk, powerful, and in the dark how general processes of disinhibition produce both prosocial and antisocial behavior. *Perspectives on Psychological Science, 6*, 415-427.
- Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P., & Donohew, R. L. (2002). Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences, 32*, 401-414.

- Huesmann, L. R., Eron, L. D., Guerra, N. G., & Crawshaw, V. B. (1994). Measuring children's aggression with teachers' predictions of peer nominations. *Psychological Assessment, 6*, 329-336.
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102-138). New York, NY: Guilford Press.
- Joireman, J., Anderson, J., & Strathman, A. (2003). The aggression paradox: Understanding links among aggression, sensation seeking, and the consideration of future consequences. *Journal of Personality and Social Psychology, 84*, 1287-1302.
- Kalmoe, N. (2015). Trait aggression in two representative U.S. surveys: Testing the generalizability of college samples. *Aggressive Behavior, 41*, 171-188.
- Lang, J. W. B., & Fries, S. (2006). A revised 10-item version of the Achievement Motives Scale: Psychometric properties in German-speaking samples. *European Journal of Psychological Assessment, 22*, 216-224. doi:10.1027/1015-5759.22.3.216
- Liebert, R. M., & Baron, R. A. (1972). Some immediate effects of televised violence on children's behavior. *Developmental Psychology, 6*, 469-475.
- Macrae, C. N., & Johnston, L. (1998). Help, I need somebody: Automatic action and inaction. *Social Cognition, 16*, 400-417.
- Malle, B. F. (2011). Time to give up the dogmas of attribution: An alternative theory of behavior explanation. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 297-352). San Diego, CA: Academic Press.
- McGregor, H. A., Lieberman, J. D., Greenberg, J., Solomon, S., Arndt, J., Simon, L., & Pyszczynski, T. (1998). Terror management and aggression: Evidence that mortality salience motivates aggression against worldview-threatening others. *Journal of Personality and Social Psychology, 74*, 590-605.
- Mikulincer, M., Shaver, P. R., Gillath, O., & Nitzberg, R. A. (2005). Attachment, caregiving, and altruism: Boosting attachment security increases compassion and helping. *Journal of Personality and Social Psychology, 89*(5), 817-839.
- Mussweiler, T., & Förster, J. (2000). The sex → aggression link: A perception-behavior dissociation. *Journal of Personality and Social Psychology, 79*, 507-520.
- Pedersen, W. C., Gonzales, C., & Miller, N. (2000). The moderating effect of trivial triggering provocation on displaced aggression. *Journal of Personality and Social Psychology, 78*, 913-927.
- Penner, L. A., Fritzsche, B. A., Craiger, J. P., & Freifeld, T. R. (1995). Measuring the prosocial personality. In J. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 10). Hillsdale, NJ: Lawrence Erlbaum.
- Piff, P. K., Kraus, M. W., Cote, S., Cheng, B. H., & Keltner, D. (2010). Having less, giving more: The influence of social class on prosocial behavior. *Journal of Personality and Social Psychology, 99*, 771-784.
- Prot, S., Gentile, D. A., Anderson, C. A., Suzuki, K., Swing, E., Lim, K. M., . . . Lam, B. C. P. (2014). Long-term relations among prosocial-media use, empathy, and prosocial behavior. *Psychological Science, 25*(2), 358-368.
- Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma: A study in conflict and cooperation*. Ann Arbor, MI: University of Michigan Press.
- Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 38*, 119-125.
- Ritter, D., & Eslea, M. (2005). Hot sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior, 31*, 407-419.
- 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*, 281-287.
- Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game. *Psychological Science, 18*, 803-809.
- Sprafkin, J., & Gadow, K. D. (2001). The immediate impact of aggressive cartoons on emotionally disturbed and learning disabled children. *The Journal of Genetic Psychology, 149*, 35-44.
- Taylor, S., Gammon, C., & Capasso, D. (1976). Aggression as a function of the interaction of alcohol and threat. *Journal of Personality and Social Psychology, 34*, 938-941.
- Tedeschi, J. T., & Quigley, B. M. (1996). Limitations of laboratory paradigms for studying aggression. *Aggression and Violent Behavior, 1*, 163-177.
- 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.
- Twenge, J. M., Baumeister, R. F., Tice, D. M., & Stucke, T. S. (2001). If you can't join them, beat them: Effects of social exclusion on aggressive behavior. *Journal of Personality and Social Psychology, 81*(6), 1058-1069.
- van Baaren, R. B., Holland, R. W., Kawakami, K., & van Knippenberg, A. (2004). Mimicry and prosocial behavior. *Psychological Science, 15*, 71-74.
- Walder, L., Abelson, R., Eron, L. D., Banta, T. J., & Laulicht, J. H. (1961). Development of a peer-rating measure of aggression. *Psychological Reports, 9*, 497-556.
- Warburton, W. A. (2007). *Control-related aggression: Evidence for a mechanism, its origins, and personality correlates*. Dissertation submitted to Macquarie University, Sydney, Australia.
- Webster, G. D., DeWall, C. N., Pond, R. S., Deckman, T., Jonason, P. K., Le, B. M., . . . Bator, R. J. (2014). The brief aggression questionnaire: Psychometric and behavioral evidence for an efficient measure of trait aggression. *Aggressive Behavior, 40*(2), 120-139.

Supplementary Analyses

PILOT STUDY 1: ASSIGNMENT OF TANGRAM PUZZLES WITHIN DIFFICULTY LEVELS

The goal for this study was to test whether the tangram puzzles we categorized as easy, medium, and hard accurately represented their difficulty levels in terms of visual perception and the amount of time it takes to complete the puzzles.

Method

Participants

One hundred fifty participants (86 female; 56 male; 8 unidentified) were recruited from introductory psychology courses at a large Midwestern University for course credit. The mean age was 20.45, ($SD = 1.01$) years.

Materials

Tangram assignment sheet. We selected 30 tangrams from a tangram puzzle book and assigned them to easy, medium, and hard difficulty levels (similar to Appendix A). Tangram difficulty was initially assigned based on the number of pieces required to complete a particular tangram shape. Specifically, tangram shapes assigned to the easy category can be completed using 1-3 tangram pieces; medium shapes can be completed using 4-6 pieces; and hard shapes require all 7 pieces.

Tangram Completion Times. Participants were recorded and monitored through a video camera placed above the participant and directed at the table on which the participant worked on the puzzles. An experimenter monitored this camera in an adjacent room. Participants were randomly assigned to complete one of three tangram packets which contained 11 randomly selected tangram puzzles of varying difficulty levels. Three sets were used to reduce the

possibility that completion times would be biased by the particular set of Tangrams with which they were paired or the order in which they were presented. Experimenters recorded the number of seconds each participant took to complete each of the assigned tangrams. Participants were given up to two minutes to solve each Tangram puzzle after which they were encouraged to move on to the next tangram. Experimenters recorded 120 seconds as the completion time for any Tangram shapes that were not completed within the two minute time limit.

Tangram Difficulty Visual Perception. Participants rated the perceived visual difficulty of 10 randomly selected Tangram shapes on a scale (1=Easy; 2=Medium; 3=Hard). Participants were assigned one of three visual perception packets which contained 10 randomly selected tangram shapes of varying difficulty levels. Again, three sets were used to reduce the possibility that perceived difficulty ratings would be biased by the particular set of Tangrams with which they were paired or the order in which they were presented.

Procedure

Participants completed an informed consent upon arrival and were told that their session would be recorded. Participants were told that: (a) they will be interacting with another participant on a puzzle task; (b) that they would choose 11 tangrams for the “other participant” to complete; (c) that they would be assigned 11 tangrams by the “other participant”; and (d) that if the other participant completed 10 of the 11 tangrams within ten minutes, that person would win a \$25 gift certificate. Participants, however, were told that they could not win a gift certificate even if they complete the assigned tangrams in time, because one of the study’s goals was to determine whether a potential prize influences performance on this tangram task.

Participants received standardized tangram instructions and a practice packet. After demonstrating understanding of the tangram task, participants were led to a cubicle and provided

demographic information. Next, they chose 11 tangrams to assign the other participant by circling their choices directly on the tangram assignment table. After that, participants received a sheet of paper with ten tangrams and were asked to rate each of them on their perceived difficulty level. Next, participants received a packet of 11 tangrams, presumably assigned to them from the other participant, and were asked to complete as many of them as possible within 10 minutes. Participants were told to press a button at the completion of each tangram so that the experimenter could accurately record their time. Finally, participants were debriefed, thanked, and dismissed.

Results

Tangram Completion Times: Average amount of time (in seconds) taken to complete each tangram puzzle is displayed in Table 1. Tangrams 1-10 were pre-assigned to the easy category, Tangrams 11-20 to the medium category, and Tangrams 21-30 to the hard category. As discussed above, this pre-classification into the three difficulty types was based on the number of pieces required to complete a particular tangram shape. Table 1 reveals that three of the tangram puzzles (Tangrams 3, 4, and 5) assigned to the easy category had higher than expected average completion times for this category (36.28; 19.95; and 41.26 seconds, respectively).

Tangram Difficulty Visual Perception. The average perceived difficulty for each tangram puzzle is displayed in Table 1. Table 1 reveals that three of the tangram puzzles (Tangrams 3, 4, and 5) assigned to the easy category were perceived to be more difficult than other puzzles in the easy category (1.56; 1.42; and 1.46, respectively). Note, these three tangrams also had higher than expected average completion times for the easy category. In addition, three tangrams (Tangrams 15, 16, and 18) from the medium category were perceived to be easier than other tangram puzzles in this category (1.27; 1.15; and 1.08, respectively). These tangrams also

had relatively lower average completion times than most of the other tangrams in the medium category. Finally, one of the tangrams (Tangram 21) assigned to hard category was perceived to be easier (1.98) than other tangrams in this category. In Study 2, we replaced these seven tangrams (Tangrams 3, 4, 5, 15, 16, 18, and 21) with new tangrams that were deemed appropriate in terms of expected actual completion time and perceived difficulty for their assigned categories.

PILOT STUDY 2: FINDING REPLACEMENT PUZZLES

The goal for this study was to assess the completion time and perceived difficulty for seven new Tangrams that would serve to replace the seven Tangrams from Study 1 that were misfit for their assigned categories. In this study, we tested the completion time and perceived difficulty of 10¹ new Tangrams (3 easy, 5 medium, and 2 hard).

Method

Participants

Twenty participants (11 female; 9 male) from a large Midwestern University participated in the current study for course credit. The mean age was 19.40, ($SD = 0.92$) years.

Materials & Procedure

The tangram materials and procedures were the same as Pilot Study 1.

Results

Tangram Completion Times: Average amount of time (in seconds) taken to complete each tangram puzzle is displayed in Table 2. Tangrams 1-3 were pre-assigned to the easy category, 4-9 to the medium category, and 10-11 to the hard category. Table 2 reveals that three Tangram puzzles (Tangrams 1, 2, and 3) assigned to the easy category had acceptable average

¹ Ten Tangrams were tested because we expected some of these might not match their pre-assigned difficulty categories based on actual completion time or perceived difficulty.

completion time for this category (6.30; 5.65; and 7.45 seconds, respectively). Note that these times are comparable to the average completion times of easy puzzles in Study 1. Amongst the medium category puzzles, four Tangrams (Tangrams 4, 5, 6, and 8) had appropriate average completion times for this category. Because we only needed to replace three of the medium category puzzles from Study 1 we used perceived difficulty to determine which of the four medium puzzles from this study should replace the puzzles from Study 1. Finally, out of the two Tangrams pre-selected for the hard category, only one (Tangram 10) had acceptable average completion time for this category (104.40 seconds).

Tangram Difficulty Visual Perception. Average perceived difficulty for each tangram puzzle is displayed in Table 2. Table 2 reveals that three tangram puzzles (Tangrams 1, 2, and 3) assigned to the easy category were in fact perceived as easy ($M_s=1.00$; 1.00; and 1.10, respectively). Amongst the medium tangrams, we selected the three (Tangrams 4, 5, and 6) that were perceived to be the most difficulty ($M_s=1.90$; 1.80; and 2.15, respectively). Finally, out of the two Tangrams pre-assigned to the hard category, only one (Tangram 9) was perceived to be difficult ($M=2.95$). Thus, seven tangrams (Tangrams 1, 2, 3, 4, 5, 6, and 9) from Study 2 replaced the seven misfit tangrams from Study 1.

Final Tangram Assignment Table. The final tangram assignment table is displayed in Appendix A. The average completion time and perceived difficulty ratings for the 30 Tangrams included in the final tangram assignment table are presented in Table 3. Additionally, we report trimmed means using a .10 (10%) trimming setting to account for some outlier participants who took 120 seconds on even the easier tangram puzzles. The 0.10 trimming setting allows us to compute means for each tangram puzzle ignoring the top and bottom 10% of scores. As displayed in Figure 2, the correlation between the average completion time and perceived visual

difficulty (N =30 tangram items) was $r = 0.90$. The correlation between the trimmed means of average completion time and perceived difficulty was $r = 0.91$. This suggests that the tangrams assigned to the easy, medium, and hard categories in the final tangram assignment table reflect true differences in difficulty as assessed by completion time and perceived visual difficulty.

STUDY 4

Main Analyses Without Control Variables

Difference score. A one-way (condition: empathy-inducing/neutral) ANOVA revealed a significant effect of condition, $F(1, 270) = 9.45, p < .01, d = 0.37, \eta^2 = 0.03$. Participants who read the empathy-inducing essay were more helpful-less hurtful, $M = 6.37, 95\% \text{ CI } [5.65, 7.09]$, compared to participants who read the neutral essay, $M = 4.77, 95\% \text{ CI } [4.02, 5.50]$.

Separate helping and hurting scores. Participants who read the empathy-inducing essay were more likely to help than participants in the neutral essay condition ($M_s = 6.87 \text{ \& } 5.83, 95\% \text{ CIs } [6.38, 7.36], [5.34, 6.32], \text{ respectively}$), $F(1, 270) = 8.70, p < .01, d = 0.36, \eta^2 = 0.03$. Conversely, participants in the neutral essay condition were significantly more likely to hurt than participants in the empathy-inducing essay condition, ($M_s = 1.06 \text{ \& } 0.50, 95\% \text{ CIs } [0.77, 1.35], [0.21, 0.78], \text{ respectively}$), $F(1, 270) = 7.54, p < .05, d = 0.33, \eta^2 = 0.03$.

Study 4 Comparison of helping indices.

To test if the experimental manipulation influenced the helping scores from the tangram task differently than the monetary reward recommended for the other participant, a 2 (condition: empathy/neutral) x 2 (outcome: helping score from tangram task/monetary reward) mixed ANOVA was conducted with outcome as the within subjects factor. Both outcomes were standardized for this analysis. As expected, the condition effect was significant, $F(1, 269) = 16.61, p < .01$. However, the condition X outcome interaction was non-significant, $F < 1, p$

>0.10 suggesting that the effect of the experimental manipulation on helping behavior was not significantly different across the two outcome measures. This provides additional evidence for convergent validity of the tangram task as an assessment of helpful behaviors.

Study 5 Main Analyses Without Control Variables

Difference score. A one-way (condition: provocation/neutral) ANOVA revealed a significant effect of condition, $F(1, 135) = 8.21, p < .01, d = 0.49, \eta^2 = 0.06$. Participants in the provocation condition were more hurtful-less helpful, $M = -1.64, 95\% \text{ CI} [-2.65, -0.64]$, compared to participants in the neutral condition, $M = 0.43, 95\% \text{ CI} [-0.59, 1.46]$.

Separate helping and hurting scores. A 2 (condition: provocation/neutral) x 2 (behavior: helping/hurting) mixed ANOVA was conducted with behavior type as the within subjects factor. The condition X behavior interaction was significant, $F(1, 135) = 8.21, p < .01, \eta^2 = 0.06$. Simple effects analyses revealed that provoked participants scored higher on hurting than participants in the neutral condition ($M_s = 3.41$ & $2.22, 95\% \text{ CIs} [2.87, 3.96], [1.66, 2.78]$, respectively), $F(1, 127) = 9.04, p < .01, d = 0.52, \eta^2 = 0.06$. Conversely, participants in the neutral condition scored higher on helping than participants in the provocation condition, ($M_s = 2.66$ & $1.77, 95\% \text{ CIs} [2.10, 3.21], [1.23, 2.31]$, respectively), $F(1, 127) = 5.03, p < .05, d = 0.39, \eta^2 = 0.04$.

Study 6 Main Analyses Without Control Variables

Difference score. A one-way (condition: provocation/neutral) ANOVA revealed a significant effect of condition, $F(1, 97) = 14.39, p < .01, d = 0.77, \eta^2 = 0.13$. Provoked participants were less helpful/more hurtful compared to participants in the neutral condition, $M_s = -0.88; 2.92, 95\% \text{ CIs} [-2.28, 0.52], [1.51, 4.33]$, respectively.

Separate helping and hurting scores. A 2 (condition: provocation/neutral) x 2 (behavior: helping/hurting) mixed ANOVA was conducted with behavior type as the within subjects factor. The condition X behavior interaction was significant, $F(1, 97) = 14.39, p < .01, \eta p^2 = 0.13$. Simple effects analyses revealed that provoked participants scored higher on hurting than participants in the neutral condition ($M_s = 3.36$ & 1.35 , 95% CIs $[2.74, 3.98]$, $[0.72, 1.97]$, respectively), $F(1, 97) = 20.50, p < .001, d = 0.92, \eta p^2 = 0.17$. Conversely, participants in the neutral condition scored higher on helping than those who were provoked, ($M_s = 4.27$ & 2.48 , 95% CIs $[3.37, 5.16]$, $[1.59, 3.37]$, respectively), $F(1, 97) = 7.87, p < .05, d = 0.57, \eta p^2 = 0.08$.

Hot-Sauce Task. Recall there were two indices, the weight in grams of the hot-sauce assigned to the other participant and the product of the hot-sauce selected and the weight in grams. Two one-way (prime: provocation/neutral) ANOVAs revealed a significant effect of essay primes on both the weight and the weight X hot-sauce product, $F_s(1, 97) = 9.72, 4.60, p_s < .05, d_s = 0.63, 0.44, \eta p^2_s = 0.09, 0.05$. Provoked participants had higher scores on the weight and weight X hot-sauce product indices, $M_s = 11.11; 24.38$, 95% CIs $[9.33, 12.89]$, $[18.36, 30.39]$, than participants in the neutral condition, $M_s = 7.14; 15.14$, 95% CIs $[5.34, 8.93]$, $[9.33, 12.89]$.

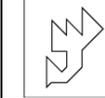
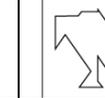
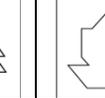
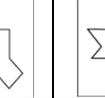
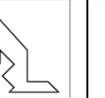
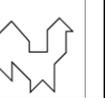
Evaluation. A one-way (prime: provocation/neutral) ANOVA revealed a significant effect of essay prime, $F(1, 96) = 37.50, p < .001, d = 1.25, \eta p^2 = 0.28$. Participants in the neutral, relative to provoked, condition gave the other participant more positive evaluations, $M_s = 5.34; 4.32$, 95% CIs $[5.11, 5.58]$, $[4.09, 4.56]$, respectively.

Monetary reward. A one-way (prime: provocation/neutral) ANOVA revealed a significant effect of essay prime, $F(1, 96) = 17.73, p < .001, d = 0.86, \eta p^2 = 0.16$. Participants in the neutral, relative to provoked, condition recommended a higher monetary reward for the other participant, $M_s = 11.04; 6.24$, 95% CIs $[9.44, 12.64]$, $[4.65, 7.84]$, respectively.

Tables

Table 1. Average completion time (CT, in seconds) and perceived difficulty (PD, 1-3) of tangrams in Study 1.

Easy										
Tangrams	1	2	3	4	5	6	7	8	9	10
										
Mean CT	5.76	5.07	36.28	19.95	41.26	6.47	19.70	10.48	16.74	8.16
SD CT	13.08	15.39	29.20	19.38	36.25	13.19	28.43	18.52	23.59	15.76
Range CT	1-120	1-120	6-120	7-120	4-120	2-120	5-120	2-120	2-120	1-100
Mean PD	1.13	1.04	1.56	1.42	1.46	1.10	1.19	1.05	1.08	1.21
SD PD	0.39	0.20	0.60	0.54	0.62	0.37	0.45	0.22	0.28	0.50
Range PD	1-3	1-2	1-3	1-3	1-3	1-3	1-3	1-2	1-2	1-3
Medium										
Tangrams	11	12	13	14	15	16	17	18	19	20
										
Mean CT	26.57	25.25	50.11	37.00	20.16	21.97	47.22	28.17	24.92	23.68
SD CT	23.60	25.49	37.43	26.19	16.81	30.23	31.89	25.71	25.17	13.15
Range CT	11-46	12-53	13-46	14-53	15-37	16-37	17-46	18-36	19-37	20-53
Mean PD	1.38	1.42	1.70	1.78	1.27	1.15	2.02	1.08	1.98	2.10
SD PD	0.53	0.50	0.61	0.62	0.45	0.36	0.68	0.27	0.56	0.63
Range PD	1-3	1-2	1-3	1-3	1-2	1-2	1-3	1-2	1-3	1-3

Hard										
Tangrams	21	22	23	24	25	26	27	28	29	30
										
Mean CT	51.11	89.22	49.89	63.47	71.85	75.14	46.41	75.93	66.43	50.41
SD CT	34.40	33.24	29.93	37.89	36.61	39.91	28.24	39.77	39.00	35.40
Range CT	15-120	2-120	15-120	16-120	24-120	15-120	19-120	17-120	2-120	8-120
Mean PD	1.98	2.55	2.15	2.67	2.53	2.60	2.54	2.45	2.51	2.33
SD PD	0.73	0.60	0.70	0.56	0.50	0.61	0.62	0.64	0.62	0.66
Range PD	1-3	1-3	1-3	1-3	2-3	1-3	1-3	1-3	1-3	1-3

Mean CT = Average Completion Time; SD CT = Standard Deviation of Completion Times; Range CT = Range of Completion Times; Mean PD = Average Perceived Difficulty; SD PD = Standard Deviation of Perceived Difficulty; Range PD = Range of Perceived Difficulty.

Table 2. The average completion time and perceived difficulty of tangrams in Study 2

Tangrams	1	2	3	4	5	6	7	8	9	10
										
Mean CT	6.30	5.65	7.45	24.95	27.00	28.15	22.05	26.00	104.40	22.05
SD CT	2.39	2.39	2.48	4.59	5.38	5.58	17.51	32.04	15.04	12.14
Range CT	2-10	1-10	3-12	18-31	20-38	20-37	7-79	3-120	83-120	7-60
Mean PD	1.00	1.00	1.10	1.90	1.80	2.15	1.40	1.35	2.95	1.65
SD PD	0.00	0.00	0.31	0.45	0.41	0.37	0.50	0.49	0.22	0.67
Range PD	1-1	1-1	1-2	1-3	1-2	2-3	1-2	1-2	2-3	1-3

Table 3. The average completion time and perceived difficulty of tangrams in the final tangram assignment table.

Easy										
Tangrams	1	2	3	4	5	6	7	8	9	10
										
Mean CT	5.76	5.07	5.65	7.45	6.30	6.47	19.70	10.48	16.74	8.16
TMean CT	4.08	2.34	5.67	7.44	6.33	4.73	15.62	7.05	12.45	5.74
SD CT	13.08	15.39	2.39	2.48	2.39	13.19	28.43	18.52	23.59	15.76
Mean PD	1.13	1.04	1.00	1.10	1.00	1.10	1.19	1.05	1.08	1.21
TMean PD	1.07	1.00	1.00	1.06	1.00	1.05	1.14	1.00	1.05	1.14
SD PD	0.39	0.20	0.00	0.31	0.00	0.37	0.45	0.22	0.28	0.50
Medium										
Tangrams	11	12	13	14	15	16	17	18	19	20
										
Mean CT	26.57	25.25	50.11	37.00	24.95	28.15	47.22	27.00	24.92	23.68
TMean CT	22.93	22.24	48.64	34.92	24.95	28.11	45.40	26.78	22.77	22.33
SD CT	23.60	25.49	37.43	26.19	4.59	5.58	31.89	5.38	25.17	13.15
Mean PD	1.38	1.42	1.70	1.78	1.90	2.15	2.02	1.80	1.98	2.10
TMean PD	1.34	1.41	1.67	1.75	1.89	2.11	2.02	1.83	1.98	2.11
SD PD	0.53	0.50	0.61	0.62	0.45	0.37	0.68	0.41	0.56	0.63

Hard										
Tangrams	21	22	23	24	25	26	27	28	29	30
										
Mean CT	104.40	89.22	49.89	63.47	71.85	75.14	46.41	75.93	66.43	50.41
TMean CT	104.72	91.41	48.12	63.10	71.83	75.74	45.09	76.50	66.56	49.09
SD CT	15.04	33.24	29.93	37.89	36.61	39.91	28.24	39.77	39.00	35.40
Mean PD	2.95	2.55	2.15	2.67	2.53	2.60	2.54	2.45	2.51	2.33
TMean PD	3.00	2.61	2.17	2.73	2.53	2.66	2.59	2.50	2.56	2.36
SD PD	0.22	0.60	0.70	0.56	0.50	0.61	0.62	0.64	0.62	0.66

Mean CT = Average Completion Time; TMean CT = Trimmed Average Completion Time; SD CT = Standard Deviation of Completion Times; Mean PD = Average Perceived Difficulty; TMean PD = Trimmed Average Perceived Difficulty; SD PD = Standard Deviation of Perceived Difficulty.

Table 4. Correlations between helping, hurting, difference score, and pre-experimental measures in Studies 4-6.

	Study 4 N = 272			Study 5 N = 152			Study 6 N = 99		
	Helping	Hurting	Help-Hurt	Helping	Hurting	Help-Hurt	Helping	Hurting	Help-Hurt
BP Physical	-0.03	0.04	-0.04	-0.14	0.30**	0.24*	-0.08	0.18	-0.13
BP Verbal	-0.09	0.15*	-0.12+				0.02	0.13	-0.05
BP Anger	-0.12*	0.14*	-0.14*				0.04	0.08	-0.01
BP Hostility	-0.12	0.07	-0.11				-0.02	0.08	-0.05
BP Composite	-0.12*	0.13*	-0.13*				-0.02	0.17	-0.09
Prosocialness	0.04	-0.04	0.04				0.29**	-0.30**	0.31**
Extraversion							0.07	-0.07	0.07
Agreeableness							0.03	-0.16	0.09
Conscientiousness							-0.22*	0.22*	-0.24*
Emotional Stability							-0.08	0.06	-0.08
Exploration							0.07	-0.09	0.08

+ $p = .05$, * $p < .05$, ** $p < .01$.