Title

Forgiveness takes place on an attitudinal continuum from hostility to friendliness: Toward a closer union of forgiveness theory and measurement.

Journal

Journal of personality and social psychology, 119(4)

ISSN

0022-3514

Authors

Forster, Daniel E
Billingsley, Joseph
Russell, V Michelle
et al.

Publication Date

2020-10-01

DOI

10.1037/pspi0000227

Peer reviewed

Forgiveness Takes Place on an Attitudinal Continuum from Malevolence to Benevolence: Toward a Closer Union of Forgiveness Theory and Measurement

Preprint · March 2019

DOI: 10.31234/osf.io/wkvhn

9 authors, including:

Daniel E Forster
University of Miami
18 PUBLICATIONS **321** CITATIONS
SEE PROFILE

Joseph Billingsley
North Carolina State University
13 PUBLICATIONS **46** CITATIONS
SEE PROFILE

V. Michelle Russell
University of North Carolina at Greensboro
14 PUBLICATIONS **324** CITATIONS
SEE PROFILE

Michael E Mccullough
University of Miami
148 PUBLICATIONS **17,165** CITATIONS
SEE PROFILE

Forgiveness Takes Place on an Attitudinal Continuum from Malevolence to Benevolence: Toward a Closer Union of Forgiveness Theory and Measurement

Daniel E. Forster[1,2], Joseph Billingsley[1], V. Michelle Russell[3,4], Jeni L. Burnette[3], Adam Smith[5,6], Yohsuke Ohtsubo[5], Thomas G. McCauley[1], Joanna Schug[7], Debra Lieberman[1], & Michael E. McCullough[1*]

[1] University of Miami
[2] U.S. Army Research Laboratory
[3] North Carolina State University
[4] University of North Carolina, Greensboro
[5] Kobe University
[6] Nagoya University
[7] College of William and Mary

*Corresponding author: Michael E. McCullough, email: mikem@miami.edu

Note: This is a non-peer-reviewed pre-print.

ABSTRACT

Researchers commonly conceptualize forgiveness as a rich complex of psychological changes involving attitudes, emotions, and behaviors. Psychometric work with the measures developed to capture this conceptual richness, however, often point to a simpler picture of the latent psychological dimensions along which forgiveness takes place. In an effort to better unite forgiveness theory and measurement, we evaluate several psychometric models for a frequently used measure of forgiveness. In doing so, we assess samples from both the United States and Japan to understand forgiveness in both non-close and close relationships. In addition, we assess the predictive utility of these models for several behavioral outcomes that traditionally have been linked to the motivations underlying forgiveness. Our results highlight models based on correlated factors models and bifactor (S-1) models. The bifactor (S-1) model evinced particular utility, identifying a single psychological dimension that spans from malevolence to benevolence while also pointing to other sources of variance that may be conceived of as method factors. The general factor of the bifactor (S-1) model consistently predicts variation in relevant criterion measures, including four different experimental economics games (when played with a transgressor), and also suffuses a second self-report measure of forgiveness. Taken together, these results suggest that forgiveness can be productively conceptualized as prosocial change along a single attitudinal continuum that ranges from malevolence to benevolence.

Keywords: forgiveness, TRIM, bifactor model, cross-cultural, reconciliation

INTRODUCTION

Because interpersonal conflict is unavoidable, forgiveness is crucial for maintaining social relationships: People must be able to move past negative motivations if they are to retain their valuable social partners. Indeed, socially reparative behaviors have been observed in multiple primate species (De Waal & Pokorny, 2005). In humans, researchers have made considerable progress in mapping the information-processing mechanisms that regulate this crucial psychological process (Fehr, Gelfand, & Nag, 2010). Noteworthy advances include: (i) an emerging consensus on a basic definition of interpersonal forgiveness as “prosocial change toward a perceived transgressor” (McCullough, Pargament, & Thoresen, 2001; p. 9); (ii) the development of theoretical models of forgiveness that expand upon this basic definition by specifying a conceptually rich and multidimensional set of psychological changes that can involve attitudes, emotions, behaviors, and physiology (Worthington et al., 2015); and (iii) the validation of multiple instruments for measuring forgiveness as conceptualized in these more complex models (see Worthington et al., 2015).

Despite these advances, progress in understanding the psychological constructs that underlie forgiveness has suffered from inconsistencies between theoretical and empirical conceptualizations of forgiveness. Indeed, the research literature provides multiple hints of possible misfit between *a priori* conceptions of forgiveness and how measurement tools are used to model those conceptions. For example, McCullough and colleagues have depicted forgiveness as a suite of motivational changes whereby a victim becomes less vengeful, less avoidant, and more benevolent toward a transgressor (McCullough, Fincham, & Tsang, 2003; McCullough & Hoyt, 2002). To

measure this three-factor conception of forgiveness, they developed the self-report Transgressor-Related Interpersonal Motivations Inventory (TRIM-18; McCullough, Root, & Cohen, 2006). In some of the work using this questionnaire, researchers have scored the TRIM Inventory as if it reflects the operation of three distinct motivations--revenge, avoidance, and benevolence (Bono, McCullough, & Root, 2008; Carmody, Gordon, & Differences, 2011; Tsang, McCullough, & Fincham, 2006). Despite the supposed conceptual distinctions among these three theoretical constructs, the subscales used to measure them are often highly intercorrelated (e.g., McCullough, Root, & Cohen, 2006). In other studies, researchers have used versions of the TRIM Inventory to model two distinct motivations—the motivation to seek revenge and a bipolar motivation that ranges from avoidance to benevolence (e.g., McCullough, Bono, & Root, 2007; McCullough et al., 2006). In still other work, researchers have simply summed the items on the TRIM Inventory as if forgiveness reflects change across a single attitudinal or continuum that runs from malevolence to benevolence (Harper et al., 2014; McCullough, Luna, Berry, Tabak, & Bono, 2010; McCullough, Pedersen, Tabak, & Carter, 2014; Ohtsubo, Yamaura, & Yagi, 2015; Tabak, McCullough, Luna, Bono, & Berry, 2012; Worthington et al., 2015). Which of these models best reflects the actual psychological dimension or dimensions in which forgiveness takes place? Are there really multiple distinct motivational changes underlying forgiveness, or does forgiveness instead mostly reflect changes across a single underlying psychological dimension?

Other researchers have faced similar challenges in matching their conceptions of forgiveness with the empirical realities of their measures. Subkoviak et al. (1995), for example, defined forgiveness as the confluence of positive and negative cognitive,

affective, and behavioral components in which “a person overcomes resentment toward an offender, but does not deny him/herself the moral right to such resentment. The forgiver tries to have a new stance of benevolence, compassion, and even love toward the offender, even though the latter has no moral right to such a response” (p. 642). Subkoviak et al. sought to measure this notion of forgiveness, in all of its conceptual richness, with the 60-item Enright Forgiveness Inventory, which was designed with six distinct subscales to reflect both positive and negative manifestations of cognition, affect, and behavior. Despite the conceptual depth of their six-dimensional model of forgiveness, a single summary score based on all 60 items behaves in every way (e.g., internal consistency, test-retest reliability, correlation with other measures) similarly to the six individual subscales (see also Orathinkal, Vansteenwegen, Enright, & Stroobants, 2007), suggesting that the conceptual distinctions between cognitive, affective, and behavioral components are not borne out empirically.

Similarly, Rye et al. (2001) developed their self-report measure of forgiveness (“The Forgiveness Scale”) with the goal of differentiating between motivations toward the offender due to the presence of positive emotions and the absence of negative emotions. Despite their efforts to separate these concepts, the two subscales exhibited a strong positive correlation, achieved similar internal consistency and test-retest reliabilities, and evinced similar correlations with a variety of validational targets—all of which suggests that the subscales may reflect a single latent continuum rather than two meaningfully different dimensions of forgiveness.

These three sets of results suggest that the underlying attitudinal dimensions upon which forgiveness takes place may be conceptually leaner than researchers have

heretofore posited. Indeed, they all suggest that subjects' scores on the sort of items that are often used on these scales (e.g., "I'll make him or her pay"; "I'd keep as much distance between us as possible"; "Even though his/her act hurt me, I still have good will for him/her") are caused at least in part by a general factor that runs from negative (i.e., malevolent) evaluative reactions to the transgressor to positive evaluative reactions to the transgressor, perhaps along with additional factors that influence scores on some of the items (e.g., items related to revenge) but not others (e.g., items related to benevolence). Since a malevolence-benevolence dimension appears to suffuse interpersonal behavior in general (as exemplified in the Interpersonal Circumplex model of interpersonal behavior; Gurtman, 2009), it would hardly be surprising to find that forgiveness also reflected movement away from a hostile attitude regarding a transgressor and movement toward a friendly or conciliatory attitude. To the extent that items are caused both by such a general factor and extraneous influences (e.g., residual motives, method effects), efforts to measure forgiveness may be impeded by psychometric models that cannot easily separate these two types of variance.

There would be additional theoretical benefit from the ability to more precisely model forgiveness as a single latent continuum that runs from negative to positive evaluations toward a transgressor. The prospect that the questionnaire items that various researchers use to measure forgiveness (which often involve self-reports of affects, cognitions, and behaviors toward a transgressor) can be neatly summarized with a single general factor would suggest that forgiveness could reasonably be described as a positive change in one's attitude toward an offender (inasmuch as recent formal accounts depict attitudes as global evaluative reactions toward an attitude object

that incorporate beliefs, feelings, and behaviors; Dalege et al., 2016). This could prove to be a theoretical boon for forgiveness research because so much is already known about the nature of attitudes and attitude change (Albarracin & Shavitt, 2018; Bohner & Dickel, 2011). Locating the concept of forgiveness within the broader conceptions of attitudes and attitude change could speed future progress substantially.

In hopes of better uniting forgiveness theory and measurement, here we systematically evaluated several modeling options for the TRIM Inventory in hopes of determining the model that best depicts the underlying psychological dimension (or dimensions) in which forgiveness takes place. To do so, we compared confirmatory models used in the existing literature (one-factor, two-factor, and three-factor models) to a model that depicts forgiveness primarily as psychological change in a single attitudinal dimension (e.g., malevolence-benevolence, or hostile-friendly), along with additional dimensions that might influence scores on some of the items but not others. These latter models, which depict a general factor and additional item-specific content (or method) factors, can be specified using variants of the bifactor modeling approach (Eid, Geiser, Koch, & Heene, 2017).

Bifactor Modeling: General and Specific Motives Underlying Forgiveness

Recent interest in measuring general factors that span specific factors has spurred the development of different so-called *G*-factor models (Eid et al., 2017), such as bifactor models (Holzinger & Swineford, 1937). In a traditional bifactor model, a general factor explains item responses across all domains while residual factors explain item responses on their specified domains (Reise, 2012). However, the traditional model is inappropriate when applied to typical survey methods, which has motivated the

development of alternative bifactor modeling approaches (Eid et al., 2017). One of the models Eid et al. (2017) proposed is referred to as the bifactor (S-1), in which the general factor loads on all indicators (as in the traditional bifactor model), but specific (S) factors are modeled for all but one of the specific domains (hence, S-1), while the non-modeled specific domain defines the scale of the general factor. When applying this model to the TRIM Inventory, one might specify a model in which a malevolence-benevolence factor explains item responses across domains of avoidance, benevolence, and revenge, while simultaneously modeling residual factors for two of the three domains, thereby leaving one domain to set the scale for the general malevolence-benevolence factor.

To apply the bifactor (S-1) model to the TRIM Inventory, one thus faces the challenge of selecting which domain best represents the general factor. We suggest that avoidance is the proper reference domain because forgiveness is most consistently conceptualized as a reduction in people's motives to avoid their transgressors; in other words, forgiveness is conciliatory movement away from malevolence toward benevolence in which victims give their transgressors a subsequent opportunity to be good social partners. On this view, the motivations to re-engage (or to reduce avoidance) are not necessarily benevolent or vengeful. The use of avoidance as a reference domain is also supported by research indicating that approach and avoidance motives are distinct, and that benevolence and revenge motives are both approach-oriented, thereby making them share something important that is not shared by avoidance (Elliot & Thrash, 2002). See Figure 2 for our depiction of the bifactor (S-1) model as applied to the TRIM Inventory.

Does the TRIM Inventory Predict Behavior toward the Transgressor?

In addition to issues of specifying models that capture our understanding of forgiveness as a construct, researchers who study forgiveness often lament the fact that self-report measures of forgiveness are rarely validated against behavioral measures that could demonstrate that they reflect interpersonally consequential psychological processes (Worthington et al., 2015; for notable exceptions, see Carlisle et al., 2012; Dorn, Hook, Davis, Van Tongeren, & Worthington Jr, 2014; Exline, Baumeister, Bushman, Campbell, & Finkel, 2004). Without such validations against behavioral measures, researchers face an uphill battle in arguing that the motivational changes they measure via self-report lead to changes in how the victim actually treats the offender. Indeed, on an evolutionary view of forgiveness, its function is to motivate behaviors that signal to a transgressor one's willingness to re-establish cooperative relations contingent on amended behavior on the transgressor's part (McCullough, Kurzban, & Tabak, 2013). If the function of forgiveness is indeed to promote prosocial interactions, and a self-report measure of those motivations to do so is valid, then measures of that motivational change from malevolence to benevolence should correlate with behavioral measures that might signal a willingness to refrain from revenge and return to cooperative relations.

A variety of standardized experimental economics games that are generally taken to reflect people's willingness to trust others, to share with others, to cooperate with others in the pursuit of mutual benefit, to uphold the principle of fairness, and to retaliate (or, conversely, to refrain from retaliating) have been extensively studied over the past several decades. Games such as the iterated prisoner's dilemma (Axelrod,

1980), the Dictator Game (Forsythe, Horowitz, Savin, & Sefton, 1994), the Trust Game (Berg, Dickhaut, & McCabe, 1995), and the pay-to-punish game (Fehr & Fischbacher, 2004) all appear to reflect the operation of a domain-general phenotype that creates covariances among these games (McAuliffe, Forster, Pedersen, & McCullough, 2018; Peysakhovich, Nowak, & Rand, 2014). Furthermore, the shared variance among these games is reflected in peer-judgments of people's altruistic and trusting tendencies (McAuliffe et al., 2018), suggesting that people's scores on these laboratory-based behavioral measures reflect a friendly or prosocial approach to real-life interpersonal interactions. These behavioral economic measures of prosocial behavior therefore present unique opportunities to assess whether forgiveness as measured by self-report is associated with affiliative interpersonal behavior toward the person whom one has forgiven.

The Present Project

Here, we sought to assess the relative fit and predictive utility of alternative psychometric models, including the bifactor (S-1) model and other more traditional confirmatory factor models used for the TRIM-18 in the past. Further, we tested these models using a new form of the TRIM-18 designed for non-close others (TRIM-NCO), as well as the original TRIM-18. To pursue our research questions in non-close others, we implemented an experimental paradigm across three settings: Online subjects from the United States, laboratory subjects from the United States, and laboratory subjects from Japan. In each of the three experiments, we also evaluated whether the factors we estimated predicted people's scores on the iterated prisoner's dilemma game, the dictator game, the trust game, and the pay-to-punish game. To then test the

generalizability of our modeling approaches to different relationship types and measurement scales, we conducted two non-experimental survey studies across two settings: Laboratory subjects from the United States and laboratory subjects from Japan. In these survey studies, we used a self-report measure of forgiveness as the criterion by which to assess each model's predictive utility, and we evaluated the bifactor (S-1) model's ability to account for scores on other researchers' measures of forgiveness as well (Hook, Worthington Jr, & Utsey, 2009; Hook, Worthington Jr, Utsey, Davis, & Burnette, 2012).

STUDY 1 METHOD

Subjects. Subjects were 1,887 workers from Amazon's Mechanical Turk (MTurk; 49% female; Age: M = 33.93 years; SD = 10.48). Per our preregistration, we removed all suspicious subjects from our analyses[1] (n = 867), resulting in a final sample of 1,020.[2] Results from analyses of all subjects (suspicious and non-suspicious) are available in the supplemental materials but did not change meaningfully from those reported in the main text. All subjects entered the experiment for a guaranteed $3.50 with the expectation of a bonus that would depend on their decisions in the study; in reality, everyone received a $6.50 bonus and therefore earned $10 for participating.

Procedure

[1] We presented analyses with non-suspicious subjects here to be consistent with our preregistration plan, but in the time since preregistering we learned that this practice has little empirical utility. One should analyze all data to obtain (1) more statistical power in the event that suspicious and non-suspicious subjects are statistically indistinguishable; (2) results that generalize beyond a sub-population that, for unknown reasons, would not report suspicion; and (3) results that do not violate the rules of causal inference based on experimentation (Gupta, 2011). We also note here that our results with suspicious subject included were consistent with those reported here (see supplemental materials).

[2] Our rates of suspicion may appear high, but we were very conservative in who we considered 'suspicious' (e.g., subjects who clearly reported suspicion because we expressly asked them about suspicion). See supplemental materials for details.

We used a series of experimental economic games to manipulate a transgression and measure forgiveness behaviorally. Complete procedures were preregistered on the Open Science Framework and are available in the supplemental materials. In this manuscript, we present only the procedures that yielded the information for the analyses conducted here. We conducted analyses to address separate research questions in Billingsley et al. (in prep). See Table 1 for an outline of the experimental design. Subjects were assigned to one of 9 conditions in a 3 (Message: Control, Apologetic, Aggravating) x 3 (Economic Game: Trust Game, Dictator Game, Pay-to-Punish Game) between-subjects design.

Subjects were told that the study was designed to examine how communication influences decision-making across a variety of tasks, and that they would be engaging in a series of authentic interactions with another MTurk worker. To enhance the believability of this interaction, we programmed the experiment using SoPHIE - the Software Platform for Human Interaction Experiments (Hendriks, 2012). SoPHIE enabled us to set up an online “waiting room,” which subjects entered after consenting to participate. Subjects remained in the waiting room until a second subject joined, whereupon the two actual subjects were paired and given the opportunity to engage in an authentic communication task. Subjects who spent seven minutes in the waiting room without being paired were dropped from the experiment, compensated $0.70 for their time, and permitted to participate in a successive session of the experiment. After being paired, subjects were told that the purpose of the communication task was to familiarize them with the program’s chat function, which would be used throughout other tasks in the experiment. During the communication task, subjects took turns sending

and receiving five short messages without any guidance about what they should discuss. Therefore, when people began engaging in experimentally manipulated interactions later in the experiment (particularly during the Prisoner's Dilemma Game, see below), they would have had some basis for believing the interactions were real.

Iterated Prisoner's Dilemma Game

Following Tabak et al. (2012), we manipulated transgressions and apologies using an iterated Prisoner's Dilemma Game (PDG; called the "Decision-Making Task" in this experiment). Subjects were introduced to the rules of the PDG and were told that they could earn money depending on their interactions with the other subject (a pre-programmed script, hereafter referred to as the "confederate"). If both subjects cooperated, they both earned $0.10 for the round; if both subjects defected, they both earned $0.05 for the round; and if one subject defected and the other cooperated, the defector earned $0.15 and the cooperator earned $0.00. The PDG lasted 29 rounds for each subject, but subjects were told that they would play for an unspecified number of rounds to prevent any end-game effects. For the first 12 rounds, the confederate played a generous tit-for-tat strategy: It always cooperated on the first round and, so long as the subject cooperated on any round, the confederate cooperated on the subsequent round. However, if the subject defected on any round, the confederate defected on the subsequent round with a 50% probability.

Because our experimental manipulations included post-transgression apologies with compensation, as a cover story we informed subjects that there would be intermittent opportunities for communication throughout the decision-making task. Subjects were told that they had been assigned to either a "Sender" or "Receiver" role,

and that the Sender could send both text and game earnings. In reality, all subjects were assigned to the “Receiver” condition. As Receivers, subjects were able to respond to the Sender’s messages, but were not able to return any money to the Sender.

Transgressions

Following the 12th round of the PDG, the game was interrupted by the first opportunity for the confederate to send a message to the subject. As in Tabak et al. (2012), every subject received the same message, which stated, “i think we should both just hit cooperate” [sic]. Subjects were then prompted to respond. This message was designed to give subjects the expectation of cooperation, regardless of the subjects’ strategies earlier in the game. Following the message, subjects returned to the Prisoner’s Dilemma and were unconditionally defected against for 7 consecutive rounds.

Apology Manipulation

Following the 7 rounds of unconditional defection, the confederate again sent a message to the subject. For Study 1, subjects received one of three messages. In the control condition, subjects read the message, “this takes more concentration than i thought it would. at least it's more interesting than the HIT i did last time” [sic]. In the apologetic condition, “sorry for defecting after i said cooperate. i won't do it again. i'll send over some money to make it up to you” [sic]. In the aggravating condition, “sucks for you, that’s just how you play the game. i’m just trying to make as much money as i can” [sic]. In the apology condition, but not in the other conditions, confederates sent $1.00 of their earnings as compensation to subjects. The offer of compensation was included based on evidence that apologies are most effective at influencing forgiveness

when the transgressor incurs some cost as part of the apology (Ohtsubo et al., 2018; Ohtsubo & Watanabe, 2009; Ohtsubo et al., 2012). After the apology manipulation, the confederate unconditionally cooperated for two rounds, then resumed a generous tit-for-tat strategy for the remaining 8 rounds.

Measures

Defections in the Prisoner's Dilemma Game

We assessed forgiveness behaviorally by analyzing the ten rounds of Prisoner's Dilemma Game behavior that followed the message manipulation. Rather than using composite scores of the number of defections, we treated each decision in the final ten rounds as an item response manifested from a latent 'propensity to defect' variable, modeled as a two parameter logistic (2PL) item response model. Such a model accounts for the possibility that item difficulties might actually vary across the final ten decisions of the Prisoner's Dilemma. In other words, it is possible that on average subjects more readily defect during earlier decisions in the sequence than during later decisions. The 2PL item response model incorporates such item-based variability into its estimate of each subject's propensity to defect, and should therefore yield more accurate assessments than a typical composite score.

One-Shot Economic Games

Following the Prisoner's Dilemma Game, we randomly assigned subjects to play one of three experimental economics games (Game: Trust Game, Dictator Game, Pay-to-Punish Game) in a between-subjects manipulation. Each game provided an opportunity to study a different forgiveness-relevant social motivation (behavioral trust, behavioral benevolence, or behavioral revenge, respectively).

In the Trust Game (TG; Berg et al., 1995), two parties are given an endowment and randomly assigned to be either ‘Truster’ or ‘Trustee’ (we used the terms ‘First Mover’ and ‘Second Mover’ in our experiment). The Truster is able to send any amount of the endowment to the Trustee. In turn, the Trustee receives a multiple of the amount sent (three times, in our experiment). Subsequently, the Trustee is able to return any proportion of the amount received (up to three times the amount sent) back to the Truster. After receiving the instructions for the game and playing an example game from the perspective of both roles, subjects were assigned to the role of the first mover (Truster), ostensibly through random assignment.

In the Dictator Game (DG; Forsythe et al., 1994), the setup is very similar to the Trust Game, in that two parties are given an endowment and are randomly assigned to be either the ‘Dictator’ or ‘Recipient’ (we used the terms ‘Decision-Maker’ and ‘Recipient’ in our experiment). In the DG, the Dictator can send any amount of the endowment to the Recipient at a 1:1 ratio (i.e., the amount gained by the Recipient is equal to the amount paid by the Dictator). The Recipient has no opportunity to influence the Dictator’s decision, and the game ends once the Dictator’s decision is made. After receiving the instructions for the game and playing an example game from the perspective of both roles, subjects were assigned to the role of the Decision-Maker (Dictator), ostensibly through random assignment.

In the Pay-to-Punish Game (PTPG; Fehr & Fischbacher, 2004), the setup is nearly identical to the DG, in that two parties are given an endowment and randomly assigned to be either ‘Decision-Maker’ or ‘Recipient’. In the PTPG, the Decision-Maker is able to spend any amount of the endowment to remove some amount from the

Recipient's endowment (in our experiment, the amount removed was four times the amount paid). As in the DG, the Recipient is not given an opportunity to influence the Decision-Maker, and the Decision-Maker cannot make any financial gains by removing money from the recipient. After receiving the instructions for the game and playing an example game from the perspective of both roles, subjects were assigned to the role of the Decision-Maker, ostensibly through random assignment.

Self-Report Measure of Forgiveness: Transgression Related Interpersonal Motivations Inventory for Non-Close Others

Immediately following subjects' decisions in the second experimental economic game, subjects completed the 18-item Transgression Related Interpersonal Motivations Inventory for Non-Close Others (TRIM-NCO; see Appendix A). Typically, researchers have measured forgiveness using items that assume an existing relationship between victim and transgressor, thereby limiting our ability to assess forgiveness between people who are interacting for the first time (as in a typical laboratory context). The TRIM-NCO Inventory is designed to overcome this limitation by re-wording items from the original TRIM-18 Inventory so that they are sensible to people who are rating their attitudes toward strangers.

Although we use the TRIM-NCO Inventory as a predictor of behavioral indicators of willingness to restore cooperative relations, we avoided assessing self-report forgiveness prior to our behavioral measures to prevent any contamination effects. We based the items in the TRIM-NCO on a prior measure, the TRIM-18, which McCullough et al. (2006) developed for measuring interpersonal forgiving in close relationships. As with the traditional TRIM-18, the TRIM-NCO provides a self-report measure of the

interpersonal motivations hypothesized to underlie forgiveness, using subscales to assess avoidant, benevolent, and vengeful inclinations.

Suspicion Probes and Debriefing

Subjects responded to a series of funnel debriefing questions designed to probe whether they were suspicious of the deception in the experiment (Aronson, Carlsmith, & Ellsworth, 1990). We asked subjects a series of "yes/no" questions to determine whether they had questions or comments about specific aspects of the experiment, followed by a free-response option if their answer indicated that they might be suspicious. Finally, we explained the true nature of the experiment and provided an explanation for our use of deception in the experiment.

Other Measures

We also measured subjects' perceptions of the transgressor's relationship value and exploitation risk during the PDG, although these data were not analyzed here because they are beyond the scope of this manuscript. Analyses incorporating these measures were used to address research questions distinct from those we addressed here and will be available in a companion manuscript (Billingsley et al., in prep).

STUDY 1 RESULTS

Analyses were conducted using Mplus version 7 (Muthén & Muthén, 1998-2012). Syntax and output are available in the supplemental materials.

Factor Analysis of TRIM-NCO

We compared confirmatory models based on previous uses of the TRIM-18 (for close others), which has been modeled using three correlated factors (avoidance, benevolence, and revenge), two correlated factors (avoidance-benevolence and

revenge), and a single factor ('forgiveness')[3]. We also analyzed the TRIM-NCO using a bifactor (S-1) model, in which a general factor explains variation in all item responses while specific factors explain unique variation among subsets of items.

Model fit for the alternative factor structures are available in Table 2a. The three-factor and bifactor (S-1) models fit best, with the fit of these two models being indistinguishable under reasonable constraints (Geiser, Eid, & Nussbeck, 2008); therefore, we opted to compare models of forgiveness as bifactor (S-1) and three factor models. Path coefficients for each of these models are displayed in Table 3a.

We also concluded that the item, "I hope he/she gets what he/she deserves," which was designed as a revenge indicator, performed poorly at both the general and specific levels. This was demonstrated by its abnormally low factor loadings across all of the models we tested. Our interpretation of this result is that the item is inherently ambiguous—people believe that those who behave prosocially deserve good things and should get what they deserve. In other words, the negative connotations typically associated with 'just deserts' is not necessarily clear in the way the item was phrased. Therefore, this item was excluded from the analyses presented here but we also report results with this item included in the supplemental materials.

Did a Three-Factor Model Predict Economic Game Behavior?

We regressed subjects' propensities to defect and their one-shot economic game behavior (TG, DG, or PTPG) on each of the three forgiveness factors (avoidance, revenge, and benevolence). We report standardized coefficients here and in Table 3a.

[3] Although previous one-factor models used a Rasch model for graded responses, we used a less restrictive graded response model (Samejima, 1969) for comparison, which does not constrain all item loadings and threshold parameters to be equal to each other.

Propensity to defect in the final ten rounds of the PDG was predicted by avoidance, b = -.409, se = .097, $p < .001$, but not by revenge, b = -.109, se = .073, $p = .134$, or by benevolence, b = -.011, se = .131, $p = .932$. Amount transferred in the TG was predicted by benevolence, b = .479, se = .181, $p = .008$, but not by avoidance, b = -.048, se = .130, $p = .712$, or by revenge, b = -.121, se = .107, $p = .258$. Amount sent in the DG was not predicted by any of the three factors (avoidance: b = 240, se = .186, $p = .198$; revenge: b = .157, se = .124, $p = .207$; benevolence: b = .004, se = .241, $p = .985$). Finally, the amount removed in the PTPG was predicted by revenge, b = -.462, se = .096, $p < .001$, but not by avoidance, b = -.074, se = .141, $p = .599$, or by benevolence, b = .157, se = .176, $p = .373$.

Overall, one could argue that the three-factor model excels at determining which motivations are underlying different types of experimental-economic behavior. However, three distinct factors are unable to capture whether a common process is underlying some aspect of these behaviors. Moreover, the three-factor model leads to ambiguity in interpreting why none of the factors predicted transfers in the DG: this result could indicate either that *none* of the three motives are related to transfers in the DG or that *all* of the motives relate to DG transfers but explain overlapping variance. Therefore, we compared these results to those from a bifactor (S-1) model, which we used to extract a general factor (forgiveness, scaled by avoidance) as well as specific factors for benevolence and revenge.

Did a Bifactor (S-1) Model Predict Economic Game Behavior?

We specified the bifactor (S-1) model with the subset of avoidance items as the reference domain. The general factor was scaled using the item ‘I would not trust

him/her' (reverse-scored), the specific domain for benevolence was scaled using the item 'I would have good will for him/her', and the specific domain for revenge was identified with 'I would want to seek revenge' (reverse-scored).

We then regressed subjects' propensities to defect and their one-shot economic game behavior (TG, DG, or PTPG, depending on their randomized condition) on the general forgiveness factor (as scaled by reverse-scored 'avoidance', making positive values represent greater forgiveness) and two specific factors (reverse-scored revenge and benevolence) in the bifactor (S-1) model. We report standardized coefficients here and in Table 3a. Propensity to defect in the PDG was predicted by the general factor, b = -.493, se = .035, $p < .001$, but not by the specific factors representing revenge (reverse-scored), b = -.080, se = .056, $p = .152$, or benevolence, b = -.011, se = .058, $p = .853$. Amount transferred in the TG was predicted by the general factor, b = 0.293, se = .051, $p < .001$, and by the benevolence specific factor, b = .253, se = .084, $p = .003$, but not by the revenge specific factor, b = -.110, se = .083, $p = .185$. Amount sent in the DG was predicted by the general factor, b = .350, se = .049, $p < .001$, but not by either the revenge specific factor, b = .161, se = .096, $p = .093$, or the benevolence specific factor, b = -.061, se = .085, $p = .475$. Finally, the amount paid to punish in the PTPG was predicted by the general factor, b = -.218, se = .056, $p < .001$, and by the revenge specific factor, b = -.343, se = .074, $p < .001$, but not by the benevolence specific factor, b = .027, se = .078, $p = .729$.

Here, we see that the general factor from the bifactor (S-1) model predicted meaningful variation in all of our behavioral outcomes, whereas the factors from the correlated model performed less consistently. Also, none of the factors from the

488 correlated model were able to explain meaningful variation in Dictator Game transfers;
489 not because the factors were unrelated, but because the correlated factors explained
490 overlapping variation in that outcome, which was made apparent by the general factor's
491 success in predicting DG transfers in the bifactor (S-1) model.

492 Finally, we correlated the general factor with a score derived by simply taking the
493 mean of all 17 items from the TRIM-NCO. The two measures were highly correlated, r =
494 .946, $p < .001$. Unsurprisingly, this simple 17-item composite was also significantly
495 correlated with scores on each of the four experimental economics games (see Table
496 5a). Jointly, these results suggest that the general factor can be estimated with
497 reasonable fidelity simply by taking the mean of all 17 items on the TRIM-NCO.

498 *Did Item Responses Support a Continuum from Malevolence to Benevolence on the*
499 *General Factor?*

500 To better understand how the general factor corresponds to our
501 conceptualization of forgiveness, we computed location indices (LIs) for each item
502 based on the item response function (LI_{IRF}; Ali, Chang, & Anderson, 2015). The LI_{IRF}
503 value represents the difficulty of a polytomous item (i.e., the probability of endorsing a
504 higher value on the scale) for a person with a specified factor score on the latent
505 construct (we based LI_{IRF} values on a latent score of 0, but the pattern of findings would
506 hold across any chosen latent score). Although one could summarize each item's
507 difficulty by computing the mean or median threshold for each item, the LI_{IRF} tends to
508 characterize item difficulty better than a crude central tendency measure because it
509 integrates information from the item's loading and all of its thresholds. We present LI_{IRF}
510 values in Table 5a, ordered from lowest (i.e., 'easiest') to highest (i.e., 'hardest'). As

shown in the table, the LI_{IRF} values are arrayed along the general factor with revenge items on the easier end and a mixture of avoidance and benevolence items on the more difficult end, suggesting that a motivational or attitudinal continuum that stretches from malevolence to benevolence underlies the general factor. We interpret this descriptive pattern to suggest that people with very low scores on the general factor can renounce revenge whereas they require higher scores on the general factor to endorse approach and benevolence.

STUDY 1 DISCUSSION

For Study 1, we collected a large online sample to test the relative fit and predictive utility of alternative TRIM-NCO models. Based on our results, we found that the three-factor and bifactor models of the TRIM-NCO fit best. Therefore, we used the results of these two measurement models to predict outcomes from four experimental economics games that measure forgiveness-relevant social motivations (Billingsley & Losin, 2017): an iterated prisoner's dilemma game and three one-shot economic games (trust game, dictator game, and pay-to-punish game). The three factors that are traditionally used to account for the covariances among the items on the TRIM (revenge, avoidance, and benevolence) were each significantly associated with scores on one of the four experimental economic games we examined. However, the general factor we derived from the bifactor (S-1) model of those covariances was significantly associated with scores on all four economic games, and in the directions one would expect if that general factor did indeed reflect a latent attitudinal continuum that runs from antagonism to friendliness, or malevolence to benevolence). The group factors representing variance specifically attributable to the revenge items and the benevolence

items predicted significant amounts of unique variance in two of the experimental economics games we examined, which suggests that they may possess incremental validity as measures of forgiveness. However, the consistent usefulness of the general factor for predicting all four behavioral outcomes suggests that it may be the factor that most reliably uncovers the behaviorally relevant motivational changes underlying forgiveness. We also found that this general factor can be dependably estimated simply by taking the mean of all 17 of the items on the TRIM-NCO.

STUDY 2

Researchers' inability to control data collection conditions in online samples has caused much concern among researchers (Buchanan & Scofield, 2018; Buhrmester, Kwang, & Gosling, 2011; Kan & Drummey, 2018). For example, scholars have noticed that substantial numbers of online research subjects provide false responses or find ways of participating multiple times, thereby clouding our understanding of the effects the wish to study (Buchanan & Scofield, 2018; Kan & Drummey, 2018). To ensure that Study 1's results were generalizable beyond online samples, we conducted Study 2 using a typical laboratory setting at a university in the United States.

STUDY 2 METHOD

Subjects

To maximize statistical power and access to subjects in Study 2, we recruited subjects from two sources. First, we recruited 342 students from the undergraduate psychology subject pool at a large university in the southeastern United States (Sex not collected; Age: $M = 19.41$ years; $SD = 4.77$). Although subjects initially believed they could earn up to $10 (implying the possibility of earning less), all subjects received the

full $10 in addition to partial course credit for participating. Second, we recruited 79 subjects from the surrounding community via email and Craigslist. Subjects from the community sample were on average 36.20 (*SD* = 11.95) years old. All community subjects received a $20 show-up fee. In addition, these subjects initially believed they would earn a bonus up to $11 (depending on their decisions in the study), although all community subjects received the full $11, resulting in a total of $31 in compensation for their participation. As in Study 1, we excluded subjects who indicated some level of suspicion (*N* = 95). We also excluded responses from one of the conditions (see procedural differences below for details), for a final sample of 228, with analyses of all subjects available in the supplemental materials.

Procedure

For Study 2, the student and community samples completed procedures similar to those of Study 1. Upon arriving at the laboratory, subjects were seated at a computer. After subjects provided informed consent, a researcher instructed them that they would be playing economic games with an anonymous partner located elsewhere on campus via a computer network. The remainder of the experiment was conducted via computer using the E-Prime 2.0 software (Psychology Software Tools, 2012).

Procedural Differences Between Study 2 and Study 1

In addition to its laboratory setting and its implementation in E-Prime (vs. SoPHIE), Study 2 diverged from Study 1 by including a No-Transgression control condition (*N* = 98). Cases from the No-Transgression control condition were excluded from analyses presented here because we were interested in understanding the utility of

579 the TRIM-NCO when a transgression occurred. Analyses including the no-transgression
580 condition appear in the companion to this manuscript (Billingsley et al., in prep).
581 Study 2 also diverged from Study 1 in that we had subjects engage in only one
582 the economic games we examined in Study 1 (the Trust Game). We chose the Trust
583 Game for its face-valid relevance to a continued interaction, whereas the other two
584 games were one-sided interactions.

585 STUDY 2 RESULTS

586 We followed the same analysis procedures as we did in Study 1. Model fit
587 statistics for the alternative factor structures are available in Table 2a. As in Study 1, the
588 three-factor and bifactor (S-1) models fit best, so we proceeded to evaluate the
589 predictive utility of the two modeling approaches. Model specification procedures were
590 identical to those of Study 1. Also consistent with Study 1, we removed the item, 'I hope
591 he/she gets what he/she deserves,' from the analyses reported here.

592 *Did a Three-Factor Model Predict Economic Game Behavior?*

593 We regressed subjects' propensities to defect and their Trust Game behavior on
594 each of the three forgiveness factors (avoidance, revenge, and benevolence). We report
595 standardized coefficients here, as well as in Table 3a.

596 Propensity to defect in the prisoner's dilemma game was not predicted by any of
597 the three factors (avoidance: b = -.250, se = .133, p = .060; revenge: b = -.154, se =
598 .122, p = .209; benevolence: b = -.067, se = .134, p = .617). Amount transferred in the
599 trust game was predicted by (reverse-scored) revenge, b = -.287, se = .107, p = .007,
600 and benevolence, b = .358, se = .117, p = .002, but not by avoidance, b = .187, se =

.127, p =.140. These results are in contrast to Study 1, in which defections in the PDG were predicted by avoidance and transfers in the TG were predicted by benevolence.

Did a Bifactor (S-1) Model Predict Economic Game Behavior?

We then regressed subjects' propensities to defect and their Trust Game behavior on the general forgiveness factor and the two specific factors (revenge and benevolence) in the bifactor (S-1) model.

Propensity to defect in the PDG was predicted by the general factor, b = -.403, se = .062, p < .001, but not by the specific factors representing revenge, b = -.101, se = .092, p = .274, or benevolence, b = -.056, se = .094, p = .554. Amount transferred in the TG was predicted by the general factor, b = .263, se = .064, p < .001, by the revenge specific factor (reverse-scored), b = -.235, se = .078, p = .002, and by the benevolence specific factor, b = .224, se = .081, p = .005. As in Study 1, these results suggest that the general factor from the bifactor (S-1) model captures behaviorally relevant variance in people's regard for someone who has recently harmed them.

Finally, we correlated the general factor with a score derived from simply taking the mean of all 17 items from the TRIM-NCO. As in Study 1, the two measures were highly correlated, r = .937, p < .001. Unsurprisingly, this simple 17-item composite was also significantly correlated with scores in the Trust Game (see Table 4a). Jointly, these results suggest that the general factor can be estimated with reasonable fidelity simply by taking the mean of all 17 items on the TRIM-NCO.

Did Item Responses Support a Continuum from Malevolence to Benevolence on the General Factor?

As in Study 1, we computed item difficulty values using Ali et al.'s (2015) LI_{IRF} method, which are presented in Table 5a. Again, we found that LI_{IRF} values were sorted along the general factor with revenge items on the easier end and a mixture of avoidance and benevolence items on the more difficult end. Further, LI_{IRF} values for each item were strongly correlated with those found in Study 1, r =.85, as were the rankings themselves, r = .61, indicating consistency across samples and providing further support that the general factor represents an attitudinal continuum that ranges from malevolence (revenge motivation) to benevolence.

STUDY 2 DISCUSSION

Here, we largely replicated the results found in Study 1, in that the general factor and the group factors from the bifactor model consistently predicted subjects' behavior in the prisoner's dilemma and the Trust Game. We note that the three-factor correlated model yielded little consistency across Studies 1 and 2. This is in contrast to the pattern observed using the bifactor (S-1) modeling approach: The general factor predicted scores on both of the games studied, and the revenge and benevolence factors both predicted unique variance in subjects' TG transfers (but not in their PDG scores). Oddly, however, the correlation between the (reverse-scored) revenge-specific factor and trust game scores was negative rather than positive—subjects who were more vengeful than their scores on the general factor (scaled by avoidance) would indicate sent more money in the trust game. It is tempting to attribute the negative relationship between two prosocial constructs to the vagaries of sampling error variance and to conclude that it casts suspicion upon the validity of the group factors as measures of forgiveness.

We also found that the general factor can be dependably estimated simply by taking the mean of all 17 of the items on the TRIM-NCO.

STUDY 3

Psychology has faced pointed criticisms for over-reliance upon samples of Western undergraduates. The material wealth, extensive education, democratic values, and highly industrialized backgrounds of these students relative to much of the world call into question the representativeness of results obtained using such samples (Henrich, Heine, & Norenzayan, 2010). Although Study 1 draws from an adult rather than an undergraduate sample, and Study 2 includes community members as well as students, our results thus far demonstrate the utility of the bifactor (S-1) model for the TRIM-NCO measure in a Western—specifically, American—context. To increase the generalizability of our results, we attempted to replicate our findings in Japan, a culture known to differ significantly from that of the United States along several dimensions pertinent to forgiveness, notably relational mobility and collectivism vs. individualism (Kashima et al., 1995; Schug, Yuki, Horikawa, & Takemura, 2009; Yamagishi & Yamagishi, 1994). If successful, our efforts would furnish researchers with preliminary evidence that the usefulness of the bifactor (S-1) for analyzing TRIM data extends beyond the geographical borders of the United States and beyond the linguistic boundaries of the English language. Such efforts would provide researchers from diverse cultures—perhaps especially those cultures where a translated version of the TRIM-18 is already in use—with a basis for modeling forgiveness as prosocial change that takes place on a latent attitudinal continuum that runs from hostility to friendliness.

STUDY 3 METHOD

Subjects

300 Japanese citizens were recruited using a Japanese crowdsourcing service, Lancers, Inc (65% female; Age: $M = 36.57$; $SD = 10.12$). As in the previous experiments, we excluded subjects who indicated any level of suspicion ($N = 114$) before conducting analyses, resulting in a final sample of $N = 186$. Results with all subjects included are available in the supplementary materials.

Procedures

As in Study 1, Study 3 was programmed using SoPHIE (Hendriks, 2012). All procedures were identical between Studies 1 and 3, except that subjects played only the Trust Game following the Prisoner's Dilemma Game, much like we did in Study 2, with endowments identical to those in Study 1. Study 3 also differed from Study 1 in how the chat was implemented. Specifically, instead of sending and receiving five open-ended chat messages, subjects simply typed the Japanese word for 'hello' and received a commensurate 'hello' from the other subjects. We made this change because Japanese subjects had difficulty chatting with a stranger under the constraints of the chat function, which prevented them from moving forward with the study.

Translation

A Japanese version of the TRIM-18 translation already existed (Ohtsubo, Yamaura, & Yagi, 2015), so we modified this existing translation to make the items more applicable to non-close others, just as we did with the existing English version for Studies 1 and 2.

STUDY 3 RESULTS

690 We followed the same analysis procedures as we did in Studies 1 and 2. Model
691 fit statistics for the alternative factor structures are available in Table 2a. As in Study 1,
692 the three-factor and bifactor (S-1) models fit best, so we proceeded to evaluate the
693 predictive utility of the two modeling approaches. Model specification procedures were
694 identical to those of Studies 1 and 2. Also consistent with Studies 1 and 2, we removed
695 the item, 'I hope he/she gets what he/she deserves,' from the analyses reported here.
696 *Did a Three-Factor Model Predict Economic Game Behavior?*
697 We regressed subjects' propensities to defect and their one-shot economic game
698 behavior in the trust game on each of the three forgiveness factors (avoidance,
699 revenge, and benevolence). We report standardized coefficients here and in Table 3a.
700 Propensity to defect in the PDG was predicted by benevolence, b = -.441, se =
701 .178, p = .013, but not by avoidance, b = -.099, se = .178, p = .580, or revenge, b =
702 .178, se = .155, p = .250. Amounts transferred in the TG were not predicted by any of
703 the three factors (avoidance: b = -.030, se = .158, p = .850; revenge: b = .102, se =
704 .128, p = .426; benevolence: b = -.007, se = .159, p = .965).
705 *Did a Bifactor (S-1) Model Predict Economic Game Behavior?*
706 We then regressed subjects' propensity to defect in the prisoner's dilemma game
707 and their scores on the trust game on the general forgiveness factor and two specific
708 factors (revenge and benevolence) in the bifactor (S-1) model.
709 Propensity to defect in the PDG was significantly associated with the general
710 factor, b = -.379, se = .066, p < .001, but not with the specific factors representing
711 revenge (reverse-scored), b = .139, se = .120, p = .244, or benevolence, b = -.148, se =
712 .097, p = .126. As was the case with the standard three-factor model for the TRIM,

amounts transferred in the trust game were not significantly associated with the general factor (b = .030, se = .070, p = .669), the revenge specific factor (b = .077, se = .100, p = .442) or the benevolence specific factor: b = -.002, se = .091, p = .986). Although we found that the general factor was able to predict subjects' propensities to defect in the PDG, we were unable to explain any of the variation in TG transfers with any of the factors from either the correlated or bifactor (S-1) models.

Finally, we correlated the general factor with a score derived from simply taking the mean of all 17 items from the TRIM-NCO. The two measures were highly correlated, r = .940, p < .001. Unsurprisingly, the simple 17-item composite was also significantly correlated with scores on the prisoner's dilemma game, but not with scores on the trust game (see Table 4a). Jointly, these results suggest that the general factor can be estimated with high fidelity with the mean of all 17 items on the TRIM-NCO.

Did Item Responses Support a Continuum from Malevolence to Benevolence on the General Factor?

As in Studies 1 and 2, we computed LI_{IRF} for each item (Ali et al., 2015), which are presented in Table 5a. Again, we found that LI_{IRF} values were sorted along the general factor with revenge items on the easier end and a mixture of avoidance and benevolence items on the more difficult end. We also found that LI_{IRF} values and LI_{IRF} ranks for each item were strongly correlated with those found in Study 1 (r =.923 and r = .838, respectively) and those found in Study 2 (r =.757 and r = .600, respectively), providing further evidence for consistency across samples and for the interpretation that the general factor represents a malevolence-benevolence continuum.

STUDY 3 DISCUSSION

In Study 3, we obtained similar results regarding the factor structure of the TRIM-NCO Inventory, with the bifactor model and three factor models exhibiting good model fit. As was the case in Studies 1 and 2, the general factor from the bifactor (S – 1) model was associated with scores on the prisoner's dilemma game. So, too, was one of the subscales that resulted from a traditional scoring of the TRIM. None of the factors derived from the standard three-factor model or the bifactor (S – 1) model predicted scores in the Trust Game. Because the general factor was significantly associated with scores on every other experimental economics game from all three experiments, the non-significant association of forgiveness with scores on the trust game may reflect a true cultural difference rather than the result of sampling error. Finally, a mean of all 17 TRIM items performed very much like the general factor, suggesting that a simple composite of all 17 items is a reasonable way to estimate the general factor.

STUDY 4

In Studies 1-3, we probed the hypothesis that a single dimension spanning malevolence and benevolence underlies forgiveness. We used the TRIM-NCO, a new version of widely-used TRIM-18 (McCullough, Cohen, & Root, 2006) that was modified to assess interpersonal motivations between individuals encountering one another for the first time. However, forgiveness often occurs in close interpersonal relationships rather than in first-time, anonymous encounters, with great significance for our personal lives and well-being (Rusbult, Hannon, Stocker, & Finkel, 2005).

To expand our findings beyond first-time interactions between strangers, we therefore report the results of a fourth study, in which we applied the bifactor (S-1) modelling technique to the more traditional 18-item Transgression Related Interpersonal

Motivations Inventory (TRIM-18; McCullough, Cohen, & Root, 2006), which was designed to measure forgiveness following a transgression committed by a familiar person in a real-world setting. In addition to expanding on Studies 1-3 by assessing forgiveness in close relationships, we also sought to generalize our findings to self-report instruments beyond variants of the TRIM. To do so, we used subjects' responses from the Decision to Forgive and Emotional Forgiveness Scales (Hook et al., 2009; Hook et al., 2012). These additional scales allowed us to determine whether a general factor reflects a general propensity to forgive that suffuses the items from other measures as well.

STUDY 4 METHOD

Subjects

Subjects were 168 undergraduate students recruited from introductory psychology classes at a Southeastern University in the United States (distinct from that of Study 2; 59.5% female; Age not collected). Subjects who were of East Asian nationality were excluded so as not to confound a broader cross-national project design (see Study 5). Subjects completed the survey outside of the lab in exchange for course credit. Unlike Studies 1-3, study 4 utilized a correlational rather than experimental design. As a result, our experiment did not involve deception, we did not probe subjects for suspicion, and no subjects were excluded from analyses.

Procedure

All data were collected using an online Qualtrics survey with standardized instructions embedded in the survey. To study forgiveness in real-world transgressions, we instructed subjects to "[t]hink of a time that a close other person did something to

upset you, hurt you, or otherwise commit an offense that caused a rift in your relationship." To increase the salience of the memory, we asked subjects to describe the context and outcome of the offense. Immediately after describing the transgression, subjects completed several self-report measures, including the TRIM-18, Decision to Forgive Scale, and Emotional Forgiveness Scale. Subjects also completed other measures beyond the scope of this paper.

Measures

Decisional Forgiveness

The Decision to Forgive Scale (DFS) measures the degree to which the victim of a transgression deliberately works to replace negative behavior towards a transgressor with positive, prosocial behavior (Hook et al., 2009; Hook et al., 2012). Subjects were instructed as follows: "Think of your current intentions toward the person who hurt you. Indicate the degree to which you agree or disagree with the following statements." For example, subjects were asked how much they agree with the statement, "I will not talk with him or her." The DFS is an 8-item scale with response options ranging from 1 ("Strongly Disagree") to 5 ("Strongly Agree"; see Appendix C). We scored subjects' DFS responses so that higher scale scores indicated greater decisional forgiveness.

Emotional Forgiveness

The Emotional Forgiveness Scale (EFS) measures the degree to which a victim replaces negative emotions towards a transgressor (e.g., anger) with positive emotions (e,g., compassion; Hook et al., 2009). Subjects were instructed as follows: "Think of your current emotions toward the person who hurt you. Indicate the degree to which you agree or disagree with the following statements." For example, subjects were asked

how much they agree with the statement, “I no longer feel upset when I think of him/her.” The EFS is an 8-item scale with response options ranging from 1 (“Strongly Disagree”) to 5 (“Strongly Agree”; see Appendix D). We scored the EFS so that higher scale scores indicated greater emotional forgiveness.

Transgression-Related Interpersonal Motivation Scale (Close Others)

Forgiveness motivations were measured using the 18-item Transgression-Related Interpersonal Motivation Scale for close others (TRIM-18; McCullough, Root, & Cohen, 2006). Subjects were instructed as follows: “For the following questions, please indicate your current thoughts and feelings about the person who hurt you; that is, we want to know how you feel about that person right now. Next to each item, circle the number that best describes your current thoughts and feelings.” Responses ranged from 1 (“Strongly Disagree”) to 5 (“Strongly Agree”). Consistent with Studies 1-3, negatively worded items (i.e., items indicating less forgiveness) were reverse scored so that higher scores indicated greater forgiveness. For example, strong agreement with the items “I will make him or her pay” from the revenge scale and “I withdraw from him/her” from the avoidance scale contributed to lower scores on the TRIM-18. In contrast, strong agreement with the item “Despite what he/she did, I want us to have a positive relationship” from the benevolence scale contributed to a higher score.

Single-item measure of forgiveness

We also used a single item, “Have you forgiven the person for the offense?” as a criterion measure of forgiveness. Subjects responded using a slider scale with response options ranging from 0 and 100. Higher scores indicated greater forgiveness.

STUDY 4 RESULTS

828 Analyses were conducted using Mplus version 7 (Muthén & Muthén, 1998-2012).
829 See supplementary materials for data and syntax. We report standardized coefficients
830 for all outcomes.
831 *Factor Analysis of the TRIM-18*
832 As in Studies 1-3, we fit three confirmatory correlated models for the TRIM-18
833 using one-factor, two-factor (avoidance-benevolence and revenge), and three-factor
834 (avoidance, benevolence, and revenge) models. Model fit for the factor structures can
835 be seen in Table 2b. Consistent with previous research, we found that the fit of the two-
836 factor model was comparable to that of the three-factor model, which is unsurprising
837 given that the avoidance and benevolence factors in the three-factor model were
838 correlated at $r = .915$. In contrast to what we found with the TRIM-NCO, which
839 consistently favored a model with three correlated factors, we thus found that people's
840 responses to items on the TRIM-18 may be informed by only two underlying constructs.
841 To better understand the implications of the relative interchangeability of the two-
842 factor and three-factor models for an understanding of forgiveness based on the bifactor
843 model, we created two alternative bifactor (S-1) models: For the first model, we created
844 only one specific factor for the revenge items, with the general factor scaled by the
845 avoidance and benevolence items (referred to as a bifactor (2-1) model). For the
846 second model, we created two specific factors—one for the revenge items and another
847 for the benevolence items—with the general factor scaled by the avoidance items (as in
848 Studies 1-3; referred to as a bifactor (3-1) model). As mentioned above, bifactor (S-1)
849 models and an S-correlated factors model (where S represents the number of factors in
850 the comparable models) yield identical model fit under reasonable restrictions (Geiser et

al., 2008); therefore, these models were compared for their predictive utility (with a single self-report item of forgiveness as our criterion) against the two correlated factors and three correlated factors models, respectively. Path coefficients can be seen in Table 3b.

Did the Two- and Three-Factor Models Predict Single-Item Forgiveness?

We regressed the single-item measure of forgiveness on each of the factors in the two- (avoidance-benevolence and revenge) and three-factor (avoidance, revenge, and benevolence) models of the TRIM-18. In the two-factor model, the single-item forgiveness measure was predicted by avoidance-benevolence, $b = 0.450$, $SE = 0.074$, $p < .001$, and revenge, $b = 0.232$, $SE = 0.086$, $p < .001$. In the three-factor model, the single-item forgiveness measure was not predicted by avoidance, b = .072, se = .187, $p = .699$, but was predicted by benevolence, $b = 0.382$, $SE = 0.167$, $p = .022$, and revenge, $b = 0.250$, $SE = 0.091$, $p = .006$.

The fact that the avoidance factor in the three-factor model did not predict unique variance in the single-item forgiveness measure lends additional support to our speculation that avoidance and benevolence are indistinguishable in close relationships. In fact, a three-factor model leads to the problematic conclusion that avoidance motivation is unrelated to self-reported forgiveness after accounting for revenge and benevolence motivations, which is not surprising in light of the high collinearity between the avoidance and benevolence factors in the three-factor model.

Did the Bifactor (S-1) Models Predict Single-Item Forgiveness?

As mentioned previously, we specified two bifactor (S-1) models: The first with a specific factor only for the revenge items (bifactor (2-1) model) and the second with

specific factors for both the revenge and the benevolence items (bifactor (3-1) model). For both of these models, we regressed the single-item measure of forgiveness on the general forgiveness factor and the specified group factor or factors. In the bifactor (2-1) model (with a specific factor only for the revenge items), scores on the single-item forgiveness measure were significantly predicted by the general factor, b = .561, se = .061, $p < .001$, and the revenge specific factor, b = -.208, se = .073, $p = .004$. The direction of the relationship between the revenge factor and the single-item measure of forgiveness was troubling, however, because the revenge items here are reverse scored; thus, the negatively signed regression coefficient suggests that disavowal of revenge motivation is associated with lower scores on the single-item measure of forgiveness. In the bifactor (3-1) model (with specific factors for both the revenge items and the benevolence items), single-item forgiveness responses were significantly predicted by the general factor, $b = 0.582$, $SE = 0.060$, $p < .001$, but not the revenge factor, $b = 0.144$, $SE = 0.113$, $p = .202$, or the benevolence factor, $b = -0.044$, $SE = 0.103$, $p = .666$. The significant negative relationship of (disavowals of) revenge motivation with the single-item measure of forgiveness in the bifactor (2-1) model therefore disappeared once the group factor for the benevolence items was also included.

Next, we correlated the general factor with a score derived by simply taking the mean of all 18 items from the TRIM-18. The two measures were almost perfectly correlated, $r = .978$, $p < .001$. Unsurprisingly, this simple 18-item composite was also significantly correlated with people's responses to the single-item forgiveness measure, $r = .586$, $p < .001$. As was the case with the TRIM-NCO in Studies 1-3, these results

suggest that the general factor can be estimated with high fidelity simply by taking the mean of all of 18 items on the TRIM-18.

Did Item Responses Support a Continuum from Malevolence to Benevolence on the General Factor?

As in Studies 1-3, we computed LI_{IRF} for each item from the general factor (Ali et al., 2015), which are presented in Table 5b. Again, we found that LI_{IRF} values were sorted along the general factor with revenge items on the easier end and a mixture of avoidance and benevolence items on the more difficult end. We also found that LI_{IRF} values and LI_{IRF} ranks for each item were strongly correlated with those found in Study 1 (r =.861 and r = .641, respectively), Study 2 (r =.719 and r = .741, respectively), and Study 3 (r =.906 and r = .520, respectively), providing further evidence for consistency in item difficulties across samples, as well as for the interpretation that the general factor represents a continuum from revenge to avoidance and benevolence.

Is the general factor of forgiveness unique to the TRIM Inventory or does it apply to other measures of forgiveness?

Next, we sought to determine whether the single malevolence-benevolence factor that we hypothesize to underlie forgiveness also explains the pattern of item responses on the Decision to Forgive Scale and the Emotional Forgiveness Scale (Hook et al., 2009; Hook et al., 2012). To do so, we fit a bifactor (S-1) model to the 18 items from the TRIM Inventory (McCullough et al., 2006), the eight items from Hook and colleague's Decision to Forgive Scale, and the eight items from the Emotional Forgiveness Scale (Hook et al., 2009; Hook et al., 2012). The bifactor (S-1) model included four group factors: two for the revenge and benevolence items from the TRIM

Inventory, and two for the items on the Decision to Forgive Scale and the Emotional Forgiveness Scale, respectively. The model exhibited good model fit (see Table 2b), with the general factor explaining common variance across all items in the TRIM Inventory, the Decision to Forgive Scale, and the Emotional Forgiveness Scale.

We then regressed the single-item measure of forgiveness on the general factor from this newly created bifactor (S-1) model. As we found with simply modeling responses from the TRIM-18, the general factor remained a strong predictor of the single-item measure of forgiveness, b = .592, se = .059, $p < .001$. The relationship of the general factor and the single-item measure of forgiveness changed very little in magnitude when we included the Decisional and Emotional Forgiveness scales in the bifactor models ($\Delta b = .031$), indicating that the general factor neither gained nor lost substantial precision by including responses from additional scales. The fact that the general factor estimated exclusively on the basis of the TRIM-18 operates essentially identically to a general factor that is estimated by also including other forgiveness scales stands as evidence that the general factor does in fact reflect subjects' standing on an attitudinal or motivational continuum that exists independently of any specific tool used to measure it.

STUDY 4 DISCUSSION

In Study 4, we used a sample of non-East Asian U.S. undergraduates to determine if our findings extended to relationships involving close others. We largely replicated the results of Studies 1-3, such that a bifactor (S-1) model of forgiveness provided good model fit across the context of close relationships, and model fit for the bifactor (S-1) remained excellent even as we added items from related scales designed

to tap emotional and decisional components of forgiveness. The latter result suggests that the same underlying malevolence-benevolence continuum suffuses the items from a second self-report measure of forgiveness. Moreover, the general factor of the bifactor models predicted a single-item measure of forgiveness, even in models that incorporated self-report measure of forgiveness beyond the TRIM. When we analyzed the three-factor model of forgiveness, the benevolence and revenge factors predicted the single-item measure of forgiveness, but avoidance did not. In Study 4, our analyses of item responses likewise replicated Studies 1 through 3 in suggesting the possibility that a single motivational continuum underlies the forgiveness process, ranging from malevolence to benevolence. In contrast to Studies 1 through 3, however, factor analyses from Study 4 did not reveal significant differences in fit between the two-factor and three-factor models when using the correlated-factors approach. Given that Studies 1 through 3 involved strangers interacting for the first time, whereas Study 4 involved close others, this result may have interesting implications for our understanding of how forgiveness operates across different relationship types—a topic to which we return in the General Discussion.

STUDY 5

Through these experiments, we have sought to replicate our results in different experimental contexts to ensure that our results are generalizable beyond a single sampling procedure or study design. In Study 5, we therefore sought to replicate the results of Study 4—with its emphasis on forgiveness in close relationships—in a sample of Japanese undergraduate students, which we expected would yield consistent results with our U.S. samples, just as Study 3 broadly replicated Studies 1 and 2.

STUDY 5 METHOD

Subjects

Subjects were 158 undergraduate students recruited using the psychology subject pool at a university in Japan (51.2% female; Age not collected). East Asian subjects were targeted for recruitment, resulting in an ethnically homogenous sample. Subjects completed the survey in exchange for 700 JPY.

Procedure

The procedure in Study 5 was highly similar to Study 4. Subjects completed back-translated versions of the scales in Study 4 (Ohtsubo et al., 2015). The only procedural deviation was in subject recruitment, as subjects completed the survey in a laboratory setting in order to receive compensation for their participation.

STUDY 5 RESULTS

Analyses were conducted using Mplus version 7 (Muthén & Muthén, 1998-2012). See supplementary materials for data and syntax. We report standardized coefficients for all outcomes.

Factor Analysis of the TRIM-18

As in Study 4, we fit three confirmatory models for the TRIM-18 using one-factor, two-factor (avoidance-benevolence and revenge), and three-factor (avoidance, benevolence, and revenge) models. Model fit for the factor structures can be seen in Table 2b. Consistent with Study 4, we found that the two-factor and three-factor models fit very comparably, which again is unsurprising in light of the fact that the avoidance and benevolence factors in the three-factor model were correlated at $r = .919$.

To be consistent with Study 4, we created two alternative bifactor (S-1) models: For the first model, we created only one specific factor—for the revenge items—with the general factor scaled by the avoidance and benevolence items. For the second model, we created two specific factors—one for the revenge items and another for the benevolence items—with the general factor scaled by the avoidance items (as in Studies 1-3). These models were then compared for their predictive utility (with a single self-report item of forgiveness as our criterion) against the two correlated factors and three correlated factors models, respectively. Path coefficients can be seen in Table 3b.

Did the Two- and Three-Factor Models Predict Single-Item Forgiveness?

We regressed the single-item measure of forgiveness on each of the factors in the two-factor (avoidance-benevolence and revenge) and three-factor (avoidance, revenge, and benevolence) models of the TRIM-18. In the two-factor model, the single-item forgiveness measure was predicted by avoidance-benevolence, $b = 0.388$, $SE = 0.073$, $p < .001$, and revenge, $b = 0.276$, $SE = 0.070$, $p < .001$. In the three-factor model, the single-item forgiveness measure was negatively predicted by avoidance, b = -.421, se = .175, $p = .016$, and positively by revenge, $b = 0.307$, $SE = 0.067$, $p < .001$, and benevolence, $b = 0.807$, $SE = 0.179$, $p < .001$.

Although these results did not yield the same problematic patterns we found in Study 4, we did find that the three-factor model created a negative relationship between (reverse-scored) avoidance and the single-item measure of forgiveness after accounting for variance in benevolence. This may be more evidence that, when measured in close relationships, the benevolence and avoidance items reflect the same underlying

construct, thereby causing problems when trying to identify their unique influence on forgiveness-relevant constructs.

Did the Bifactor (S-1) Models Predict Single-Item Forgiveness?

As mentioned previously, we specified two bifactor (S-1) models: The first with only one specific factor for the revenge items (bifactor (2-1) model) and the second with specific factors for both the revenge and the benevolence items (bifactor (3-1) model). For both of these models, we regressed the single-item measure of forgiveness on the general forgiveness factor and the respective group factors. In the bifactor (2-1) model (with a specific factor only for the revenge items), scores on the single-item forgiveness measure were significantly predicted by the general factor, b = .532, se = .063, $p < .001$, and the revenge factor, b = .243, se = .055, $p < .001$. Unlike in Study 4, the relationship between the (reverse-scored) revenge factor was positively related to forgiveness, as one would expect. In the bifactor (3-1) model (with specific factors for both the revenge items and the benevolence items), single-item forgiveness responses were significantly predicted by the general factor, $b = 0.539$, $SE = 0.063$, $p < .001$, the revenge factor, $b = 0.198$, $SE = 0.055$, $p < .001$, and the benevolence factor, $b = 0.178$, $SE = 0.055$, $p = .001$.

Next, we correlated the general factor with a score derived from simply taking the mean of all 18 items from the TRIM-18. The two measures were highly correlated, $r = .816$, $p < .001$. Unsurprisingly, this simple 18-item composite was also significantly correlated with people's responses to the single-item forgiveness measure, $r = .589$, $p < .001$. As was the case with the TRIM-NCO in Studies 1-3 and the TRIM-18 in Study 4,

these results suggest that the general factor can be estimated with reasonably high fidelity simply by taking the mean of all of 18 items on the TRIM-18.

Did Item Responses Support a Continuum from Malevolence to Benevolence on the General Factor?

As in Studies 1-3, we computed LI_{IRF} for each item (Ali et al., 2015), which are presented in Table 5b. Again, we found that LI_{IRF} values were sorted along the general factor with revenge items on the easier end and a mixture of avoidance and benevolence items on the more difficult end. We also found that LI_{IRF} values and LI_{IRF} ranks for each item were strongly correlated with those found in Study 1 (r =.768 and r = .525, respectively), in Study 2 (r =.730 and r = .748, respectively), in Study 3 (r =.826 and r = .699, respectively), and in Study 4 (r =.874 and r = .794, respectively), providing further evidence for consistency in item difficulties across samples, as well as for the interpretation that the general factor represents a continuum from malevolence to benevolence.

Does the general factor of forgiveness capture dimensionality in other forgiveness measures?

Next, we sought to determine whether the single malevolence-benevolence factor that we hypothesize to underlie forgiveness also explains the pattern of item responses on other forgiveness measures—specifically here the Decision to Forgive Scale and Emotional Forgiveness Scale (Hook et al., 2009; Hook et al., 2012). As in Study 4, we fit a bifactor (S-1) model to the 18 items from the TRIM Inventory, the eight items from Hook and colleagues' Decision to Forgive Scale, and the eight items Emotional Forgiveness Scale (Hook et al., 2009; Hook et al., 2012). The bifactor (S-1)

model included four group factors: two for the revenge and benevolence items from the TRIM Inventory, and two for the items on the Decision to Forgive Scale and the Emotional Forgiveness Scale, respectively. The model exhibited good model fit (see Table 2b), with the general factor explaining common variance across all items in the TRIM, Decision to Forgive Scale, and Emotional Forgiveness Scale.

We then regressed the single-item measure of forgiveness on the general factor from this newly created bifactor (S-1) model. As we found with modeling responses from the TRIM-18 only, the general factor remained a strong predictor of single-item forgiveness, b = .556, se = .064, $p < .001$. As we found in Study 4, the relationship of the general factor and the single-item measure of forgiveness changed very little in magnitude when we included the Decisional and Emotional Forgiveness scales in the bifactor models ($\Delta b = .017$), which stands as further evidence that the general factor does in fact reflect subjects' standing on an attitudinal or motivational continuum that exists independently of any specific tool used to measure it.

STUDY 5 DISCUSSION

In Study 5, we sampled from Japanese students at a university in Japan to replicate the main results we obtained in Study 4 (which were obtained from non-East Asian students in the U.S.). In concert with the results of Study 4, we found that the bifactor modeling approach yielded good fit, with a general factor that predicted single-item forgiveness scores. As in Study 4, we also found that the general factor captured variance underlying not only the TRIM-18 but also the Decision to Forgive and Emotional Forgiveness Scales, which suggests that the same underlying malevolence-benevolence continuum suffuses the items from a second self-report measure of

forgiveness. Together with similar results from Studies 1-3 in which non-close others were involved, item response analyses suggest that such a motivational or attiudinal continuum underlies forgiveness in both close and non-close relationships. However, as in Study 4, factor analyses using the correlated factors approach found no significant difference in fit between the two-factor and three-factor models. The contrast with the results of Studies 1 through 3 in this regard provides additional grounds to suspect that the structure of forgiveness may differ in close vs. non-close relationships. We address this possibility further in the General Discussion.

GENERAL DISCUSSION

Forgiveness has long been conceptualized as a process of psychological change regarding a harmdoer, but change in what? Through the years, scholars have proposed a variety of theoretical models of the psychological changes that constitute forgiveness (e.g., McCullough, Fincham, & Tsang, 2002; Subkoviak et al., 1995), but the tools they have developed to measure forgiveness have not always succeeded in capturing those complexities and nuances. Generally, the constructs defined by theory often appear to be more complex that the underlying structure of the psychometric instruments designed to capture those theoretical constructs. Here, with the goal of obtaining a better theoretical understanding of the underlying structure of the psychological phenomena that change when people forgive, we tested a variety of modeling approaches for a commonly used measure of forgiveness—the Transgression-Related Interpersonal Motivations (TRIM) Inventory (McCullough et al., 2006)—with separate variants of the scale for close and non-close relationships. In doing so, we specifically sought to determine whether a bifactor modeling approach, which specifies forgiveness

as a single underlying continuum that ranges from malevolence to benevolence, along with other potentially substantive item-specific factors, would provide additional clarity about the psychological dimensions that underlie forgiveness.

In three experiments involving staged transgressions between non-close others, we found that the bifactor model of responses on the newly developed TRIM-NCO yielded substantially better model-data fit than did more traditional single-factor and two-factor confirmatory models. Although the bifactor model and three-factor model fit the data equally well (which is true by definition, under reasonable constraints; Geiser et al., 2008), none of the factors in the three-factor model stood out as robust predictors of any behavioral measure of forgiveness, whereas the general factor of the bifactor model predicted nearly every behavioral outcome we measured. Indeed, every criterion variable that was correlated with at least one factor from the three-factor model was also significantly correlated with the general factor in the bi-factor model—and in the theoretically expected direction (which was not always the case with the factors derived from the three-factor model).

Similarly, in two additional studies involving recalled transgressions between close others, bifactor models of responses to the original TRIM-18 fit the data well, and the general factor of those bifactor models consistently predicted a single-item self-report measure of forgiveness—even when we added other self-report measures of forgiveness (beyond the TRIM-18) into the model. Critically, the general factor that we obtained from a bifactor model of the TRIM-18 was uncovered with near-perfect precision from a bifactor model that also included items from two independently developed measures of forgiveness (Hook et al., 2009, 2012). Across all five studies,

the reliably good fit of the bifactor model, together with the fact that its general factor consistently predicted relevant behavioral or self-reported criterion measures, strongly align with an understanding of forgiveness as a general process of change along a single attitudinal or motivational continuum, perhaps along with one or more group factors that reflect method variance or some other substantive source of forgiveness-relevant variance (for example, personality-based response sets that reflect individual differences in aversion to harming strangers). The psychological meanings and theoretical importance of the group factors was rendered somewhat uncertain by their unreliable associations with the various criterion variables with which we sought to correlate them (for a clear discussion of G and S factors and their meanings in regressions, see Heinrich, Zagorscak, Eid, & Knaevelsrud, 2018).

Our hypothesis that the trait underlying the measures of forgiveness we examined here reflects an attitudinal continuum running from malevolence to benevolence is reinforced by the correlations of the general factor recovered here with subjects' scores on four different laboratory experimental economics games, including games that reflect both punitiveness and cooperativeness. Other work has shown that the variance shared among people's scores in the cooperative games appears to be caused by a common tendency to cooperate that manifests itself both in the lab and in real-life social interactions (McAuliffe et al., 2018; Peysakhovich et al., 2014; Wilhelm, Kaltwasser, & Hildebrandt, 2018), so we are inclined to conclude that the general factor we found here reflects variation in a single broad motivational or attitudinal construct that, on the positive end of the continuum, manifests itself through cooperative behavior in daily life. This conclusion is strengthened considerably by the fact that the item

difficulties of the items on both versions of the TRIM inventory align themselves along a continuum that ranges from motivation to seek vengeance on one end to motivation to restore friendly relations on the other. Because a malevolence-benevolence dimension appears to suffuse interpersonal behavior in general (as exemplified in the Interpersonal Circumplex model of interpersonal behavior; Gurtman, 2009), it is to some extent unsurprising to discover that a similar continuum underlies forgiveness, though it has the potential to be critically important for future theory and research on forgiveness.

Measuring Forgiveness in Close vs. Non-Close Relationships

Prior research involving measures of transgression-related interpersonal motivations has used the TRIM-18 (McCullough et al., 2006) or its earlier 12-item variant (McCullough et al., 1998), which were designed to gauge the forgiveness-relevant motivational changes experienced by individuals in close relationships (e.g., friendships, family relationships, and mating relationships). In our first three experiments, we analyzed a new variant of the measure, the TRIM-NCO, which involved modifying the TRIM-18 to make it applicable to non-close relationships of the sort that researchers frequently examine in experiments. The novelty of the TRIM-NCO, with its focus upon interpersonal motivations in new relationships rather than established ones, raises the concern that the conclusions derived from research that use it might not generalize across relationship types. We addressed this concern by reporting the results of two correlational studies that used the standard TRIM-18 to assess transgression-related motivations in the context of close relationships. For both close and non-close relationships, we found that the bifactor model showed excellent fit and that its general factor reliably exhibited predictive utility.

Although the patterns of results involving the bifactor model were largely consistent across close and non-close relationships, the same cannot be said for results involving the more standard two-factor and three-factor correlated-factors models. In studies involving non-close others (Studies 1-3), the two-factor correlated-factor model (avoidance-benevolence and revenge) fit the data significantly worse than the three-factor model. But in studies involving close others (Studies 4-5), the two-factor model fit the data as well as the three-factor model did—a finding consistent with prior work (McCullough et al., 2007; McCullough et al., 2006). Moreover, in studies involving close others, both factors of the two-factor model accounted for unique variance in our single-item criterion measure of forgiveness, whereas the avoidance factor of the three-factor model did not significantly predict the single item measure after accounting for revenge and benevolence.

Collectively, therefore, the results of our five studies indicate a discrepancy in the number and nature of specific factors that underlie forgiveness in close relationships vs. forgiveness in non-close relationships. The possible causes and implications of this discrepancy invite consideration. Do avoidance and benevolence motives become indistinguishable as relationships become close? Does the conflation of avoidance and benevolence in studies involving close others result from the fact that transgressions involving close others occurred in the more distant past or because transgressions in close relationships may already have been resolved at the time of measurement? Is it because transgressions involving close others are more harmful than those manufactured in experiments involving strangers? These and other possibilities provide the basis for future inquiry into the process of forgiveness, although it appears that most

of these questions could be easily elided by concentrating future theoretical efforts on the general factor (whether estimated explicitly with a bifactor model or, more approximately, with a sum score of all of the items on the scale).

Do Specific Factors Reflect Method or Motive?

We argued from the results of the bifactor (S-1) model of the TRIM Inventory (and in Studies 4 and 5, both the Decisional and Emotional Forgiveness Scales) that the general factor reflects a motivational or attitudinal continuum that spans from malevolence to benevolence. We argued further that the additional group factors reflect either substantive variation due to theoretically meaningful features of forgiveness or methodological factors that might best be characterized as nuisance variance. Although we found some modest correlations between the specific factors and our outcomes, they were not wholly consistent across experiments, possibly undermining their utility beyond improving model-data fit. It is also plausible that the specific factors reflect systematic responses to particular methods (such as positive vs. negative wording) that distinguish items pertaining to revenge and benevolence, which may also undermine their predictive utility. In fact, some treatments of a bifactor model are designed to account for differences in wording (e.g., reverse-scored items) due to inherent differences in how people respond to items that are worded to be the inverse of their construct, and scholars have even noted that these applications may be doing most of their work by accounting for implausible response patterns (Reise, Kim, Mansolf, & Widaman, 2016). Therefore, we still have some important open questions regarding the *predictive* utility of modeling specific factors of forgiveness, specifically with regard to responses on the TRIM Inventory. Again, however, the importance of these findings

seem to us considerably overshadowed by the prominence of the general factor and its broad conceptual and empirical utility.

One possible limitation of this work is that our first three experiments involved only a single behavioral measure for each of the constructs representing trust, benevolence, and punitiveness, which attenuated their reliability. On one hand, this concern may be somewhat minor for several reasons. First, the iterated prisoner's dilemma game evinced extremely high internal consistency (McDonald's ωs > .97; see supplemental materials) in all three studies, and was positively and significantly correlated with the general factor (but not the specific factors) in each of them. Second, previous research has obtained internal consistency estimates for six-item composites of scores on the Dictator Game and the Trust game of that ranged from .91 to .95 for the Dictator Game and .95 to .96 for the Trust Game (McAuliffe et al. 2018b). Applying the Spearman-Brown prophecy formula to the mean of the respective reliability estimates implies that a one-item measure of the Dictator Game could possess a reliability as high as .69 and the Trust Game could possess a reliability as high as .76. These estimates suggest that the one-item constitution of these two measures in our studies might attenuate true score correlations by as little as $1 - \sqrt{.69} = 17\%$ for the Dictator Game and $1 - \sqrt{.76} = 13\%$ for the Trust Game. The accuracy of these prophesied estimates cannot be verified with our data, of course, but they do suggest that attenuations in the magnitude of the associations we reported here are smaller than one might imagine. We also remind readers that in Studies 1-3 the general factor was a robust predictor of most behavioral measures—even the single-item measures—

perhaps because it represents a general propensity to act prosocially, which is in contrast to the more nuanced interpretations of the specific factors (Geiser et al., 2008).

Even so, our understanding of the relationships between the specific factors of the TRIM-NCO and behavioral measures of forgiveness could be improved in future research by assessing forgiveness-relevant cooperative behavior using multiple indicators per construct (e.g., multiple 'benevolence', 'trust', and 'punitive' behavioral assessments). We also note that Studies 4 and 5—which used the TRIM-18 and focused on forgiveness in close relationships—included no behavioral measures, a limitation that future research might also address.

Does the Utility of the General Factor Justify a Single-Factor Model?

Because we were able to gain most of our predictive utility for forgiveness-relevant outcomes from the general factor of the bifactor (S-1) model, we considered how well the general factor from this more complex model compares with composite scores often used in forgiveness research. In each of our studies, we found that the general factor correlated with an average 'TRIM' score to a very high degree (*r*s > 0.80; see supplemental materials), indicating that researchers could retain a great deal of predictive power by considering only a rudimentary model of forgiveness. Although we do not recommend this practice, we do think researchers could justify their use of a simple model of TRIM scores *because* we demonstrated that the general factor of the bifactor (S-1) is effective at representing the underlying malevolence-benevolence continuum in which forgiveness evidently takes place. For researchers interested in applying the bifactor (S-1) model that we advocate here, we make the relevant software code available in supplemental materials.

Conclusions

Although researchers have been studying forgiveness for more than 25 years, the rich and sometimes multifaceted conceptualizations of forgiveness that have guided this work have not been tightly moored to the psychometric realities of the tools used to measure those conceptualizations. The studies presented here strongly support the hypothesis that forgiveness is a process of attitudinal change, incorporating thoughts, feelings, and behavioral tendencies, along a dimension that ranges from malevolence to benevolence. We recovered evidence for this underlying continuum from studies in laboratory studies as well as in online studies, in the United States as well as in Japan, with transgressors who are strangers as well as with transgressors who are existing relationship partners, and with the TRIM Inventory (McCullough et al., 2006) as well as with the Emotional and Decisional Forgiveness Scales (Hook et al., 2009; Hook et al., 2012). Perhaps most notably, the underlying malevolence-benevolence dimension we identified here is behaviorally relevant inasmuch as it is consistently and positively related to laboratory behaviors that reflect the propensity to cooperate in daily life (McAuliffe et al., 2018; Peysakovich, Nowak, & Rand, 2014).

Although other factors from the bifactor model apparently underlie the self-report measures we studied here, their value for inspiring future theoretical and empirical work remains unclear. For instance, these group factors never uncovered correlates of forgiveness that the general factor failed to identify, they manifested themselves in slightly different ways in studies of laboratory transgressions than in studies of real-life transgressions, and they sometimes yielded nonsensical correlations with the behavioral targets and single-item measure of forgiveness used here. In contrast, the

general factor explains the huge preponderance of the variance in the individual questionnaire items, as well as in an external single-item measure of forgiveness. It also reliably predicts subjects' cooperative behavior, and it never yields nonsensical correlations with external validation criteria. In other words, the general factor of responses on self-report forgiveness scales behaves like a measure of forgiveness should.

Going forward, we believe that substantial theoretical and empirical insights might be gained by viewing forgiveness as prosocial change along a malevolence-benevolence attitudinal continuum. In particular, the proposition that forgiveness reflects attitude change implies that much could be learned about forgiveness by applying the basic principles that social psychologists have already discovered about attitudes and how to change them (Albarracin & Shavitt, 2018; Bohner & Dickel, 2011; Dalege et al., 2016). More generally, we believe and hope that these results demonstrate the promise of seeking to achieve a closer union between our theories of forgiveness and the empirical realities of the measures with which we seek to measure the constructs those theories invoke.

Table 1. Order of events for Studies 1, 2, and 3, including the partner's preprogrammed behavior, the timing of assessments for relationship value and exploitation risk, and the confederate's messages to subjects.

Experiment Stage	**Description**	
Prisoner's Dilemma Game	Round	Event
	1	Confederate Cooperates
	2-12	Confederate plays generous tit-for-tat
		Message encouraging continued cooperation
	13-19	Confederate defects
		Message: Apology Manipulation
	20-21	Confederate Cooperates
	22-29	Confederate plays generous tit-for-tat
Second Economic Game	Trust Game (Studies 1, 2, and 3) or Dictator Game (Study 1) or Pay-to-Punish Game (Study 1)	
Forgiveness	Transgression-Related Interpersonal Motivations	

Note. This table of the experimental protocols includes only the events that pertain to this manuscript. Other measures were taken during the PDG but are not reported here because they are beyond the scope of this manuscript; results using these measures are available in the manuscript written to be a complement to this (Billingsley et al., in prep).

1300

Table 2a. Model fit for different factor structures of the TRIM-NCO.

	χ^{2*} (df)	RMSEA [90% CI]	CFI	TLI
	Study 1 (U.S. Mechanical Turk)			
One Factor	4493.282 (119)	.190 [.185, .195]	0.921	0.909
Two Factor	2748.931 (118)	.148 [.143, .153]	0.952	0.945
Three Factor	1616.899 (116)	.113 [.108, .118]	0.973	0.968
Bifactor (S-1)	915.451 (108)	.086 [.081, .091]	0.985	0.982
	Study 2 (U.S. University/Community)			
One Factor	849.289 (119)	.164 [.154, .175]	0.879	0.862
Two Factor	543.428 (118)	.126 [.115, .137]	0.930	0.919
Three Factor	392.458 (116)	.102 [.091, .113]	0.954	0.946
Bifactor (S-1)	352.619 (108)	.100 [.088, .111]	0.960	0.949
	Study 3 (Japanese University)			
One Factor	818.650 (119)	.178 [.166, .189]	0.890	0.875
Two Factor	589.404 (118)	.147 [.135, .158]	0.926	0.915
Three Factor	522.736 (116)	.137 [.125, .149]	0.936	0.925
Bifactor (S-1)	328.437 (108)	.105 [.092, .118]	0.965	0.957

Note: χ^{2*} differences cannot be compared directly in a typical χ^2 difference test.

1301

Table 2b. Model fit for different factor structures of the TRIM-18.

	χ^{2*} (df)	RMSEA [90% CI]	CFI	TLI
	Study 4 (U.S. University)			
One Factor	777.548 (135)	.168 [.157, .180]	0.926	0.916
Two Factor	305.932 (134)	.087 [.074, .100]	0.980	0.977
Three Factor	271.662 (132)	.079 [.066, .093]	0.984	0.981
Bifactor (2-1)	315.680 (130)	.092 [.079, .105]	0.979	0.975
Bifactor (3-1)	261.947 (123)	.082 [.068, .096]	0.984	0.980
Bifactor (w/ DFS and EFS)	821.947 (523)	.058 [.051, .066]	0.977	0.974
	Study 5 (Japanese University)			
One Factor	898.322 (135)	.189 [.178, .201]	0.881	0.865
Two Factor	466.095 (134)	.125 [.113, .138]	0.948	0.941
Three Factor	447.372 (132)	.123 [.111, .136]	0.951	0.943
Bifactor (2-1)	466.167 (130)	.128 [.116, .141]	0.947	0.938
Bifactor (3-1)	389.809 (123)	.117 [.104, .130]	0.958	0.948
Bifactor (w/ DFS and EFS)	948.251 (523)	.072 [.064, .079]	0.960	0.955

Note: χ^{2*} differences cannot be compared directly in a typical χ^2 difference test.

1302

1303

Table 3a. Path coefficients for economic game decisions predicted by three-factor and bifactor (S-1) models of the TRIM-NCO in Studies 1-3.

		Three-Factor Model			Bifactor (S-1) Model		
		Avoid	Revenge	Benevolence	General Factor	Revenge-Specific	Benevolence-Specific
	Outcome						
Study 1	PDG	-.409 (.097)***	-.109 (.073)	-.011 (.131)	-.493 (.035)***	-.080 (.056)	-.011 (.058)
	TG	-.048 (.130)	-.121 (.107)	.479 (.181)**	.293 (.051)***	-.110 (.083)	.253 (.084)**
	DG	.240 (.186)	.157 (.124)	.004 (.241)	.350 (.049)***	.161 (.096)	-.061 (.085)
	PTPG	-.074 (.141)	-.462 (.096)***	.157 (.176)	-.218 (.056)***	-.343 (.074)***	.027 (.078)
Study 2							
	PDG	-.250 (.133)	-.154 (.122)	-.067 (.134)	-.403 (.062)***	-.101 (.092)	-.056 (.094)
	TG	.187 (.127)	-.287 (.107)**	.358 (.117)**	.263 (.064)***	-.235 (.078)**	.224 (.081)**
Study 3							
	PDG	-.099 (.178)	.178 (.155)	-.441 (.178)*	-.379 (.066)***	.139 (.120)	-.148 (.097)
	TG	-.030 (.158)	.102 (.128)	-.007 (.159)	.030 (.070)	.077 (.100)	-.002 (.091)

Notes: Estimates are reported as standardized with standard errors. Asterisks indicate p-value ranges: *$p < .05$, **$p < .01$, ***$p < .001$

Table 3b. Path coefficients for single-item forgiveness predicted by three-factor and bifactor (S-1) models of the TRIM-18 in Studies 4 and 5.

	Three-Factor Model			Bifactor (3-1) Model		
	Avoid	Revenge	Benevolence	General Factor	Revenge-Specific	Benevolence-Specific
Study 4	.072 (.187)	.250 (.091)**	.382 (.167)*	.582 (.060)***	.144 (.113)	-.044 (.103)
Study 5	-.421 (.175)*	.307 (.067)***	.807 (.179)***	.539 (.063)***	.198 (.055)***	.178 (.055)**
	Two-Factor Model			Bifactor (2-1) Model		
	Avoidance-Benevolence		Revenge	General Factor		Revenge-Specific
Study 4	.450 (.074)***		.232 (.086)**	.561 (.061)***		-.208 (.073)**
Study 5	.388 (.073)***		.276 (.070)***	.532 (.063)***		.243 (.055)***

Notes: Estimates are reported as standardized with standard errors. Asterisks indicate p-value ranges: *$p < .05$, **$p < .01$, ***$p < .001$.

1306

Table 4a. Correlations between simple composite of 17 items from TRIM-NCO, the bifactor (S-1) model, and behavioral outcomes.

	Study 1			Study 2			Study 3		
	r	se	*p*	*r*	se	*p*	*r*	se	*p*
General	0.946	0.004	<.001	0.937	0.009	<.001	0.940	0.012	<.001
Revenge-Specific	0.279	0.011	<.001	0.313	0.026	<.001	0.248	0.032	<.001
Benevolence-Specific	0.275	0.009	<.001	0.365	0.023	<.001	0.306	0.026	<.001
PDG	-0.476	0.033	<.001	-0.414	0.057	<.001	-0.318	0.069	<.001
TG	0.313	0.052	<.001	0.232	0.062	<.001	0.043	0.073	0.557
DG	0.360	0.050	<.001	-	-	-	-	-	-
PTP	-0.300	0.051	<.001	-	-	-	-	-	-

Table 4b. Correlations between simple composite of 18 items from TRIM-18, the bifactor (S-1) model, and single-item forgiveness.

	Study 4			Study 5		
	r	se	*p*	*r*	se	*p*
General	0.978	0.009	<.001	0.953	0.011	<.001
Revenge-Specific	0.092	0.029	.001	0.189	0.025	<.001
Benevolence-Specific	0.033	0.024	.174	-0.003	0.027	.903
Single-Item	0.586	0.051	<.001	0.589	0.052	<.001

Table 5a. Raw means for item responses and item difficulties based on the Location Index from the Item Response Function in the three Experiments. Each experiment is sorted by item difficulty (easies to hardest).

Study 1				Study 2				Study 3			
Item	Difficulty	Mean	Scale	*Item*	Difficulty	Mean	Scale	*Item*	Difficulty	Mean	Sc
17	-3.91	4.35	REV	*17*	-2.75	3.95	REV	*17*	-4.85	4.39	RE
4	-3.09	4.49	REV	*1*	-1.78	3.9	REV	*13*	-4.16	4.34	RE
1	-2.55	4.2	REV	*4*	-1.67	4.1	REV	*1*	-3.07	4.35	RE
13	-2.33	4.15	REV	*8*	-1.16	3.71	BEN	*4*	-2.70	4.49	RE
16	-1.40	3.86	BEN	*2*	-1.14	3.66	AVO	*16*	-1.08	3.69	BE
8	-0.92	3.53	BEN	*18*	-1.05	3.56	AVO	*18*	-0.97	3.57	AV
3	-0.90	3.56	BEN	*11*	-0.99	3.57	AVO	*10*	-0.90	3.52	BE
14	-0.81	3.67	BEN	*5*	-0.91	3.53	AVO	*14*	-0.88	3.53	BE
10	-0.71	3.41	BEN	*13*	-0.80	3.62	REV	*8*	-0.86	3.52	BE
18	-0.65	3.31	AVO	*15*	-0.75	3.49	AVO	*5*	-0.81	3.42	AV
11	-0.59	3.3	AVO	*3*	-0.75	3.59	BEN	*11*	-0.80	3.47	AV
5	-0.58	3.35	AVO	*10*	-0.64	3.51	BEN	*2*	-0.77	3.42	AV
6	-0.56	3.37	BEN	*14*	-0.63	3.53	BEN	*6*	-0.54	3.46	BE
2	-0.55	3.33	AVO	*16*	-0.59	3.48	BEN	*15*	-0.50	3.32	AV
15	-0.19	3.12	AVO	*6*	-0.54	3.45	BEN	*7*	-0.50	3.28	AV
12	-0.07	3.03	AVO	*12*	-0.41	3.31	AVO	*3*	-0.46	3.25	BE
7	0.37	2.74	AVO	*7*	-0.05	3.17	AVO	*12*	-0.02	2.99	AV

Note: Item prompts can be referenced using Appendix A. Difficulty values represent the Location Inde based on the Item Response Function (LI_{IRF}; see supplemental materials for raw values and computation). Mean values represents the average observed item response, scored so that larger values indicate greater forgiveness. Item Scales represent *a priori* conceptualizations of the items: RE = Revenge; AVO = Avoidance; BEN = Benevolence.

Table 5b. Raw means for item responses and item difficulties based on the Location Index from the Item Response Function in the two studies. Items within each study are sorted by difficulty (easiest to hardest).

Study 4				Study 5			
Item	Difficulty	Mean	Scale	*Item*	Difficulty	Mean	Scale
13	-1.918		REV	*1*	-1.565		REV
17	-1.798		REV	*17*	-1.245		REV
1	-1.611		REV	*13*	-1.239		REV
4	-1.49		REV	*4*	-1.028		REV
9	-1.103		REV	*9*	-0.981		REV
3	-0.91		BEN	*15*	-0.89		AVO
15	-0.633		AVO	*18*	-0.7		AVO
8	-0.508		BEN	*10*	-0.532		BEN
5	-0.505		AVO	*11*	-0.514		AVO
11	-0.504		AVO	*5*	-0.504		AVO
18	-0.424		AVO	*2*	-0.451		AVO
10	-0.384		BEN	*8*	-0.414		BEN
2	-0.328		AVO	*12*	-0.38		AVO
6	-0.281		BEN	*14*	-0.308		BEN
14	-0.257		BEN	*6*	-0.303		BEN
12	-0.187		AVO	*16*	-0.255		BEN
7	-0.123		AVO	*3*	-0.214		BEN
16	-0.055		BEN	*7*	-0.059		AVO

Note: Item prompts can be referenced using Appendix B. Difficulty values represent the Location Index based on the Item Response Function (LIIRF; see supplemental materials for raw values and computation). Mean values represents the average observed item response, scored so that larger values indicate greater forgiveness. Item Scales represent *a priori* conceptualizations of the items: REV = Revenge; AVO = Avoidance; BEN = Benevolence. Item 9 was removed from the three experiments due to poor fit; however, it was retained in the two non-experimental studies.

1312

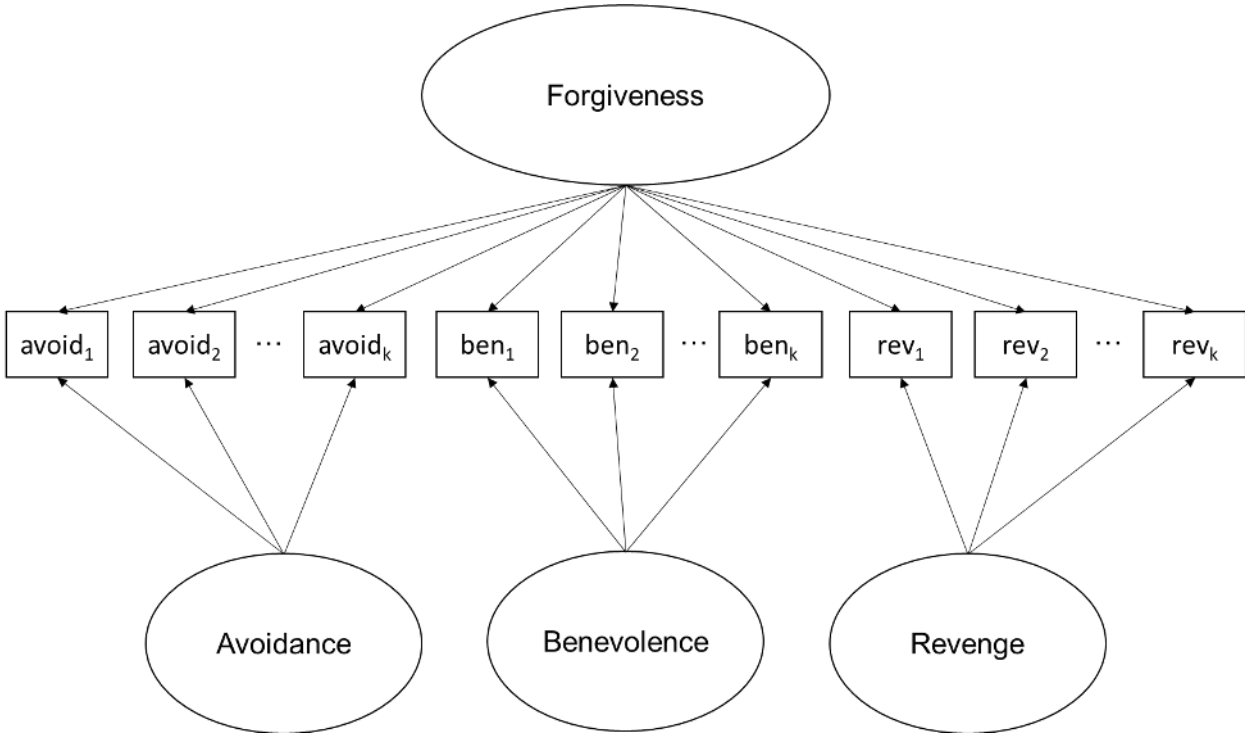

Figure 1. Traditional bifactor model applied to item responses on the TRIM-18, which measures three specific motives (avoidance, benevolence, and revenge) to construct a general forgiveness measure.

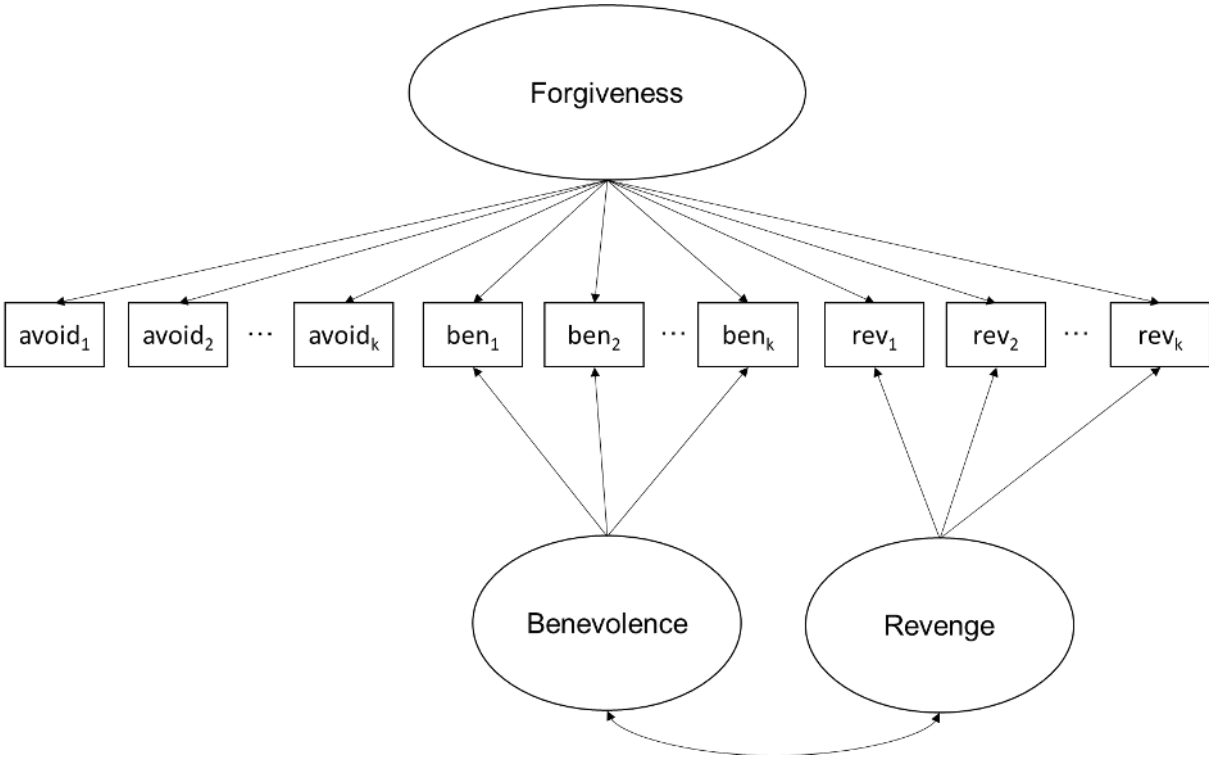

Figure 2. A bifactor (S-1) model applied to the TRIM-18. Specific domains are modeled separately from the general factor, with the exception of a reference domain, which is used to define the scale of the general factor.

Appendix A. Items from the Transgression Related Interpersonal Motivations questionnaire for Non-Close Others.

For the following questions, please indicate your current thoughts and feelings about [Target] using the scale below. Even though you will never encounter [Target] again, we are interested in how you think you would respond if you were to encounter [Target] in your daily life.

1 = strongly disagree
2 = mildly disagree
3 = agree and disagree equally
4 = mildly agree
5 = strongly agree
6 = I prefer not to answer

1. I would want to make him/her pay for treating me badly today.
2. I would try to keep as much distance between the two of us as possible.
3. I would have good will for him/her.
4. I would hope for something bad to happen to him/her.
5. I would have nothing to do with him/her.
6. I would try to put aside any reservations I had in order to develop a good relationship with him/her.
7. I would not trust him/her.
8. I would be willing to work toward a positive relationship with him/her.
9. I would want to see him/her get what he/she deserves.
10. I would act warmly towards him/her.
11. I would avoid contact with him/her.
12. I would be very happy to interact with him/her.
13. I would want to get even with him/her.
14. I would try to give up negative feelings toward him/her.
15. I would avoid working with him/her.
16. I would be willing to let go of my anger towards him/her.
17. I would want to seek revenge.
18. I would try to avoid him/her.

Note: The 9th item exhibited low factor loadings across all modeling techniques and experiments; therefore, we suggest that future research omit this item to create a scale of 17 items.

Appendix B. Items from the Transgression Related Interpersonal Motivations questionnaire for Close Others (TRIM-18).

For the following questions, please indicate your current thoughts and feelings about the person who hurt you; that is, we want to know how you feel about that person ***right now****. Next to each item, circle the number that best describes your current thoughts and feelings.*

1 = strongly disagree
2 = disagree
3 = neutral
4 = agree
5 = strongly agree

1. I'll make him or her pay.
2. I am trying to keep as much distance between us as possible.
3. Even though his/her actions have hurt me, I have good will for him/her.
4. I wish that something bad would happen to him/her.
5. I am living as if he/she doesn't exist, isn't around.
6. I want us to bury the hatchet and move forward with our relationship.
7. I don't trust him/her.
8. Despite what he/she did, I want us to have a positive relationship.
9. I want him/her to get what he/she deserves.
10. I am finding it difficult to act warmly towards him/her.
11. I am avoiding him/her.
12. Although he/she hurt me, I am putting the hurt aside so we can resume our relationship.
13. I'm going to get even.
14. I have given up my hurt and resentment.
15. I cut off the relationship with him/her.
16. I have released my anger so I can work on restoring our relationship to health.
17. I want to see him or her hurt and miserable.
18. I withdraw from him/her.

Appendix C. Items from the Decisional Forgiveness Scale (DFS).

Think of your current intentions toward the person who hurt you. Indicate the degree to which you agree or disagree with the following statements.

1 = strongly disagree
2 = disagree
3 = neutral
4 = agree
5 = strongly agree

1. I intend to try to hurt him or her in the same way he or she hurt me.
2. I will not try to help him or her if he or she needs something.
3. If I see him or her, I will act friendly.
4. I will try to get back at him or her.
5. I will try to act toward him or her in the same way I did before he or she hurt me.
6. If there is an opportunity to get back at him or her, I will take it.
7. I will not talk with him or her.
8. I will not seek revenge upon him or her.

Appendix D. Items from the Emotional Forgiveness Scale (EFS).

Think of your current emotions toward the person who hurt you. Indicate the degree to which you agree or disagree with the following statements.

1 = strongly disagree
2 = disagree
3 = neutral
4 = agree
5 = strongly agree

1. I care about him or her.
2. I no longer feel upset when I think of him or her.
3. I'm bitter about what he or she did to me.
4. I feel sympathy toward him or her.
5. I'm mad about what happened.
6. I like him or her.
7. I resent what he or she did to me.
8. I feel love toward him or her.

References

Albarracin, D., & Shavitt, S. (2018). Attitudes and attitude change. *Annual Review of Psychology, 69*, 299-327.

Ali, U. S., Chang, H. H., & Anderson, C. (2015). Location Indices for Ordinal Polytomous Items Based on Item Response Theory. *ETS Research Report Series, 2015*(2), 1-13.

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

Axelrod, R. (1980). More effective choice in the prisoner's dilemma. *Journal of Conflict Resolution, 24*(3), 379-403.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior, 10*(1), 122-142.

Billingsley, J., Forster, D. E., Russell, V. M., Burnette, J. L., Ohtsubo, Y., Lieberman, D., & McCullough, M. E. (in prep). Apologies promote forgiveness and restore cooperation by increasing perceived relationship value.

Billingsley, J., & Losin, E. A. (2017). The neural systems of forgiveness: an evolutionary psychological perspective. *Frontiers in Psychology, 8*, 737.

Bohner, G., & Dickel, N. (2011). Attitudes and attitude change. *Annual Review of Psychology, 62*, 391-417.

Bono, G., McCullough, M. E., & Root, L. M. (2008). Forgiveness, feeling connected to others, and well-being: Two longitudinal studies. *Personality and Social Psychology Bulletin, 34*(2), 182-195.

Buchanan, E. M., & Scofield, J. E. J. B. r. m. (2018). Methods to detect low quality data and its implication for psychological research. 1-11.

Buhrmester, M., Kwang, T., & Gosling, S. D. J. P. o. p. s. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *, 6*(1), 3-5.

Carlisle, R. D., Tsang, J.-A., Ahmad, N. Y., Worthington Jr, E. L., Witvliet, C. V., & Wade, N. (2012). Do actions speak louder than words? Differential effects of apology and restitution on behavioral and self-report measures of forgiveness. *The Journal of Positive Psychology, 7*(4), 294-305.

Carmody, P., Gordon, K. J. P., & Differences, I. (2011). Offender variables: Unique predictors of benevolence, avoidance, and revenge? *, 50*(7), 1012-1017.

Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychological Review, 123*(1), 2.

De Waal, F., & Pokorny, J. (2005). Primate questions about the art and science of forgiving. In E. L. Worthington Jr (Ed.), *Handbook of Forgiveness* (pp. 17-32). New York, NY: Routledge.

Dorn, K., Hook, J. N., Davis, D. E., Van Tongeren, D. R., & Worthington Jr, E. L. (2014). Behavioral methods of assessing forgiveness. *The Journal of Positive Psychology, 9*(1), 75-80.

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods, 22*(3), 541.

Elliot, A. J., & Thrash, T. M. (2002). Approach-avoidance motivation in personality: approach and avoidance temperaments and goals. *Journal of personality and social psychology, 82*(5), 804.

1484 Exline, J. J., Baumeister, R. F., Bushman, B. J., Campbell, W. K., & Finkel, E. J. (2004).
1485 Too proud to let go: narcissistic entitlement as a barrier to forgiveness. *Journal of*
1486 *personality and social psychology, 87*(6), 894.
1487 Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution*
1488 *and Human Behavior, 25*(2), 63-87.
1489 Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple
1490 bargaining experiments. *Games and economic behavior, 6*(3), 347-369.
1491 Geiser, C., Eid, M., & Nussbeck, F. W. (2008). On the meaning of the latent variables in
1492 the CT-C (M-1) model: A comment on Maydeu-Olivares and Coffman (2006).
1493 *Psychological Methods, 13*(1), 49.
1494 Gupta, S. K. (2011). Intention-to-treat concept: a review. *Perspectives in Clinical*
1495 *Research, 2*(3), 109.
1496 Gurtman, M. B. (2009). Exploring personality with the interpersonal circumplex. *Social*
1497 *and Personality Psychology Compass, 3*(4), 601-619.
1498 Harper, Q., Worthington Jr, E. L., Griffin, B. J., Lavelock, C. R., Hook, J. N., Vrana, S.
1499 R., & Greer, C. L. J. J. o. C. P. (2014). Efficacy of a workbook to promote
1500 forgiveness: A randomized controlled trial with university students. *70*(12), 1158-
1501 1169.
1502 Heinrich, M., Zagorscak, P., Eid, M., & Knaevelsrud, C. (2018). Giving G a Meaning: An
1503 Application of the Bifactor-(S-1) Approach to Realize a More Symptom-Oriented
1504 Modeling of the Beck Depression Inventory–II. *Assessment*, 1073191118803738.
1505 Hendriks, A. (2012). SoPHIE—Software platform for human interaction experiments.
1506 *University of Osnabrück, Osnabrück*.
1507 Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?
1508 *Behavioral and Brain Sciences, 33*(2-3), 61-83.
1509 Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2*(1), 41-
1510 54.
1511 Hook, J. N., Worthington Jr, E. L., & Utsey, S. O. (2009). Collectivism, forgiveness, and
1512 social harmony. *The Counseling Psychologist, 37*(6), 821-847.
1513 Hook, J. N., Worthington Jr, E. L., Utsey, S. O., Davis, D. E., & Burnette, J. L. (2012).
1514 Collectivistic self-construal and forgiveness. *Counseling and Values, 57*(1), 109-
1515 124.
1516 Kan, I. P., & Drummey, A. B. J. C. i. H. B. (2018). Do imposters threaten data quality?
1517 An examination of worker misrepresentation and downstream consequences in
1518 Amazon's Mechanical Turk workforce. *83*, 243-253.
1519 Kashima, Y., Yamaguchi, S., Kim, U., Choi, S.-C., Gelfand, M. J., & Yuki, M. (1995).
1520 Culture, gender, and self: a perspective from individualism-collectivism research.
1521 *Journal of personality and social psychology, 69*(5), 925.
1522 McAuliffe, W. H., Forster, D. E., Pedersen, E. J., & McCullough, M. E. (2018). Does
1523 Cooperation in the Laboratory Reflect the Operation of a Broad Trait? *European*
1524 *Journal of Personality*.
1525 McCullough, M. E., Bono, G., & Root, L. M. (2007). Rumination, emotion, and
1526 forgiveness: three longitudinal studies. *Journal of personality and social*
1527 *psychology, 92*(3), 490.

McCullough, M. E., Fincham, F. D., & Tsang, J.-A. (2003). Forgiveness, forbearance, and time: the temporal unfolding of transgression-related interpersonal motivations. *Journal of personality and social psychology, 84*(3), 540.

McCullough, M. E., & Hoyt, W. T. (2002). Transgression-related motivational dispositions: Personality substrates of forgiveness and their links to the Big Five. *Personality and Social Psychology Bulletin, 28*(11), 1556-1573.

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2013). Putting revenge and forgiveness in an evolutionary context. *Behavioral and Brain Sciences, 36*(1), 41-58.

McCullough, M. E., Luna, L. R., Berry, J. W., Tabak, B. A., & Bono, G. (2010). On the form and function of forgiving: modeling the time-forgiveness relationship and testing the valuable relationships hypothesis. *Emotion, 10*(3), 358.

McCullough, M. E., Pargament, K. I., & Thoresen, C. E. (2001). *Forgiveness: Theory, research, and practice*: Guilford Press.

McCullough, M. E., Pedersen, E. J., Tabak, B. A., & Carter, E. C. (2014). Conciliatory gestures promote forgiveness and reduce anger in humans. *Proceedings of the National Academy of Sciences, 111*(30), 11211-11216.

McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington Jr, E. L., Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal of personality and social psychology, 75*(6), 1586.

McCullough, M. E., Root, L. M., & Cohen, A. D. (2006). Writing about the benefits of an interpersonal transgression facilitates forgiveness. *Journal of consulting and clinical psychology, 74*(5), 887.

Muthén, B. O., & Muthén, L. K. (1998-2012). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Ohtsubo, Y., Matsunaga, M., Tanaka, H., Suzuki, K., Kobayashi, F., Shibata, E., . . . Ohira, H. (2018). Costly apologies communicate conciliatory intention: an fMRI study on forgiveness in response to costly apologies. *Evolution and Human Behavior, 39*(2), 249-256.

Ohtsubo, Y., & Watanabe, E. (2009). Do sincere apologies need to be costly? Test of a costly signaling model of apology. *Evolution and Human Behavior, 30*(2), 114-123.

Ohtsubo, Y., Watanabe, E., Kim, J., Kulas, J. T., Muluk, H., Nazar, G., . . . Zhang, J. (2012). Are costly apologies universally perceived as being sincere? A test of the costly apology-perceived sincerity relationship in seven countries. *Journal of Evolutionary Psychology, 10*(4), 187-204.

Ohtsubo, Y., Yamaura, K., & Yagi, A. J. J. J. o. S. P. (2015). Development of Japanese measures of reconciliatory tendencies. *31*(2), 135-142.

Orathinkal, J., Vansteenwegen, A., Enright, R. D., & Stroobants, R. (2007). Further validation of the Dutch version of the Enright Forgiveness Inventory. *Community Mental Health Journal, 43*(2), 109-128.

Peysakhovich, A., Nowak, M. A., & Rand, D. G. (2014). Humans display a ‘cooperative phenotype’that is domain general and temporally stable. *Nature Communications, 5*, 4939.

Psychology Software Tools, I. (2012). E-Prime (Version 2.0). Retrieved from http://www.pstnet.com.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Resesarch, 47*(5), 667-696.

Reise, S. P., Kim, D. S., Mansolf, M., & Widaman, K. F. (2016). Is the bifactor model a better model or is it just better at modeling implausible responses? Application of iteratively reweighted least squares to the Rosenberg Self-Esteem Scale. *Multivariate Behavioral Resesarch, 51*(6), 818-838.

Rusbult, C. E., Hannon, P. A., Stocker, S. L., & Finkel, E. J. (2005). Forgiveness and relational repair. In E. L. Worthington Jr (Ed.), *Handbook of Forgiveness* (pp. 185-205). New York, NY: Routledge.

Rye, M. S., Loiacono, D. M., Folck, C. D., Olszewski, B. T., Heim, T. A., & Madia, B. P. (2001). Evaluation of the psychometric properties of two forgiveness scales. *Current Psychology, 20*(3), 260-277.

Schug, J., Yuki, M., Horikawa, H., & Takemura, K. (2009). Similarity attraction and actually selecting similar others: How cross-societal differences in relational mobility affect interpersonal similarity in Japan and the USA. *Asian Journal of Social Psychology, 12*(2), 95-103.

Subkoviak, M. J., Enright, R. D., Wu, C.-R., Gassin, E. A., Freedman, S., Olson, L. M., & Sarinopoulos, I. (1995). Measuring interpersonal forgiveness in late adolescence and middle adulthood. *Journal of Adolescence, 18*(6), 641-655.

Tabak, B. A., McCullough, M. E., Luna, L. R., Bono, G., & Berry, J. W. (2012). Conciliatory gestures facilitate forgiveness and feelings of friendship by making transgressors appear more agreeable. *Journal of personality, 80*(2), 503-536.

Tsang, J. a., McCullough, M. E., & Fincham, F. D. (2006). The longitudinal association between forgiveness and relationship closeness and commitment. *Journal of Clinical Psychology, 25*(4), 448-472.

Wilhelm, O., Kaltwasser, L., & Hildebrandt, A. (2018). Will the real factors of prosociality please stand up? a comment on böckler, tusche, and singer (2016). *Social Psychological and Personality Science, 9*(4), 493-499.

Worthington, E. L., Lavelock, C., vanOyen Witvliet, C., Rye, M. S., Tsang, J.-A., & Toussaint, L. (2015). Measures of forgiveness: Self-report, physiological, chemical, and behavioral indicators. In *Measures of personality and social psychological constructs* (pp. 474-502): Elsevier.

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18*(2), 129-166.