

Psychological Assessment

The Psychology of Spite and the Measurement of Spitefulness

David K. Marcus, Virgil Zeigler-Hill, Sterett H. Mercer, and Alyssa L. Norris

Online First Publication, February 17, 2014. <http://dx.doi.org/10.1037/a0036039>

CITATION

Marcus, D. K., Zeigler-Hill, V., Mercer, S. H., & Norris, A. L. (2014, February 17). The Psychology of Spite and the Measurement of Spitefulness. *Psychological Assessment*. Advance online publication. <http://dx.doi.org/10.1037/a0036039>

The Psychology of Spite and the Measurement of Spitefulness

David K. Marcus
Washington State University

Virgil Zeigler-Hill
Oakland University

Sterett H. Mercer
University of British Columbia

Alyssa L. Norris
Washington State University

Spite is an understudied construct that has been virtually ignored within the personality, social, and clinical psychology literatures. This study introduces a self-report Spitefulness Scale to assess individual differences in spitefulness. The scale was initially tested on a large sample of 946 college students and cross-validated on a national sample of 297 adults. The scale was internally consistent in both samples. Factor analysis supported a 1-factor solution for the initial pool of 31 items. Item response theory analysis was used to identify the best performing of the original 31 items in the university sample and reduce the scale to 17 items. Tests of measurement invariance indicated that the items functioned similarly across both university and national samples, across both men and women, and across both ethnic majority and minority groups. Men reported higher levels of spitefulness than women, younger people were more spiteful than older people, and ethnic minority members reported higher levels of spitefulness than ethnic majority members. Across both samples, spitefulness was positively associated with aggression, psychopathy, Machiavellianism, narcissism, and guilt-free shame, and negatively correlated with self-esteem, guilt-proneness, agreeableness, and conscientiousness. Ideally, this Spitefulness Scale will be able to predict behavior in both laboratory settings (e.g., ultimatum games, aggression paradigms) and everyday life, contribute to the diagnosis of personality disorders and oppositional defiant disorder, and encourage further study of this neglected, often destructive, trait.

Keywords: spite, personality assessment, item response theory, dark personality traits

Spite and spitefulness are understudied constructs in social, personality, and clinical psychology. None of the major social, personality, or clinical psychology journals, including the *Journal of Personality and Social Psychology*, *Personality and Social Psychology Bulletin*, the *Journal of Personality*, *Psychological Assessment*, and the *Journal of Abnormal Psychology*, have published a single article examining spiteful behavior or individual differences in spitefulness. Spitefulness is not directly represented among the facets of the five-factor model of personality and we are unaware of any assessment instruments designed to assess spitefulness. Furthermore, spitefulness only appears once in the *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition (*DSM-5*; American Psychiatric Association [APA], 2013) as a symptom of oppositional defiant disorder (ODD; “Has been spiteful or vindictive at least twice within the past 6 months” p. 462).

Although ODD can now be diagnosed in adults, it is generally considered a childhood disorder. Spitefulness is not listed as a symptom of any of the personality disorders or any other psychological disorders typically found in adulthood.

Although social, personality, and clinical psychologists have not studied spite, evolutionary biologists and behavioral economists have conducted empirical research on spite, psychoanalysts have speculated about the origin and functions of spiteful behavior, and legal theorists have considered the role of spite in litigation. Starting with Hamilton (1970), evolutionary biologists have defined spite as behaviors that have negative consequences for both the actor and the recipient. Similarly, game theorists (Hamburger, 1979) have defined spiteful acts as those that maximize differences, even at personal costs (e.g., it is better to receive \$4 and have the other player receive \$1 than to each receive \$5). According to behavioral economists, a spiteful person is “willing to decrease the economic payoff of a reference agent at a personal cost to himself” (Fehr & Fischbacher, 2005, p. 154), and spiteful acts “incur costs that generate neither present nor future material rewards” (Cullis, Jones, & Soliman, 2012, p. 418). From a different perspective, psychoanalysts have written about the “the self-destructiveness of spite” (Shabad, 2000, p. 690) and define spite to refer to instances in which people harm themselves to punish another (e.g., people with borderline personality disorder hurting themselves to punish those who care about them; Critchfield, Levy, Clarkin, & Kernberg, 2008). From a legal perspective, spite has been defined as “the willingness of a litigant to reduce his

David K. Marcus, Department of Psychology, Washington State University; Virgil Zeigler-Hill, Department of Psychology, Oakland University; Sterett H. Mercer, Department of Educational & Counselling Psychology and Special Education, University of British Columbia, Vancouver, British Columbia, Canada; Alyssa L. Norris, Department of Psychology, Washington State University.

Correspondence concerning this article should be addressed to David K. Marcus, Department of Psychology, Washington State University, Pullman, WA 99164-4820. E-mail: david.marcus@wsu.edu

payoff in order to reduce his opponent's payoff" (Cooter, Marks, & Mnookin, 1982, p. 239). All of these definitions are consistent with the saying "cutting off one's nose to spite one's face." Such spiteful behavior is not necessarily masochistic because the self-harm is in the service of harming the other and not an end in itself. Although spite may be defined more broadly to include any acts that involve vindictively harming another, there are advantages to applying a narrower definition derived from these other disciplines. Specifically, the requirement that spite must involve harm to the actor (and not just to the other) can distinguish spite from other selfish, sadistic, hostile, or aggressive behaviors. Thus, although measures of spitefulness and aggression should be positively correlated, based on this conceptualization of spite, they should be distinct constructs.

Spitefulness that includes an element of self-harm can be a powerful motive with potentially serious and often negative psychological, interpersonal, and societal consequences. Because psychologists have yet to measure spitefulness, the evidence that spitefulness is an important motive that merits study is largely circumstantial and anecdotal. Consider the various domains of suicide, terrorism, divorce negotiations, and tax policy. Most suicides may be motivated by dysphoric states such as hopelessness, shame, a sense of perceived burdensomeness, and a lack of belonging (Joiner, 2010; Joiner & Rudd, 2000). However, there also appears to be a smaller subset of spiteful suicides that are motivated, at least in part, by the desire to take revenge on others, for example, by killing themselves in a way that will traumatize the person with whom they were angry (Joiner, 2010). There have also been documented cases of people staging their suicides to look like homicides (Prahlow, Long, & Barnard, 1998). With respect to terrorism, it appears that one of the primary motivations of suicide bombers is to get revenge on a hated enemy (Gambetta, 2005; Ricolfi, 2005). In all of these instances, individuals are willing to engage in the ultimate form of self-harm to harm others emotionally, legally, or physically. At the interpersonal level, spiteful behavior "is a familiar aspect of divorce negotiations" (Scott, 1992, p. 646). These negotiations have been likened to a negative sum game, in which out of spite the parties in conflict "diminish or destroy the value of the benefits to be distributed" (Scott, 1992, p. 644). Furthermore, spiteful motives during child custody negotiations may damage children's relationships with their parents, ultimately harming all of the parties involved (Johnston, 2003). At the societal level, using an analogue experiment, Cullis et al. (2012) found that increasing the audit rate and penalties can elicit spiteful tax evasion in which people at the greatest risk of being caught hide more income.

The possibility that spiteful actors can behave in a manner that incurs costs with no direct gains (whether in terms of fitness or finances) poses a challenge, at least superficially, to the basic assumptions that underlie traditional evolutionary or economic theory. Among evolutionary biologists, Hamilton (1970) hypothesized that spiteful behavior may occur when the actor is negatively related to the recipient (i.e., the actor shares fewer genes with the recipient than with the average member of the population). Given the necessary conditions for spite (including accurate kin determination to identify negative relatedness), Hamilton (1970) acknowledged that genuinely spiteful behavior is probably rare among nonhuman animals (except perhaps social insects) but that humans may be capable of such behaviors. Behavioral econ-

omists (e.g., Fehr & Fischbacher, 2005) have used examples of spiteful behavior to refute the self-interest hypothesis that is central to classical economics. For example, in an ultimatum game, a self-interested individual acting purely to maximize profit should accept \$1 regardless of whether the other player receives \$1 or \$20, yet players often reject such unfair divisions to spite the other player even though it costs them money to do so.

Most of the empirical research on spiteful behavior by humans has been conducted by behavioral economists and game theorists who have focused on the question of when do people behave spitefully. Using the ultimatum game, prisoner's dilemma, and other games, economists have found that certain situations are more likely to elicit spiteful behavior than other situations. For example, Pillutla and Murnighan (1996) examined spiteful behavior using an ultimatum game in which participants had the opportunity to accept or reject an offer of \$1 or \$2 from an anonymous other. If the participant rejected the offer, neither received money. Further, in a modification of the ultimatum design, in some conditions the participant still received \$1 or \$2 even if he or she rejected the offer. The authors varied whether the participants knew how much money was being divided and whether the participants believed that the other knew how much they would receive if they rejected the offer. Participants were angriest and most likely to reject the offer when they knew that the division was unequal (i.e., \$20 was being divided) and when they thought that the other knew their fallback option (i.e., how much they would get for rejecting the offer).

Although the economic researchers have not attempted to identify personality traits or characteristics associated with spitefulness, their research finds that a sizable minority of participants is consistently spiteful and thus suggests large individual differences in spitefulness. For example, in the Pillutla and Murnighan (1996) study, over one third of the participants rejected an offer of \$2 when they knew that \$20 was being divided and they believed that the other knew that their fallback option was getting \$1 for rejecting the offer. These individuals behaved spitefully, costing themselves \$1 in order to prevent the other from receiving \$18. Using a bidding game¹ task, Kimbrough and Reiss (2012) found that spitefulness was bimodally distributed, with the majority of participants behaving in a nonspiteful manner across trials and a minority of individuals (roughly 25%) who were consistently spiteful. Similarly, Levine (1998) developed a mathematical model to explain the findings from a variety of resource allocation games and concluded that 20% of the players were spiteful. Perhaps most strikingly, Falk, Fehr, and Fischbacher (2005) designed a three-person prisoner's dilemma in which after each trial, participants were given the opportunity to spitefully punish the other players by paying to reduce the other players' winnings. Not surprisingly, cooperators often punished defectors, but unexpectedly, it was not uncommon for defectors to pay to punish both other defectors and cooperators. Reviewing the resource allocation literature, Fehr and Fischbacher (2005) suggested that there are individuals with a spiteful or envious preference who "always values the economic

¹ Unlike other spite studies in this discipline, this study adopted a broader definition of spite and used a bidding task game in which participants had the opportunity to raise the cost of the item for their opponents without incurring any risks or costs to themselves.

payoff of relevant reference agents negatively” (p. 154). Is there a way to identify who these spiteful individuals are and how they differ from those willing to accept unequal offers? More importantly, is there a way to identify individuals who are spiteful across a wide range of everyday activities and not just when competing in resource allocation games?

A few psychological studies have attempted to link behavior in ultimatum games to various emotions or psychological states, but (unlike the behavioral economic literature) these studies have not interpreted rejection of offers as spiteful. The findings from these studies have generally been inconsistent with induced sadness leading to greater rejection of unequal offers (Harlé & Sanfey, 2007, 2010), but trait negative affect (Dunn, Makarova, Evans, & Clark, 2010) or clinical depression (Harlé, Allen, & Sanfey, 2010) resulting in greater acceptance of unequal offers. Conversely, induced amusement (Harlé & Sanfey, 2007, 2010) led to increased acceptance of unequal offers, whereas trait positive affect was associated with rejection of unequal offers (Dunn et al., 2010). Perhaps most relevant to our attempt to develop a self-report measure of spitefulness, Almakias and Weiss (2012) found that individuals with a dismissing avoidant attachment style (high avoidance with low anxiety) were the most likely to reject unequal offers in an ultimatum game.

Measuring Spitefulness: The Nomological Network

We suspect that a primary reason why spitefulness has not received attention within the psychological literature is because there has been no instrument to measure this trait. There are a number of emotions and associated personality variables that are likely to be related to trait spitefulness. Not surprisingly, anger has been implicated in spitefulness. In their ultimatum game, Pillutla and Murnighan (1996) found that anger was a strong predictor of offer rejections, even more than perceptions of unfairness. Therefore, trait spitefulness is likely to be associated with trait hostility and aggression. This supposition is consistent with the view that spiteful behavior may often result from a tendency to externalize and a desire to punish others for their perceived transgressions. Relatedly, the dismissing avoidant attachment style that predicted spiteful rejections in an ultimatum game (Almakias & Weiss, 2012) is consistent with psychopathic personality traits, including a lack of empathy, callousness, impulsivity, and low levels of fear and anxiety (e.g., Patrick, Fowles, & Krueger, 2009).

Psychoanalysts have not conducted empirical research on spite, but based on their clinical experiences, a few psychoanalytic theorists have made observations about the personality characteristics of their spiteful patients. These analysts (Gottlieb, 2004; Shabad, 2000; Stern, 2004) have consistently linked spitefulness to experiences of shame. Considering that shame has been linked to anger, resentment, blame externalization, and indirect acts of aggression (Tangney, Stuewig, & Mashek, 2007), a positive association between shame-proneness and trait spitefulness is likely. In contrast, spitefulness should be associated with low levels of guilt because the presence of this social emotion should reduce the likelihood that individuals would deliberately inflict harm on others. Psychoanalysts have also observed envy and narcissism in their spiteful clients (Gottlieb, 2004; Shabad, 2000; Stern, 2004). These clinical observations are consistent with Kirchsteiger’s (1994) mathematical model, which demonstrated how envy (as

opposed to fairness) explains why participants in ultimatum games do not behave as rational actors using strategies that maximize their earnings. On the basis of these clinical observations and because envy is a component of narcissism, especially pathological narcissism (Krizan & Johar, 2012), it is likely that narcissistic individuals are spiteful. Machiavellianism is often associated with psychopathy and narcissism as the third component of the “dark triad” (Paulhus & Williams, 2002). Because Machiavellianism is defined by a high level of self-interest and spitefulness entails self-harm, we might expect that unlike psychopathy and narcissism, Machiavellianism would not be strongly associated with spitefulness. However, in Stern’s (2004) clinical observations about spiteful clients, he suggested that they are highly manipulative, which suggests that spitefulness may be associated with Machiavellianism.

In terms of general personality functioning, it seems likely that spiteful individuals will demonstrate low levels of agreeableness because agreeableness is antithetical to spitefulness. Given the inconsistent findings from the ultimatum game studies, we did not have a substantive basis for making clear predictions in other areas of personality and psychological functioning. Therefore, the associations with the other personality traits and measures of psychological functioning were considered exploratory.

The aim of the current study was to develop a psychometrically sound measure of spitefulness. After generating a set of candidate items for the spitefulness measure, we used item response theory (IRT) analysis to identify and retain the best performing items. We then tested for invariance of the IRT findings in an independent sample, as well as by sex and ethnicity. Finally, we examined the associations between this measure of spite and criterion variables in both samples.

Method

Participants

This study included three samples: two undergraduate samples and a national sample recruited through Amazon Mechanical Turk (MTurk). Participants in Sample 1 were 556 undergraduate students attending a public university in the Pacific Northwest, and Sample 2 included 390 undergraduate students from a public university in the Midwest. Students in both samples received research credit in exchange for their participation. Across these two samples, the mean age of the student participants was 19.9 ($SD = 3.5$), and 79.3% were women. Regarding racial/ethnic identity, 74.4% identified as Caucasian, 7.8% as Hispanic, 6.8% as Black, 5.3% as Asian, 1.2% as Native Hawaiian/Pacific Islander, 0.2% as American Indian, and 4.3% as “other.”

Sample 3 was composed of 297 MTurk participants. MTurk is an online system that allows individuals to earn small amounts of money in exchange for their completion of surveys. MTurk samples tend to be more representative of the U.S. population than convenience samples (Berinsky, Huber, & Lenz, 2012) and representative of Internet users in the United States (Ross, Irani, Silberman, Zaldivar, & Tomlinson, 2010). MTurk samples are typically more diverse in age, ethnicity, and geography than the typical undergraduate students recruited for research (Buhrmester, Kwang, & Gosling, 2011). We restricted participation to individuals in the United States, which is verified through IP addresses in

MTurk. For this study, individuals received \$1 in exchange for their participation. Participants ranged in age from 18 to 82 ($M = 36.4$, $SD = 13.1$), and 62.3% were women. Participant racial/ethnic representation was similar to that of the student samples, with 75.4% Caucasian, 9.8% Black, 6.1% Hispanic, 4.7% Asian, 1.7% American Indian, and 2.4% "other."

Materials and Procedure

The participants provided informed consent and completed all of the study measures online. They provided demographic information, completed the Spitefulness Scale, and a set of criterion measures. The measures were administered in random order. Unless otherwise noted, these criterion measures were completed by all three samples.

Spitefulness Scale. We generated 31 one-sentence items describing situations in which a person might behave spitefully. The items were generated by the first two authors, with some assistance from colleagues and graduate students. Each situation involved engaging in some behavior or expressing a preference that would harm another but that would also entail harm to oneself. This harm could be social, financial, physical, or an inconvenience (see the Appendix). For some of the items, the harm to the respondent was obvious (e.g., getting punched, paying a fine), whereas for other items, the harm was subtler (e.g., taking extra time to leave a parking lot or a test, harm to one's community, living with an unattractive yard). However, unlike in an aggression scale (e.g., "If others make me mad or upset, I often hurt them"), all of the spite items involved at least a minimal negative consequence for the respondent. Most of these scenarios involved personal behavior, but four of them were political. Ten of the items were reverse scored. After the items were generated, they were randomly ordered. Participants provided responses for each item using scales ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). This 31-item version of the scale yielded coefficient alphas of .88 in the university sample and .89 in the MTurk sample. Further details about the refinement and psychometric properties of the scale are provided in the Results section.

Forms and Functions of Aggression Scale. The Forms and Functions of Aggression Scale (Little, Henrich, Jones, & Hawley, 2003) was used to capture aggressive tendencies. This measure is based on the idea that aggressive behavior can be understood as the combination of the function of aggressive behavior (i.e., pure, reactive, or instrumental) and the behavioral form of aggression (i.e., overt or relational). Participants were asked to rate their level of agreement with each of the 36 items using scales ranging from 1 (*not at all*) to 4 (*completely true*). For the current analysis, we used the total score on the scale, which was internally consistent ($\alpha = .96$).²

Self-Report Psychopathy Scale. Psychopathy was measured in Samples 1 and 3 via the Self-Report Psychopathy Scale (SRP-III; Paulhus, Neumann, & Hare, in press; Williams, Paulhus, & Hare, 2007). The SRP-III was based on the revised version of Hare's Psychopathy Checklist (Hare, 2003) and was intended to serve as a measure of psychopathy in noncriminal samples. The version of the SRP-III used in the current study consists of 34 items and is based on the factor analysis conducted by Mahmut, Menictas, Stevenson, and Homewood (2011). Participants were instructed to indicate their agreement with each of the 34 state-

ments on scales ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). This version of the SRP-III consists of four subscales: Callous Affect (eight items; e.g., "I am often rude to people" [$\alpha = .80$]), Erratic Lifestyle (eight items; e.g., "I've often done something dangerous just for the thrill of it" [$\alpha = .78$]), Interpersonal Manipulation (eight items; e.g., "I think I could 'beat' a lie detector" [$\alpha = .74$]), and Criminal Tendencies (10 items; e.g., "Broken into a building or vehicle in order to steal something or vandalize" [$\alpha = .80$]).

Levenson Self-Report Psychopathy Scale. Psychopathy was measured in Sample 2 using the Levenson Self-Report Psychopathy Scale (LSRP; Levenson, Kiehl, & Fitzpatrick, 1995). The LSRP was designed to measure psychopathy in the general population. The LSRP consists of 26 items, and responses are provided on scales ranging from 1 (*disagree strongly*) to 4 (*agree strongly*). This version of the LSRP consists of two subscales: Primary Psychopathy (16 items; e.g., "For me, what's right is whatever I can get away with" [$\alpha = .75$]) and Secondary Psychopathy (10 items; e.g., "I find myself in the same kinds of trouble, time after time" [$\alpha = .74$]).

Narcissistic Personality Inventory. Narcissism was assessed using the 40-item version of the Narcissistic Personality Inventory (NPI; Raskin & Hall, 1979, 1981). The NPI is well validated and is often considered to be the standard measure of subclinical narcissistic personality features (Paulhus & Williams, 2002). Items on the NPI are in a forced-choice format such that participants must choose between a narcissistic and a nonnarcissistic statement for each item (e.g., "I like having authority over other people" or "I don't mind following orders"). Although there has been controversy regarding the underlying factor structure of the 40-item NPI (see Brown, Budzek, & Tamborski, 2009, for a review), Ackerman et al. (2010) has suggested that the NPI may consist of three factors: Leadership/Authority (11 items; e.g., "If I ruled the world, it would be a much better place" [$\alpha = .66$]), Grandiose Exhibitionism (10 items; e.g., "I really like to be the center of attention" [$\alpha = .77$]), and Exploitativeness/Entitlement (four items; e.g., "I find it easy to manipulate people" [$\alpha = .47$]).

Pathological Narcissism Inventory. The Pathological Narcissism Inventory (PNI; Pincus et al., 2009) was used to assess grandiose and vulnerable aspects of pathological narcissism. The PNI is a 52-item measure for which responses were made on scales ranging from 0 (*not at all like me*) to 5 (*very much like me*). The PNI measures seven dimensions of pathological narcissism: contingent self-esteem (e.g., "It's hard for me to feel good about myself unless I know other people like me"), exploitative tendencies (e.g., "I can make anyone believe anything I want them to"), self-sacrificing self-enhancement (e.g., "I try to show what a good person I am through my sacrifices"), hiding of the self (e.g., "When others get a glimpse of my needs, I feel anxious and ashamed"), grandiose fantasy (e.g., "I often fantasize about being recognized for my accomplishments"), devaluing (e.g., "When others don't meet my expectations, I often feel ashamed about what I wanted"), and entitlement rage (e.g., "It irritates me when people don't notice how good a person I am"). These seven dimensions, in turn, load onto the two higher order factors referred

² Alpha values for the criterion variables were computed using combined data from all of the samples who completed the measure.

to as grandiose narcissism (i.e., exploitative, self-sacrificing self-enhancement, grandiose fantasy, and entitlement rage) and vulnerable narcissism (i.e., contingent self-esteem, hiding of the self, and devaluing). Initial information concerning the reliability and validity of the PNI has shown that it correlates in the expected direction with other measures of narcissism and related constructs such as level of self-esteem, interpersonal style, clinical outcomes, and contingent self-esteem (Pincus et al., 2009). The internal consistencies of the PNI grandiosity and vulnerability subscales were .90 and .94, respectively.

Mach-IV. Machiavellianism was measured via the Mach-IV (Christie & Geis, 1970). The Mach-IV is a 20-item instrument that was developed to measure manipulative and deceitful tendencies as well as cynical and immoral beliefs (e.g., “The best way to handle people is to tell them what they want to hear”). Participants were asked to rate their level of agreement with the items of the Mach-IV using scales that ranged from 1 (*strongly disagree*) to 5 (*strongly agree*). The Mach-IV has been found to possess adequate psychometric properties, and it is the most widely used measure of Machiavellianism (McHoskey, Worzel, & Szyarto, 1998). In the current sample, the internal consistency for the Mach-IV was $\alpha = .75$.

Test of Self-Conscious Affect. The Test of Self-Conscious Affect (TOSCA-3; Tangney, Dearing, Wagner, & Gramzow, 2000) consists of 16 brief scenarios (e.g., “You break something at work and then hide it”) that are followed by either four or five possible responses that are rated with regard to their likelihood on scales ranging from 1 (*not likely*) to 5 (*very likely*). These possible responses capture shame-proneness (16 items; e.g., “You would think: ‘I’m inconsiderate’” [$\alpha = .79$]), guilt-proneness (16 items; e.g., “You would think: ‘This is making me anxious. I need to either fix it or get someone else to’” [$\alpha = .86$]), and four other subscales that were not used in the current study (i.e., externalization, detachment, alpha-pride, and beta-pride). The TOSCA-3 has been shown to possess adequate psychometric properties (e.g., Tangney & Dearing, 2002).

Big Five Inventory. The Big Five Inventory (BFI; John, Donahue, & Kentle, 1991) was used to capture personality features in Sample 2. The BFI is a 44-item questionnaire that assesses the Big Five personality dimensions of Extraversion (eight items; e.g., “I see myself as someone who is talkative” [$\alpha = .84$]), Emotional Stability (eight items; e.g., “I see myself as someone who is relaxed, handles stress well” [$\alpha = .81$]), Agreeableness (nine items; e.g., “I see myself as someone who is considerate and kind to almost everyone” [$\alpha = .77$]), Conscientiousness (nine items; e.g., “I see myself as someone who does a thorough job” [$\alpha = .77$]), and Openness (10 items; e.g., “I see myself as someone who is original, comes up with new ideas” [$\alpha = .76$]). The BFI has been shown to possess adequate psychometric properties in previous studies (e.g., Benet-Martínez & John, 1998).

HEXACO-60. Personality was assessed in Sample 3 using the HEXACO-60 (Ashton & Lee, 2009), which is a 60-item measure of six basic personality dimensions: Extraversion (10 items; e.g., “The first thing that I always do in a new place is to make friends” [$\alpha = .86$]), Emotionality (10 items; e.g., “I sometimes can’t help worrying about little things” [$\alpha = .75$]), Agreeableness (10 items; e.g., “Most people tend to get angry more quickly than I do” [$\alpha = .79$]), Conscientiousness (10 items; e.g., “I plan ahead and organize things, to avoid scrambling at the last minute” [$\alpha = .81$]),

Openness to Experience (10 items; e.g., “I like people who have unconventional views” [$\alpha = .79$]), and Honesty-Humility (10 items; e.g., “I would never accept a bribe, even if it were very large” [$\alpha = .75$]). Responses for each item were provided using response scales that ranged from 1 (*strongly disagree*) to 5 (*strongly agree*). The HEXACO-60 has been shown to possess adequate psychometric properties (e.g., Ashton & Lee, 2009).

Rosenberg Self-Esteem Scale. The Rosenberg Self-Esteem Scale (Rosenberg, 1965) is a 10-item measure of global self-esteem (e.g., “On the whole, I am satisfied with myself”). Participants were instructed to complete the instrument according to how they typically or generally feel about themselves. Responses were made on scales ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). This instrument is regarded as a well-validated and reliable measure of global self-regard (e.g., Blaskovich & Tomaka, 1991). The internal consistency of this measure for the current study was $\alpha = .90$.

Brief Symptom Inventory. The Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983) is a 53-item short form of the Symptom Checklist 90-Revised that was used in Sample 2 to assess nine areas of potential dysfunction: somatization (seven items; e.g., “Faintness or dizziness”), obsessive compulsive (six items; e.g., “Having to check and double check what you do”), interpersonal sensitivity (four items; e.g., “Your feelings being easily hurt”), depression (six items; e.g., “Feeling blue”), anxiety (six items; e.g., “Feeling fearful”), hostility (five items; e.g., “Temper outbursts that you could not control”), phobic anxiety (five items; e.g., “Feeling afraid in open spaces”), paranoid ideation (five items; e.g., “Feeling that most people cannot be trusted”), and psychoticism (five items; e.g., “The idea that something is wrong with your mind”). Respondents were asked to indicate how much they were distressed by symptoms from each area of dysfunction during the past week on scales ranging from 0 (*not at all*) to 4 (*extremely*). A composite score was used as an indicator of global distress. The internal consistency for the global severity index was .97 for the current study.

Data Analysis

To examine the degree to which the Spitefulness Scale met assumptions of IRT analyses, we first conducted an exploratory factor analysis (EFA) with data from the combined undergraduate sample in Mplus 7.11 (Muthén & Muthén, 2012) using the mean- and variance-adjusted weighted least squares (WLSMV) estimator to account for the ordered, categorical item format. Unidimensionality was examined by considering the magnitude of eigenvalues for extracted factors as well as the ratio of the first and second eigenvalues. To examine local independence, residual correlations from a one-factor confirmatory factor analysis (CFA) model, using the theta parameterization and WLSMV estimator, were examined, with any values greater than .20 warranting special consideration (Yen, 1984, 1993). Following these tests of assumptions, we fit a graded-response IRT model with marginal maximum likelihood estimation using the *grm* function of the *ltm* package (Rizopoulos, 2006) in R 2.15 (R Development Core Team, 2012) to the data from the combined undergraduate sample. The IRT results were used to select a final set of items based primarily on magnitude of slope (*a*) values. First, items that poorly discriminated ($a < .65$; Baker, 2001) between high and low levels of latent spitefulness

Table 1
Factor Loadings, Slope and Location Parameters, and Tests of Differential Item Functioning (DIF) Across Samples and by Sex for Final Spitefulness Scale Items

Item	λ	a	b_1	b_2	b_3	b_4	Sample	Sex	Ethnicity
							χ^2_{LR}	χ^2_{LR}	χ^2_{LR}
1	.66	1.69	0.32	1.51	2.30	3.27	3.29	9.13	10.43
2	.60	1.39	-0.18	1.37	2.33	3.57	8.55	2.71	2.14
3	.54	1.06	-0.69	1.61	3.18	4.48	3.19	7.85	12.82
4	.67	1.61	-0.15	1.29	2.20	3.37	7.07	2.95	3.27
5	.77	2.41	0.41	1.22	1.88	2.54	3.69	10.35	4.55
6	.71	1.84	-0.18	0.84	1.64	2.95	13.51	11.50	7.65
7	.63	1.45	-0.39	1.02	1.97	3.17	9.60	6.67	3.18
8	.74	2.06	0.05	1.17	2.24	3.31	3.69	4.12	4.00
9	.76	2.24	0.06	1.02	1.63	2.57	2.24	5.56	5.75
10	.82	2.84	0.04	1.09	1.79	2.64	3.26	3.65	2.18
11	.64	1.50	-0.17	1.31	2.38	3.33	5.16	3.90	8.09
12	.72	1.95	-0.09	1.24	2.08	3.32	3.10	1.75	6.36
13	.63	1.55	-0.58	0.96	2.01	3.43	7.70	3.28	4.84
14	.53	1.09	-0.88	1.36	2.43	3.56	5.29	1.42	4.95
15	.65	1.51	-0.78	0.71	1.93	3.74	21.31	3.85	4.93
16	.66	1.61	-0.45	1.06	1.84	3.27	4.11	8.97	9.81
17	.70	1.66	-0.56	0.66	1.61	3.22	14.97	14.00	7.66
a	.24	0.46	-3.83	-0.80	2.95	6.92			
b ^a	.01	—	—	—	—	—			
c	.24	0.35	-5.88	-1.71	0.31	5.38			
d	.45	0.84	-1.92	0.60	2.23	4.56			
e	.41	0.81	-1.41	0.33	1.71	4.39			
f	.35	0.63	-2.37	0.87	2.44	4.66			
g	.44	0.81	-1.67	1.47	3.27	5.00			
h	.31	0.50	-3.89	-0.93	0.90	4.70			
i	.37	0.73	-1.64	1.70	3.23	5.13			
j	.29	0.51	-4.37	-0.10	2.38	5.52			
k	.43	0.84	-2.06	1.04	2.94	4.75			
l	.44	0.87	-0.75	0.71	2.39	4.07			
m	.13	0.28	-6.41	-1.50	2.71	7.28			
n	.39	0.79	-2.23	0.69	2.72	4.41			

Note. χ^2_{LR} = value of the likelihood ratio (LR) DIF test comparing models of retained items (identified by numbers instead of letters) with item slope and location parameters freely estimated compared with a model with all slope and location parameters constrained to equality across university and MTurk samples, males and females, or ethnic majority and minority groups ($n = 1,243$). Dashes indicate that the parameter was not estimated in the model.

^a This item was excluded from the item response theory (IRT) model due to the extremely low factor loading in the exploratory factor analysis (EFA). Presented factor loadings are based on an EFAs of the polychoric correlation matrix in the university samples ($n = 946$). Slope and location parameters are from a graded-response IRT model in the university samples ($n = 946$).

were excluded from consideration. Then, the remaining items were ordered in declining magnitude of slope values, and the distribution of slope values was examined for apparent discontinuities (i.e., sudden decreases in discrimination), with the aim of reducing the scale to 20 or fewer items. Four location parameters (b) per item were available in the model results, but were less useful for item selection because most items with adequate discrimination provided more information at higher levels of latent spitefulness.

Following selection of a final set of Spitefulness Scale items, differential item functioning (DIF) was tested across the university and MTurk samples³ and then by sex and ethnicity (ethnic majority compared with ethnic minority) using the backward selection procedure described in Kim and Yoon (2011). Specifically, we compared the fit of a baseline model with slope and location parameters constrained to be equal across groups, as well as residual variances constrained to equal one and the factor mean constrained to equal zero in the reference group for identification

purposes, with the fit of models that freed slopes and location parameters across groups one item at a time, with that item's residual variance constrained to equal one across groups for model identification. Change in model fit was determined on the basis of likelihood ratio (LR) chi-square tests conducted using the DIFFTEST option in Mplus. To account for inflation in Type I error rates related to multiple LR tests and potential misspecification in the baseline models, we adjusted the critical chi-square values using two methods that have been supported in statistical simula-

³ Because there were age differences between these two samples and because some of the items addressed school-related issues (extra credit, grading), it was possible that some items may have functioned differently across the two groups.

tions (Kim & Yoon, 2011). The Bonferroni- and Oort (1998) adjusted critical value for the 4 *df* LR test was 83.06 for the comparisons across samples, 76.98 for comparisons across groups of males and females, and 79.08 for comparisons across ethnic majority and ethnic minority participants. Following tests of DIF, latent mean differences on the Spitefulness Scale were compared across samples, sex, and ethnicity, with standardized mean difference effect sizes calculated using the formula of Hancock (2001).

To investigate relations of scores on the Spitefulness Scale to the criterion variables, zero-order correlations were calculated with bootstrapped standard errors (10,000 draws) to account for positive skew exhibited on several measures. Sample sizes for these calculations varied on the basis of the specific measures administered in each sample.

Results

IRT Analyses

Overall, EFA results supported that the IRT assumption of unidimensionality was met for the 31-item Spitefulness Scale in the combined university samples. The first four eigenvalues were 9.70, 2.47, 2.14, and 1.54, yielding a ratio of the first to second eigenvalue of 3.93, which is above the commonly used ratio of 3:1 that suggests a scale is unidimensional enough for IRT. Regarding local independence, of the 465 possible residual correlations in the one-factor CFA model, 11 were greater than the .20 rule of thumb (Yen, 1993) that can indicate local dependence. Items exhibiting local dependence tended to be items with small factor loadings on the first factor, and none exhibited large slope parameters in subsequent IRT results, suggesting that these local dependencies had minimal impact on IRT results. On the basis of slope (*a*) parameters in the graded-response IRT model, 17 items were retained for the final Spitefulness Scale. There was a clear break between the lowest *a* parameter for a retained item (1.06) and the highest *a* parameter for a nonretained item (.87). Factor loadings based on the EFA results and *a* and *b* parameters for the 31 items are presented in Table 1.⁴ Internal consistency of the retained items was very high, with an estimated ordinal alpha (Zumbo, Gadermann, & Zeisser, 2007) of .94 in the university samples.

DIF and Mean Differences

None of the 17 items demonstrated LR values greater than the adjusted critical value of 83.06 (see Table 1), indicating no DIF by sample. The largest LR value was 21.31. Internal consistency was identical in the MTurk sample, with an ordinal alpha of .94. In an ordinal CFA model with all *a* and *b* values constrained to equality across university and MTurk samples, spitefulness factor scores were significantly lower ($p < .001$, $d = .34$) in the MTurk sample ($n = 297$, $M = -.34$, $SD = 1.18$) than university samples ($n = 946$, $M = .00$, $SD = .94$). For the analyses of DIF by sex, no item LR values exceeded the adjusted critical value of 76.97; the largest LR value was 14.00. With all *a* and *b* values constrained to equality across men and women, men ($n = 307$, $M = .00$, $SD = .89$) scored significantly higher ($p < .001$, $d = .30$) on latent spitefulness than women ($n = 936$, $M = -.28$, $SD = .91$). Similarly, no item LR values indicated DIF by ethnicity; the largest LR value was 12.82, well short of the adjusted critical value

of 79.08. With *a* and *b* values constrained to equality, ethnic majority (i.e., Caucasian) participants ($n = 928$, $M = -.27$, $SD = .91$) scored significantly lower ($p < .001$, $d = .30$) on latent spitefulness than ethnic minority participants ($n = 315$, $M = .00$, $SD = .87$). In sum, there was no evidence of DIF for any Spitefulness Scale item across university or MTurk samples, by sex or by ethnicity.

Demographic Differences in Spitefulness

Not only did men score higher in spitefulness when both samples were combined, but similar significant sex differences were found when the university and MTurk samples were analyzed separately. Because of the restricted age range in the university sample, we only examined the association between age and spitefulness in the MTurk sample. Older individuals reported being less spiteful than younger individuals, $r(295) = -.27$, $p < .001$. Furthermore, when age was included as a covariate, the difference in spitefulness between the university and MTurk samples was no longer significant, $F(1, 1235) = 1.96$, $p = .16$, $\eta_p^2 = .002$.

Criterion-Related Validity

Aggression and the dark triad. As expected, spitefulness was highly correlated with aggression in both the university ($r = .52$) and MTurk samples ($r = .58$). Spitefulness was also significantly correlated with all of the facets of psychopathy assessed by both the SRP-III and the LSRP, with correlations ranging from .22 to .71. Most notable were the very large correlations with the SRP-III Callous Affect subscale. Thus, individuals who lack empathy and kindness also report high levels of spitefulness. In fact, in both samples, the correlations between the Spitefulness Scale and the SRP-III Callous Affect subscale were significantly larger than correlations between the Spitefulness Scale and any of the other three SRP-III subscales (in the university sample, t s between 3.97 and 10.83, all p s $< .001$; in the MTurk sample, t s between 2.43 and 6.06, p s between .016 and .001). In contrast, although the SRP-III Erratic Lifestyle subscale was significantly associated with psychopathy, this correlation was only .22 in the university sample, which was significantly smaller than the correlations with any of the SRP-III subscales (t s between 6.17 and 10.83, all p s $< .001$). In the MTurk sample, the correlation that spitefulness had with Erratic Lifestyle was significantly smaller than the correlations it had with Interpersonal Manipulation, $t(294) = 4.13$, $p < .001$, or Callous Affect, $t(294) = 6.06$, $p < .001$, but not Criminal Tendencies, $t(294) = 1.37$, $p = .17$. For the LSRP, spitefulness was more strongly associated with Primary Psychopathy than with Secondary Psychopathy, $t(387) = 3.40$, $p = .001$. Spiteful behavior typically involves premeditation and deliberate actions (i.e., callousness, primary psychopathy), whereas the Erratic Lifestyle subscale of the SRP-III and the Secondary Psychopathy scale of the LSRP appear to capture impulsive and thrill-seeking tendencies, so it makes sense that this facet of psychopathy yielded smaller correlations with the Spitefulness Scale.

⁴ Full results of IRT analyses are only presented for the combined university samples. In an IRT analysis of the community sample only, no items excluded based on the university sample analyses had *a* values larger than the 17 retained items.

We hypothesized that spitefulness would be associated with narcissism, especially pathological narcissism. Generally, the associations between spitefulness and the narcissism scales were more modest than the correlations between spitefulness and psychopathy. Despite its questionable internal reliability, the NPI factor that was most strongly associated with spitefulness was Exploitativeness/Entitlement, which is consistent with the notion that envy and entitlement motivate spiteful behavior. In both samples, spitefulness was more strongly correlated with Exploitativeness/Entitlement than it was with either Leadership/Authority or Grandiose Exhibitionism (t s between 3.66 and 5.92, all p s < .001). Also notable were the correlations between vulnerable narcissism from the PNI and spitefulness, which demonstrates a link between fragile or pathological narcissism and spite. We were unsure about whether Machiavellianism would be associated with spitefulness, but these correlations were actually quite robust (.46 and .40 in the university and MTurk samples, respectively). We may have underestimated the deliberative and manipulative nature of spitefulness (at least as assessed with our Spitefulness Scale), which is also consistent with the correlations between spitefulness and Interpersonal Manipulation from the SRP-III and Primary Psychopathy from the LSRP.

Shame and guilt. As predicted, guilt-proneness was negatively associated with spitefulness. On the basis of the correlations reported in Table 2, contrary to our hypothesis, it initially appears that shame-proneness is unrelated to spitefulness. However, because shame and guilt are related negative emotions and the TOSCA guilt-proneness and shame-proneness scales are highly correlated (.49 and .50 in the university and MTurk samples, respectively), research with the TOSCA has typically used partial correlations to assess shame-free guilt and guilt-free shame (e.g., Stuewig, Tangney, Heigel, Harty, & McCloskey, 2010). Partialing out shame did not appreciably change the correlations between guilt and spitefulness ($-.56$ and $-.58$ in the university and MTurk samples, respectively, both partial correlations were significant at $p < .001$). However, unlike the zero-order correlations between shame-proneness and spitefulness, guilt-free shame had significant positive associations with spitefulness (.27 and .38 in the university and MTurk samples, respectively, both partial correlations were significant at $p < .001$).

General personality traits and psychological adjustment. As expected, spitefulness was negatively associated with agreeableness, regardless of whether this trait was assessed with the BFI ($r = -.40$) or the HEXACO-60 ($r = -.37$). It also appears that spitefulness is associated with low levels of conscientiousness and low self-esteem. Furthermore, individuals higher in spitefulness also report higher levels of psychological distress on the BSI. Spitefulness was not closely related to BFI Neuroticism ($r = .08$) or HEXACO-60 emotionality ($r = -.13$).

Discussion

Spite is an understudied construct in psychology, and until the current study, there has not been a self-report measure to assess individual differences in spitefulness. The primary aim of the current study was to develop a psychometrically sound measure of spitefulness, and the 17-item Spitefulness Scale appears promising. Although the items in the scale describe a wide variety of situations, including politics, work and salary, academics, physical

Table 2
Correlations of Spitefulness Scale Scores With Criterion Variables in University and MTurk Samples

Criterion	University		MTurk	
	<i>r</i>	<i>n</i>	<i>r</i>	<i>n</i>
Aggression	.52	946	.58	297
SRP-III Callous Affect	.65	556	.71	297
SRP-III Erratic Lifestyle	.22	556	.43	297
SRP-III Interpersonal Manipulation	.48	556	.61	297
SRP-III Criminal Tendencies	.53	556	.50	297
LSRP Primary Psychopathy	.57	390	—	—
LSRP Secondary Psychopathy	.43	390	—	—
NPI Leadership/Authority	.11**	946	.06, <i>ns</i>	297
NPI Grandiose Exhibitionism	.14	946	.18**	297
NPI Exploitativeness/Entitlement	.33	946	.39	297
PNI Grandiosity	.16	946	.37	297
PNI Vulnerability	.27	946	.44	297
Machiavellianism	.47	946	.40	297
TOSCA Guilt-Proneness	-.51	946	-.48	297
TOSCA Shame-Proneness	-.05, <i>ns</i>	946	.06, <i>ns</i>	297
BFI Agreeableness	-.40	390	—	—
BFI Extraversion	-.02, <i>ns</i>	390	—	—
BFI Conscientiousness	-.33	390	—	—
BFI Neuroticism	.08, <i>ns</i>	390	—	—
BFI Openness	-.10*	390	—	—
HEXACO Honesty-Humility	—	—	-.46	297
HEXACO Emotionality	—	—	-.13**	297
HEXACO Extraversion	—	—	-.18**	297
HEXACO Agreeableness	—	—	-.37	297
HEXACO Openness	—	—	-.35	297
HEXACO Conscientiousness	—	—	-.50	297
Self-Esteem	-.23	946	-.32	297
BSI Global Severity Index	.44	390	—	—

Note. All correlations are $p < .001$ unless otherwise indicated. MTurk = Amazon's Mechanical Turk; SRP-III = Self-Report Psychopathy Scale; LSRP = Levenson Self-Report Psychopathy Scale; NPI = Narcissistic Personality Inventory; PNI = Pathological Narcissism Inventory; TOSCA = Test of Self-Conscious Affect; BFI = Big Five Inventory; BSI = Brief Symptom Inventory. Dashes indicate that the criterion measure was not administered in the sample.

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

conflict, and problems with neighbors, the Spitefulness Scale was unidimensional and internally consistent, suggesting that it measures a single coherent construct. The 17 items in the scale had slope parameters that had moderate to very high levels of discrimination, indicating that the items appropriately distinguished between individuals with low and high levels of latent spitefulness. Although college students scored higher on the scale than an MTurk sample, men scored higher than women, and ethnic minority members scored higher than ethnic majority members, it appears that the items functioned similarly across both university and MTurk samples, across both men and women, and across ethnic minority and ethnic majority members.

Given the dearth of psychological research on individual differences in spitefulness, our hypotheses regarding the associations between spitefulness and other personality traits were drawn from diverse sources including economic research and psychoanalytic case studies. The pattern of correlations with these criterion variables was generally as we predicted. Most of the expected correlations were large enough to support our hypotheses, but not so large as to suggest that the Spitefulness Scale was inadvertently measuring an identical construct as one of the criterion variables.

For example, the correlations with the Aggression Scale (.52 and .58 in the university and MTurk samples, respectively) demonstrated the expected association between spitefulness and aggression, but the magnitude of these associations indicated that these scales are capturing distinct constructs. Many of the largest correlations with the Spitefulness Scale were with the psychopathy measures, especially those facets of psychopathy that involved callousness, a lack of empathy, manipulateness, and exploitiveness, which was consistent with the finding that a dismissing avoidant attachment style predicted spiteful rejections in an ultimatum game (Almakias & Weiss, 2012). Furthermore, individuals higher in spitefulness reported lower levels of agreeableness and conscientiousness, which is also characteristic of psychopathy (Lynam & Derefinko, 2006). In contrast, spitefulness was not as strongly correlated with those facets of psychopathy that involve impulsivity and erratic behavior.

Spitefulness was associated with Machiavellianism and negatively correlated with the TOSCA guilt-proneness scale. Although we were uncertain as to whether individuals high in Machiavellianism would be willing to incur harm to themselves, the spiteful tactics described in the Spitefulness Scale may help them achieve their long-term goals. This association between spitefulness and Machiavellianism was also consistent with Stern's (2004) clinical observation that spiteful patients "are capable of observing what others feel and need but only to exploit them" (p. 658). Spitefulness was also associated with both shame and narcissism, especially those aspects of narcissism that involve vulnerability and exploitiveness (a characteristic shared by narcissism, Machiavellianism, and psychopathy). That spitefulness was also negatively correlated with self-esteem and positively associated with general psychological distress suggests a possible connection between fragile self-esteem and spitefulness. Given that many of the traits that most strongly correlated with spitefulness were also traits that men endorse more than women and that tend to decrease with age, it was not unexpected that younger men scored the highest on the Spitefulness Scale.

Limitations and Future Directions

Unlike other "dark" personality traits, such as psychopathy, Machiavellianism, and narcissism, spitefulness has received little empirical attention in the psychological literature. Perhaps this lack of attention is because there are currently no self-report measures of spitefulness. In contrast, multiple measures are available to assess these other dark personality traits. Our aim in constructing this scale is to encourage research on spitefulness and its relation to both psychopathology and social behavior. For example, although spiteful behavior is a symptom of ODD, these behaviors are not actually defined or specified in the *DSM-5* (APA, 2013). This ambiguity may account for why factor analyses of the ODD symptoms have yielded inconsistent findings, with some researchers finding that spiteful behavior loads on a negative affect factor (Burke, Hipwell, & Loeber, 2010) and others finding it loads on either a headstrong (Rowe, Costello, Angold, Copeland, & Maughan, 2010) or conduct-disordered factor (Lahey et al., 2004). Given the increasing interest in ODD in adulthood (Harpold et al., 2007), and the changes to the *DSM-5* criteria to better identify adult ODD (e.g., changing "adults" to "authority figures"), a spitefulness measure may contribute to the diagnosis of ODD.

Furthermore, items from our scale may serve as the basis for a child and adolescent measure of spitefulness.

Given the substantial correlations between the Spitefulness Scale and both psychopathy and narcissism, research with this scale may help inform whether spitefulness should be considered a symptom or associated feature of either narcissistic or antisocial personality disorder. Does the assessment of spitefulness improve the accuracy of diagnosing either of these personality disorders? Although we did not measure borderline personality traits in the current study, the degree to which self-harm in borderline personality disorder includes a spiteful component may also merit study.

The Spitefulness Scale may also prove useful in social psychological and behavioral economic research. The few studies that have attempted to identify personality traits or affective states that could predict behavior in ultimatum games have yielded inconsistent results. Perhaps a scale that assesses spitefulness directly will be a better predictor of behavior in these competitive economic games. More broadly, spite does not occur only in financial situations, so we hope that this Spitefulness Scale may prove useful for studying a wide range of social behaviors, including aggressive behavior (e.g., Will high scorers on the Spitefulness Scale hurt themselves in order to hurt a competitor?), daily interactions (e.g., Do high scorers go out of their way to inconvenience others? Are they more litigious and less willing to settle cases?), and perhaps even political preferences (e.g., Do individuals who support policies that hurt the poor and the general economy also score high in spitefulness?). Although few would consider spite a desirable motive, it is likely an important one. Perhaps the development of this Spitefulness Scale will serve as an impetus for further study of this neglected trait.

Although a strength of the study was that we cross-validated findings from a large college student sample, using data from an older nonstudent sample, the study is not without limitations. Most notably, both samples were largely Caucasian and female (the MTurk sample somewhat less so than the university sample). Additionally, given the associations between spitefulness and psychopathy, narcissism, low self-esteem, and general psychological distress, future research with the Spitefulness Scale should be conducted with clinical samples. Finally, because we needed to collect sufficient data for the IRT analysis and DIF analyses, we relied entirely on self-report data, which could have resulted in mono-method bias. Future research that includes informant reports and behavioral tasks (e.g., ultimatum games) may contribute to the further validation of this Spitefulness Scale.

References

- Ackerman, R. A., Witt, E. A., Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., & Kashy, D. A. (2011). What does the Narcissistic Personality Inventory really measure? *Assessment, 18*, 67–87. doi: 10.1177/1073191110382845
- Almakias, S., & Weiss, A. (2012). Ultimatum Game behavior in light of attachment theory. *Journal of Economic Psychology, 33*, 515–526. doi: 10.1016/j.joep.2011.12.012
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340–345. doi:10.1080/00223890902935878

- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
- Benet-Martínez, V., & John, O. P. (1998). *Los Cinco Grandes* across cultures and ethnic groups: Multitrait multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75*, 729–750. doi:10.1037/0022-3514.75.3.729
- Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis, 20*, 351–368. doi:10.1093/pan/mpr057
- Blascovich, J., & Tomaka, J. (1991). Measures of self-esteem. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (Vol. 1, pp. 115–160). New York, NY: Academic Press. doi:10.1016/B978-0-12-590241-0.50008-3
- Brown, R. P., Budzek, K., & Tamborski, M. (2009). On the meaning and measure of narcissism. *Personality and Social Psychology Bulletin, 35*, 951–964. doi:10.1177/0146167209335461
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3–5. doi:10.1177/1745691610393980
- Burke, J. D., Hipwell, A. E., & Loeber, R. (2010). Dimensions of oppositional defiant disorder as predictors of depression and conduct disorder in preadolescent girls. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*, 484–492.
- Christie, R., & Geis, R. L. (1970). *Studies in Machiavellianism*. New York, NY: Academic Press.
- Cooter, R., Marks, S., & Mnookin, R. (1982). Bargaining in the shadow of the law: A testable model of strategic behavior. *Journal of Legal Studies, 11*, 225–251. doi:10.1086/467699
- Critchfield, K. L., Levy, K. N., Clarkin, J. F., & Kernberg, O. F. (2008). The relational context of aggression in borderline personality disorder: Using adult attachment style to predict forms of hostility. *Journal of Clinical Psychology, 64*, 67–82. doi:10.1002/jclp.20434
- Cullis, J., Jones, P., & Soliman, A. (2012). 'Spite effects' in tax evasion experiments. *Journal of Socio-Economics, 41*, 418–423. doi:10.1016/j.socec.2011.05.011
- Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13*, 595–605. doi:10.1017/S0033291700048017
- Dunn, B. D., Makarova, D., Evans, D., & Clark, L. (2010). "I'm worth more than that": Trait positivity predicts increased rejection of unfair financial offers. *PLoS ONE, 5*, e15095. doi:10.1371/journal.pone.0015095
- Falk, A., Fehr, E., & Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica, 73*, 2017–2030. doi:10.1111/j.1468-0262.2005.00644.x
- Fehr, E., & Fischbacher, U. (2005). The economics of strong reciprocity. In H. Gintis, S. Bowles, R. Boyd, & E. Fehr (Eds.), *Moral sentiments and material interests: The foundations of cooperation in economic life* (Vol. 6, pp. 151–191). Cambridge, MA: MIT Press.
- Gambetta, D. (2005). Can we make sense of suicide missions? In D. Gambetta (Ed.), *Making sense of suicide missions* (pp. 259–299). Oxford, England: Oxford University Press. doi:10.1093/acprof:oso/978019276998.003.0008
- Gottlieb, R. M. (2004). Refusing the cure: Sophocles's Philoctetes and the clinical problems of self-injurious spite, shame and forgiveness. *International Journal of Psychoanalysis, 85*, 669–690. doi:10.1516/A750-YQQL-NB4C-LLGC
- Hamburger, H. (1979). *Games as models of social phenomena*. San Francisco, CA: W.H. Freeman.
- Hamilton, W. D. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature, 228*, 1218–1220. doi:10.1038/2281218a0
- Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika, 66*, 373–388. doi:10.1007/BF02294440
- Hare, R. D. (2003). *Manual for the Revised Psychopathy Checklist* (2nd ed.). Toronto, ON, Canada: Multi-Health Systems.
- Harlé, K. M., Allen, J. J., & Sanfey, A. G. (2010). The impact of depression on social economic decision making. *Journal of Abnormal Psychology, 119*, 440–446. doi:10.1037/a0018612
- Harlé, K. M., & Sanfey, A. G. (2007). Incidental sadness biases social economic decisions in the Ultimatum Game. *Emotion, 7*, 876–881. doi:10.1037/1528-3542.7.4.876
- Harlé, K. M., & Sanfey, A. G. (2010). Effects of approach and withdrawal motivation on interactive economic decisions. *Cognition & Emotion, 24*, 1456–1465. doi:10.1080/02699930903510220
- Harpold, T., Biederman, J., Gignac, M., Hammerness, P., Surman, C., Potter, A., & Mick, E. (2007). Is oppositional defiant disorder a meaningful diagnosis in adults? Results from a large sample of adults with ADHD. *Journal of Nervous and Mental Disease, 195*, 601–605. doi:10.1097/NMD.0b013e318093f448
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory – Versions 4a and 54*. Berkeley: University of California, Berkeley, Institute of Personality and Social Research.
- Johnston, J. R. (2003). Parental alignments and rejection: An empirical study of alienation in children of divorce. *Journal of the American Academy of Psychiatry and the Law, 31*, 158–170.
- Joiner, T. (2010). *Myths about suicide*. Cambridge, MA: Harvard University Press.
- Joiner, T., & Rudd, M. D. (Eds.). (2000). *Suicide science: Expanding the boundaries*. Boston, MA: Springer.
- Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling, 18*, 212–228. doi:10.1080/10705511.2011.557337
- Kimbrough, E. O., & Reiss, J. P. (2012). Measuring the distribution of spitefulness. *PLoS ONE, 7*, e41812. doi:10.1371/journal.pone.0041812
- Kirchsteiger, G. (1994). The role of envy in ultimatum games. *Journal of Economic Behavior and Organization, 25*, 373–389. doi:10.1016/0167-2681(94)90106-6
- Krizan, Z., & Johar, O. (2012). Envy divides the two faces of narcissism. *Journal of Personality, 80*, 1415–1451. doi:10.1111/j.1467-6494.2012.00767.x
- Lahey, B. B., Applegate, B., Waldman, I. D., Loft, J. D., Hankin, B. L., & Rick, J. (2004). The structure of child and adolescent psychopathology: Generating new hypotheses. *Journal of Abnormal Psychology, 113*, 358–385. doi:10.1037/0021-843X.113.3.358
- Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology, 68*, 151–158. doi:10.1037/0022-3514.68.1.151
- Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics, 1*, 593–622. doi:10.1006/redy.1998.0023
- Little, T. D., Henrich, C. C., Jones, S. M., & Hawley, P. H. (2003). Disentangling the 'whys' from the 'whats' of aggressive behaviour. *International Journal of Behavioral Research, 27*, 122–133. doi:10.1080/01650250244000128
- Lynam, R. R., & Derefinko, K. J. (2006). Psychopathy and personality. In C. J. Patrick (Ed.), *Handbook of psychopathy* (pp. 133–155). New York, NY: Guilford Press.
- Mahmut, M. K., Menictas, C., Stevenson, R. J., & Homewood, J. (2011). Validating the factor structure of the Self-Report Psychopathy Scale in a community sample. *Psychological Assessment, 23*, 670–678. doi:10.1037/a0023090

- McHoskey, J. W., Worzel, W., & Szyarto, C. (1998). Machiavellianism and psychopathy. *Journal of Personality and Social Psychology, 74*, 192–210. doi:10.1037/0022-3514.74.1.192
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.
- Oort, F. J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 5*, 107–124. doi:10.1080/10705519809540095
- Patrick, C. J., Fowles, D. C., & Krueger, R. F. (2009). Triarchic conceptualization of psychopathy: Developmental origins of disinhibition, boldness, and meanness. *Development and Psychopathology, 21*, 913–938. doi:10.1017/S0954579409000492
- Paulhus, D. L., Neumann, C. S., & Hare, R. D. (in press). *Manual for the Self-Report Psychopathy Scale* (4th ed.). Toronto, ON, Canada: Multi-Health Systems.
- Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism and psychopathy. *Journal of Research in Personality, 36*, 556–563. doi:10.1016/S0092-6566(02)00505-6
- Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes, 68*, 208–224. doi:10.1006/obhd.1996.0100
- Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G. C., & Levy, K. N. (2009). Initial construction and validation of the Pathological Narcissism Inventory. *Psychological Assessment, 21*, 365–379. doi:10.1037/a0016530
- Prahlw, J. A., Long, S., & Barnard, J. J. (1998). A suicide disguised as a homicide: Return to Thor Bridge. *The American Journal of Forensic Medicine and Pathology, 19*, 186–189. doi:10.1097/00000433-199806000-00018
- Raskin, R. N., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports, 45*, 590. doi:10.2466/pr0.1979.45.2.590
- Raskin, R. N., & Hall, C. S. (1981). The Narcissistic Personality Inventory: Alternate form reliability and further evidence of construct validity. *Journal of Personality Assessment, 45*, 159–162. doi:10.1207/s15327752jpa4502_10
- R Development Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Ricolfi, L. (2005). Palestinians, 1981–2003. In D. Gambetta (Ed.), *Making sense of suicide missions* (pp. 77–129). Oxford, England: Oxford University Press. doi:10.1093/acprof:oso/9780199276998.003.0003
- Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*, 1–25.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Ross, J., Irani, I., Silberman, M., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers?: Shifting demographics in Amazon Mechanical Turk. In the *Proceedings of the CHI EA 2010* (pp. 2863–2872).
- Rowe, R., Costello, E. J., Angold, A., Copeland, W. E., & Maughan, B. (2010). Developmental pathways in oppositional defiant disorder and conduct disorder. *Journal of Abnormal Psychology, 119*, 726–738. doi:10.1037/a0020798
- Scott, E. S. (1992). Pluralism, parental preference, and child custody. *California Law Review, 80*, 615–672. doi:10.2307/3480710
- Shabad, P. (2000). Giving the devil his due: Spite and the struggle for individual dignity. *Psychoanalytic Psychology, 17*, 690–705. doi:10.1037/0736-9735.17.4.690
- Stern, J. (2004). Of “sympathy for the devil” and a question of conscious versus unconscious spite: Comment on Shabad (2000). *Psychoanalytic Psychology, 21*, 655–661. doi:10.1037/0736-9735.21.4.655
- Stuewig, J., Tangney, J. P., Heigel, C., Harty, L., & McCloskey, L. (2010). Shaming, blaming, and maiming: Functional links among the moral emotions, externalization of blame, and aggression. *Journal of Research in Personality, 44*, 91–102. doi:10.1016/j.jrp.2009.12.005
- Tangney, J. P., & Dearing, R. L. (2002). Assessing shame and guilt. In J. P. Tangney & R. L. Dearing (Eds.), *Shame and guilt* (pp. 26–51). New York, NY: Guilford Press.
- Tangney, J. P., Dearing, R. L., Wagner, P. E., & Gramzow, R. (2000). *The Test of Self-Conscious Affect-3 (TOSCA-3)*. Fairfax, VA: George Mason University.
- Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology, 58*, 345–372. doi:10.1146/annurev.psych.56.091103.070145
- Williams, K. M., Paulhus, D. L., & Hare, R. D. (2007). Capturing the four-factor structure of psychopathy in college students via self-report. *Journal of Personality Assessment, 88*, 205–219. doi:10.1080/00223890701268074
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*, 125–145. doi:10.1177/014662168400800201
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*, 187–213. doi:10.1111/j.1745-3984.1993.tb00423.x
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*, 21–29.

Appendix

Spitefulness Scale

1. It might be worth risking my reputation in order to spread gossip about someone I did not like.
2. If I am going to my car in a crowded parking lot and it appears that another driver wants my parking space, then I will make sure to take my time pulling out of the parking space.
3. I hope that elected officials are successful in their efforts to improve my community even if I opposed their election. (reverse scored)
4. If my neighbor complained that I was playing my music too loud, then I might turn up the music even louder just to irritate him or her, even if meant I could get fined.

(Appendix continues)

5. If I had the opportunity, then I would gladly pay a small sum of money to see a classmate who I do not like fail his or her final exam.
 6. There have been times when I was willing to suffer some small harm so that I could punish someone else who deserved it.
 7. I would rather no one get extra credit in a class if it meant that others would receive more credit than me.
 8. If I opposed the election of an official, then I would be glad to see him or her fail even if their failure hurt my community.
 9. I would be willing to take a punch if it meant that someone I did not like would receive two punches.
 10. I would be willing to pay more for some goods and services if other people I did not like had to pay even more.
 11. If I was one of the last students in a classroom taking an exam and I noticed that the instructor looked impatient, I would be sure to take my time finishing the exam just to irritate him or her.
 12. If my neighbor complained about the appearance of my front yard, I would be tempted to make it look worse just to annoy him or her.
 13. I would take on extra work at my job if it meant that one of my co-workers who I did not like would also have to do extra work.
 14. I would be happy receiving extra credit in a class even if other students received more points than me. (reverse scored)
 15. Part of me enjoys seeing the people I do not like fail even if their failure hurts me in some way.
 16. If I am checking out at a store and I feel like the person in line behind me is rushing me, then I will sometimes slow down and take extra time to pay.
 17. It is sometimes worth a little suffering on my part to see others receive the punishment they deserve.
- b. I would try to be quick if I was the last student taking an exam so that I would not inconvenience the instructor. (reverse scored)
 - c. I would consider tapping on my brakes to scare a driver who was tailgating me.
 - d. If my co-workers were going to get larger raises than me, then I would prefer it if none of us received raises.
 - e. I can think of times when I have intentionally done something to hurt someone else even though I knew it would also hurt me somehow.
 - f. I would probably remove a humorous cartoon from my desk if one of my co-workers found it offensive. (reverse scored)
 - g. I would be happy with getting a raise even if my co-workers received raises that were bigger than mine. (reverse scored)
 - h. If I notice a car driving too close behind me, I will sometimes make a point to slow down in order to irritate the other driver.
 - i. I would try to be as fast as possible at an ATM if the person behind me looked as if he or she was in a hurry. (reverse scored)
 - j. If I notice a car too close behind me, I would make it easy for the other driver to pass me because he or she must be in a hurry. (reverse scored)
 - k. I would try to clean up my yard if my neighbor complained about it. (reverse scored)
 - l. I would oppose funding for a government program that provided illegal immigrants with free vaccinations for a contagious disease, even if this program would reduce the chances of a population-wide epidemic.
 - m. I would not be willing to suffer a little just to see someone else receive punishment they deserved. (reverse scored)
 - n. Even if I was not selected for a team, I would still support the team and be happy if they were successful. (reverse scored)

Excluded Items

- a. I would oppose a law that cut taxes for me if I found out that other people would receive a larger tax break than me.

Received July 9, 2013
Revision received December 31, 2013
Accepted January 15, 2014 ■