

# Measuring Hostile Interpretation Bias: The WSAP-Hostility Scale

Assessment  
2016, Vol. 23(6) 707–719  
© The Author(s) 2015  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/1073191115599052  
asm.sagepub.com



Kirsten H. Dillon<sup>1</sup>, Nicholas P. Allan<sup>1</sup>, Jesse R. Cogle<sup>1</sup>, and Frank D. Fincham<sup>1</sup>

## Abstract

Hostile interpretation biases are central to the development and maintenance of anger, yet have been inconsistently assessed. The Word Sentence Association Paradigm (WSAP) was used to develop a new measure of hostile interpretation biases, the WSAP-Hostility. Study 1 examined the factor structure and internal consistency of the WSAP-Hostility, as well as its relationship with trait anger. Study 2 provided convergent and divergent validity data by examining its associations with trait anger, aggression, depression, and anxiety. Study 3 examined the relationship between WSAP-Hostility and another measure of hostile interpretation biases, as well as another word sentence association measure, in a sample of community participants. Study 4 also used a sample of community participants to offer further evidence of convergent validity. Across the studies, the WSAP-Hostility demonstrated convergent and divergent validity and internal consistency, supporting its use as a measure of hostile interpretation biases.

## Keywords

anger, hostility, aggression, hostile interpretation bias

Trait anger is associated with numerous negative consequences including cardiovascular disease (Smith, Glazer, Ruiz, & Gallo, 2004; Williams et al., 2000), aggressive behavior (Berkowitz, 1993; Deffenbacher, 1993; Tafrate, Kassino, & Dundin, 2002), nicotine dependence (Cogle, Zvolensky, & Hawkins, 2013), excessive alcohol use (Deffenbacher, 1993; Litt, Cooney, & Morse, 2000), relationship problems (Baron et al., 2007; Miller, Markides, Chiriboga, & Ray, 1995; Tafrate et al., 2002), and increased suicide risk (Hawkins & Cogle, 2013a; Hawkins et al., 2014). Furthermore, trait anger is associated with many psychological disorders (e.g., intermittent explosive disorder, major depression, posttraumatic stress disorder, borderline personality disorder). Thus, anger is a significant clinical problem which is worth assessing and understanding further.

## The Hostile Attribution Bias

Cognitive models of anger propose that individuals with high trait anger possess a cognitive processing bias that makes them more likely to interpret ambiguous situations as hostile and less likely to adopt a benign interpretation (see Wilkowski & Robinson, 2010). Wilkowski and Robinson (2010) recently argued that hostile interpretations are a necessary link between hostile situations and the elicitation of anger and subsequent reactive aggression. Individuals with high levels of trait anger have a greater

tendency to react aggressively to provocation (Bettencourt, Talley, Benjamin, & Valentine, 2006). In fact, according to Wilkowski and Robinson (2010), anger is the necessary link between hostile interpretation and reactive aggression. The tendency for angry and aggressive individuals to perceive hostile intent in ambiguous situations, also known as the hostile attribution bias, was originally documented by Dodge (1980). Dodge found that when aggressive and non-aggressive boys were exposed to frustrating situations in which a peer had behaved with ambiguous intent the aggressive boys were likely to respond as though the peer had displayed hostile intent. However, the nonaggressive boys reacted as though the peer had possessed benign intent. Thus, without cues to indicate otherwise, aggressive boys were interpreting ambiguous situations as hostile.

Since the 1980s, a large body of research has documented the hostile attribution bias in children and adolescents (see Orobio de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002), and more recently this bias has also been documented in adult samples (Epps & Kendall, 1995; Hazebroek, Howells, & Day, 2001; Wenzel & Lystad, 2005). A meta-analysis of 41 studies found a significant

<sup>1</sup>Florida State University, Tallahassee, FL, USA

## Corresponding Author:

Jesse R. Cogle, Department of Psychology, Florida State University,  
P.O. Box 3064301, Tallahassee, FL 32306, USA.  
Email: coglej@psy.fsu.edu

association (weighted mean effect size  $r = .17$ ) between hostile attribution biases and aggressive behavior (Orobio de Castro et al., 2002).

**Measurement Issues.** The methods used to measure hostile attribution bias have varied greatly. Most studies present individuals with ambiguous scenarios that could be interpreted as either hostile or benign and have used a variety of modes to present these situations (video, audio, text, pictures, or staging), a variety of types of situations, different response options (rating scales, open-ended responses, etc.), and different levels of personal involvement in the situation (spectator vs. first person). Additionally, the methods used to score (open answer vs. multiple choice) and calculate bias (e.g., proportion of items with hostile intent selected, difference between hostile and benign attributions, etc.) have been inconsistent across studies. Existing measures may also be limited by the inclusion of a small number of items and may not be ideal for repeated administration.

Thus, even though the hostile interpretation bias is considered central to the development and maintenance of anger (Wilkowski & Robinson, 2010), it is assessed inconsistently making comparisons across studies difficult. Within the field of anger research and treatment, there is therefore a need to develop a standard self-report questionnaire for measuring and tracking hostile interpretations of ambiguous situations that can help establish the foundation for a more cohesive and cumulative literature. One such measure of this bias is the Social Information Processing-Attribution and Emotional Response Questionnaire (SIP-AEQ; Coccaro, Noblett, & McCloskey, 2009). This measure, though found to be reliable and valid, is somewhat cumbersome to be used as a quick measure of hostile interpretations, as it requires participants to read eight short stories and then answer a series of questions that pertain to each story. Thus, there is a need for a more efficient method of assessing hostile interpretations that is psychometrically sound. Such a measure could be used to examine the effectiveness of anger treatments that target cognitive biases, such as cognitive behavioral therapy and to determine whether reduction of hostile attribution biases mediates the effects of cognitive behavioral treatments on anger reduction. This is particularly important as treatments for anger are very heterogeneous and little is known about the mechanisms by which specific factors of treatment reduce specific aspects of anger (DiGiuseppe & Tafrate, 2003). Thus, the development of a standard measure to assess interpretation biases could help elucidate the mechanisms underlying anger reduction and may thereby lead to more parsimonious and effective treatment protocols.

**A Lesson From Anxiety Research?** Recently, there has been an increased focus on developing assessments to detect

interpretation biases prevalent in anxious populations (Amir, Prouvost, & Kuckertz, 2012; Beard & Amir, 2008; Kuckertz, Amir, Tobin, & Najmi, 2013). One method that has been used is the Word Sentence Association Paradigm (WSAP; Beard & Amir, 2008). This paradigm was initially created as a computerized reaction time task (Beard & Amir, 2008), but has more recently been modified to be used as a scale to assess biases (see Kuckertz et al., 2013). To assess biases, participants are presented with ambiguous sentences and either threat or benign words. They are then instructed to rate the similarity of the word and the sentence. Thus, this method can be used to calculate a threat interpretation score, a benign interpretation score, and a bias score (the difference between threat and benign scores).

The WSAP has been used to assess interpretation biases associated with obsessive-compulsive symptoms and is able to both differentiate between individuals with and without obsessive-compulsive symptoms and predict behavioral approach on a contamination task (Kuckertz et al., 2013). The WSAP has also been used to differentiate between individuals with and without social anxiety disorder (Amir et al., 2012).

The progress facilitated by the existence of the WSAP in understanding anxiety is noteworthy and prompts the question of whether a similar approach might be used to measure interpretation bias in regard to anger. To explore this possibility, the current studies examine the use of the WSAP to assess the hostile interpretation bias. We developed the WSAP-Hostility and tested its psychometric properties in four separate studies. We predicted that scores on the WSAP-Hostility would be uniquely related to trait anger and other anger-relevant variables (aggression, hostility, anger expression, and anger control).

## Study I

The goals of the present study were to examine the underlying structure of the WSAP-Hostility, refine the scale, document its internal consistency, and examine its relationship with trait anger.

## Method

**Participants and Procedure.** Participants were recruited through introductory courses at a large southeastern university and completed this study as partial fulfillment of course requirements. After giving informed consent, participants completed a battery of online questionnaires. The sample consisted of 517 participants (82.8% female) ranging in age from 18 to 44 years ( $M = 19.51$ ,  $SD = 2.0$ ), and consisted of the following ethnic groups: White (69.4%), Black or African American (10.4%), Hispanic (14.3%), Asian or Pacific Islander (2.5%), American Indian or Alaskan Native (0.4%), and other (2.9%).

## Measures

*State-Trait Anger Expression Inventory–2 (STAXI-2; Spielberger, 1999)*. The trait subscale of the STAXI-2 was used to measure trait anger. The trait anger subscale of the STAXI-2 is composed of 10 items. The STAXI-2 has been found to demonstrate good reliability and validity (Spielberger, 1999). In a college sample, it correlates highly with the Buss–Durkee Hostility Inventory (males = .71 and females = .66) and (Minnesota Multiphasic Personality Inventory) MMPI hostility ( $H_0$ , males = .59 and females = .43, see Spielberger, 1999). Internal consistency in the present sample was  $\alpha = .89$ . Furthermore, the scale yielded a *T*-score of 50 for the sample mean.<sup>1</sup>

*The Word Sentence Association Paradigm for Hostility (WSAP-Hostility)*. The WSAP-Hostility was adapted from the Word Sentence Association Test for OCD (WSAO; Kuckertz et al., 2013) and consists of distinct ambiguous sentences (e.g., “Someone is in your way”), followed by either a hostility-related word (e.g., “inconsiderate”) or a benign word (e.g., “unaware”). These sentences were phrased in such a way that the participant was meant to be an active participant in the scenario described; therefore, general versus specific relationships were referenced in each situation in an effort to be inclusive. Additionally, each scenario depicted a situation that was potentially anger provoking. Thus, a number of these ambiguous situations could be presented to the participant in order to quickly assess their general tendency to make a hostile versus a benign interpretation. Participants were asked to rate how similar the sentence and the word were on a scale of 1 (*not at all similar*) to 6 (*extremely similar*). This response scale was selected, in part, to dissuade participants from simply selecting a “neutral” (neither similar nor dissimilar) rating and, thus increased variability in responses. Additionally, by asking participants to rate the similarity between sentences and words of either hostile or benign valence, rather than asking them to answer a question such as “How angry would this situation make you?”, we were able to limit response bias and potentially obtain a more immediate assessment of their tendency to ascribe hostile versus benign intent to various situations. Each sentence was presented twice non-consecutively, once with the hostility-related word and once with the benign word. Next, average ratings for the hostile and benign words were calculated to yield two subscales (hostile and benign).<sup>2</sup>

Initially, 40 sentences were created (each with both a hostile and benign word pair). These sentences were generated by researchers familiar with the anger literature and situations which would tend to provoke anger in individuals with high levels of trait anger. In an effort to be as inclusive as possible of ambiguous situations that may lead to hostile interpretations, the experimenters developed a list of themes of anger provocation with guidance from Novaco’s Provocation Inventory

(2003). Themes used in the sentences included perceived unfairness, feeling ignored, disrespected, argued with, unappreciated, or that others are angry, thinking others are stealing from you, driving-related situations, physical encounters, and annoying traits of others. Pilot testing was conducted with these 80 word–sentence pairs and item-total correlations were examined to determine which scenarios to retain in the final measure. Seven sentences were removed due to poor item-total correlations and lack of variability in responses. Thus, in the present study, 33 sentences (66 items total) were used for further analysis.

Pilot testing of the WSAP-Hostility on 31 undergraduate students found the measure was relatively brief to complete (it took participants roughly 6.5 minutes to complete the measure, range: 3.5–9 minutes). Furthermore, the WSAP-Hostility was included in a larger study using an unselected sample of undergraduate students to collect test–retest reliability data with administrations 1 month apart and test–retest reliability was measured as  $r = .65$  (see Hawkins, Macatee, Guthrie, & Cougle, 2013; Macatee, Capron, Schmidt, & Cougle, 2013, for more information about this study).

## Results and Discussion

*Exploratory Factor Analysis and Item Response Theory Analysis for Scale Refinement*. A two-step approach was conducted for developing a brief and informative WSAP-Hostility measure. The first step involved the use of exploratory factor analysis (EFA) to remove item pairs that failed to show unidimensionality within each item. The second step involved using item response theory (IRT; Lord, 1980; Lord & Novick, 1968) to eliminate poorly discriminating items, redundant items, and to ensure that the WSAP-Hostility captured a broad trait range (referred to as ability level or  $\theta$  in IRT; Embretson & Reise, 2000).

To examine the factor structure of the 66 WSAP-Hostility items EFA was conducted in *Mplus* version 7.31 (1998–2012) using the geomin oblique rotation. The data were treated as categorical, using robust weighted least squares estimator, to account for the ordinal nature of the data (Flora & Curran, 2004). The purpose of the EFA was to eliminate item pairs that did not load on separate (presumably Hostile and Benign) factors and retain item pairs that loaded on separate factors and also produced low cross-loadings. As suggested by Tabachnick and Fidell (2001), loadings of .32 or higher were considered substantive. However, it was decided to retain an item pair if a hostile item loaded uniquely on the Hostile factor and the paired benign item loaded highest on the Benign factor with a cross-loading not on the Hostile factor. This approach was taken as it was in line with the goal of creating a scale maximizing the measurement of a hostile attribution bias. Examination of the scree plot revealed a significant bend in the elbow at the four-factor solution. Furthermore, model fit indices, including the comparative fit index

(CFI = .91), Tucker–Lewis index (.90), and root mean square error of approximation (RMSEA = .04; 90% confidence interval [.04, .05]) were within generally accepted rule-of-thumb estimates of acceptable fit (Bentler, 1990; Browne & Cudeck, 1993). Highlighting the essential independence of the Hostile (Factor 1) and Benign (Factor 2) factors, the correlation between these factors was  $-.10$ . Model parameters are provided for the four-factor EFA in Table S1 (all supplemental materials available online at <http://asm.sagepub.com/content/by/supplemental-data>). Using the aforementioned approach for scale reduction, 19-item pairs were retained.

IRT analyses (Embretson & Reise, 2000) were then conducted on the Hostile and Benign factors separately. Graded response models (Samejima, 1969) were fit to the data as the responses in the WSAP-Hostility scale are polytomous. The graded response model provides a single discrimination ( $a$ ) parameter, which can be calculated directly from *Mplus* using theta parameterization, or indirectly by dividing the factor loading of the item by the square root of the residual variance of the item (Brown, 2015). This model also provides difficulty ( $b$ ) parameters equal to  $n - 1$ , where  $n$  is the number of possible response options. These parameters were computed indirectly using *Mplus*-provided factor loadings and item thresholds (Brown, 2015). Trait levels, or  $\theta$ , are standardized such that mean trait level is 0 and an increase of 1 represents an increase of 1 standard deviation ( $SD$ ) across the trait spectrum.

The discrimination parameter indicates how well the item distinguishes between individuals with varying levels of the trait of interest (i.e., hostile or benign interpretation). Although there are no agreed on benchmarks for acceptable discrimination parameters, higher discrimination parameters are considered better. In line with Baker (2001), we considered discrimination parameters of .65 or higher as indicating at least moderate discrimination and parameters below this as indicating low to no discrimination. Again in line with maximizing hostile interpretation bias, we prioritized removing items from the Hostile factor with low-discrimination parameters. Using this criteria, six items were identified with  $a$  parameters below .60 (i.e., Items 1, 2, 6, 42, 53, and 59 corresponding to benign Items 30, 12, 36, 49, 39, and 51, respectively). Whereas two items had  $a$  parameters below the .65 threshold, they were each above .63, and were retained. Only one item from the Benign factor had an  $a$  parameter well below the .65 threshold (i.e., Item 47), and this item and its corresponding item pair (Item 55) were removed.

The resulting Hostile and Benign factors comprised 18 items each (see Table 1). These factors were examined for model fit and to determine whether they captured information acceptably across hostile and benign traits, respectively. Regarding model fit, the Hostile ( $\chi^2 = 542.36$ ,  $p < .001$ ,

CFI = .91, RMSEA = .09) and Benign ( $\chi^2 = 542.36$ ,  $p < .001$ , CFI = .91, RMSEA = .11) factors provided low to adequate model fit, although examination of modification

indices did not reveal any modifications that could improve model fit. Regarding the information captured by the Hostile and Benign factors, using the  $a$ ,  $b$ , and  $\theta$  parameters, item information functions can be calculated to show the amount of information obtained from an item. In turn, item information functions can be averaged to provide a test information function (TIF) and corresponding standard errors. When a scale is being developed to capture a broad trait range, a TIF should be produced that covers a broad range of a particular trait (here we focused on  $\pm 3 SD$ ) and therefore look relatively flat across the range of the trait. Furthermore, as a demonstration of precision across this range, standard error values (calculated as the inverse square root of the TIF) should be below .5 (Hambleton, Swaminathan, & Rogers, 1991; Nguyen, Han, Kim, & Chan, 2014). Examination of the TIFs (see Figure 1a) and standard errors of the TIFs (see Figure 1b) for the hostile and benign scales revealed that the hostile scale captured similar levels of information across the ability spectrum. Furthermore, this information was captured with precision, as the standard errors remained below .5. For the most part, the benign scale also captured similar levels of information across the ability spectrum, although somewhat less information was captured at high levels of the benign scale, as demonstrated by the drop-off in information from 2  $SD$ s above the mean; however, even with this drop-off in information captured, an acceptable level of precision was present as the standard errors remained below .5 even above 2  $SD$ s from the mean.

**Internal Consistency and Convergent Validity.**<sup>3</sup> Internal consistencies for the new 32-item scale were  $\alpha = .90$  for the benign words and  $\alpha = .87$  for the hostility-related words. Table S2 shows the means and standard deviations for all study variables. Zero-order correlations were computed between average hostile word ratings, average benign word ratings, and STAXI-2 trait anger (see Table 2). Trait anger was significantly associated with hostile and benign word ratings. This study shows that the WSAP-Hostility is a reliable measure for assessing hostile interpretations and provides initial evidence on its convergent validity.

**Gender Differences.** Analyses of variance (ANOVAs) were performed to examine gender differences across the WSAP-Hostility subscales. We found evidence of gender differences on the ratings of benign words, such that females rated similarity of benign words more highly,  $F_{(1, 468)} = 11.00$ ,  $p < .001$ . Differences in hostile word ratings were not significantly different,  $F_{(1, 468)} = 0.05$ ,  $p = .83$ . Next, we sought to examine whether gender moderated the relationship between WSAP-Hostility and trait anger. Separate regressions were run (one for each WSAP-Hostility subscale: hostile words and benign words). There was a significant interaction between gender and hostile word ratings in predicting trait anger ( $\beta = -.140$ ,  $p < .001$ ), but not for

**Table 1.** Item Response Theory Discrimination and Difficulty Parameters.

| Hostile words      | <i>a</i> | <i>SE</i> | <i>b</i> <sub>1</sub> | <i>b</i> <sub>2</sub> | <i>b</i> <sub>3</sub> | <i>b</i> <sub>4</sub> | <i>b</i> <sub>5</sub> |
|--------------------|----------|-----------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Insulting (11)     | 0.65     | .05       | -2.32                 | -1.06                 | -0.15                 | 0.88                  | 2.13                  |
| Abusive (14)       | 0.67     | .05       | -1.54                 | 0.10                  | 1.24                  | 2.18                  | 3.31                  |
| Inconsiderate (15) | 0.71     | .05       | -1.80                 | -0.44                 | 0.70                  | 1.55                  | 2.63                  |
| Aggressive (16)    | 0.77     | .05       | -1.80                 | -0.35                 | 0.71                  | 1.85                  | 2.92                  |
| Disrespectful (17) | 0.71     | .05       | -1.32                 | 0.25                  | 1.41                  | 2.34                  | 3.05                  |
| Inconsiderate (21) | 1.06     | .06       | -1.34                 | -0.14                 | 0.82                  | 1.70                  | 2.49                  |
| Disrespectful (26) | 0.71     | .05       | -2.32                 | -1.07                 | 0.11                  | 1.07                  | 2.26                  |
| Thoughtless (27)   | 0.63     | .05       | -2.72                 | -0.90                 | 0.55                  | 1.90                  | 3.08                  |
| Hostile (28)       | 0.68     | .05       | -1.56                 | -0.09                 | 1.14                  | 2.28                  | 3.21                  |
| Rude (29)          | 0.65     | .05       | -3.01                 | -1.74                 | -0.78                 | 0.45                  | 1.84                  |
| Rude (31)          | 0.63     | .05       | -2.72                 | -1.18                 | 0.01                  | 1.08                  | 2.44                  |
| Rude (32)          | 0.84     | .05       | -1.29                 | 0.39                  | 1.35                  | 2.14                  | 3.04                  |
| Disrespectful (40) | 0.75     | .06       | -1.40                 | 0.12                  | 1.20                  | 1.92                  | 2.82                  |
| Aggressive (48)    | 0.71     | .05       | -1.37                 | 0.24                  | 1.43                  | 2.32                  | 3.41                  |
| Inconsiderate (56) | 0.91     | .06       | -1.45                 | 0.10                  | 1.32                  | 2.23                  | 3.38                  |
| Ignoring (65)      | 0.64     | .05       | -2.04                 | -0.60                 | 0.75                  | 1.88                  | 3.08                  |
| Benign words       | <i>a</i> | <i>SE</i> | <i>b</i> <sub>1</sub> | <i>b</i> <sub>2</sub> | <i>b</i> <sub>3</sub> | <i>b</i> <sub>4</sub> | <i>b</i> <sub>5</sub> |
| Unintentional (58) | 0.88     | .05       | -2.62                 | -1.33                 | -0.20                 | 0.82                  | 0.82                  |
| Alerting (43)      | 0.75     | .05       | -3.06                 | -2.38                 | -1.07                 | 0.37                  | 0.37                  |
| Cautious (45)      | 0.57     | .04       | -3.66                 | -2.02                 | -0.54                 | 0.86                  | 0.86                  |
| Accidental (24)    | 1.20     | .07       | -2.87                 | -2.23                 | -1.24                 | -0.36                 | -0.36                 |
| Amused (52)        | 0.91     | .06       | -3.06                 | -2.22                 | -1.27                 | -0.23                 | -0.23                 |
| Unaware (23)       | 0.86     | .05       | -3.31                 | -2.00                 | -0.90                 | 0.10                  | 0.10                  |
| Unknowing (54)     | 0.80     | .05       | -2.44                 | -1.43                 | -0.32                 | 0.54                  | 0.54                  |
| Distracted (44)    | 1.12     | .06       | -2.76                 | -1.93                 | -0.79                 | 0.28                  | 0.28                  |
| Unhappy (61)       | 0.84     | .06       | -3.02                 | -2.23                 | -1.25                 | -0.18                 | -0.18                 |
| Forgetful (60)     | 0.82     | .05       | -2.73                 | -1.77                 | -0.67                 | 0.38                  | 0.38                  |
| Unnoticing (46)    | 0.74     | .05       | -2.70                 | -1.51                 | -0.34                 | 0.88                  | 0.88                  |
| Busy (8)           | 0.59     | .05       | -4.13                 | -2.80                 | -1.51                 | -0.09                 | -0.09                 |
| Accidental (33)    | 0.96     | .06       | -2.73                 | -2.07                 | -0.92                 | -0.01                 | -0.01                 |
| Accidental (50)    | 1.27     | .08       | -2.57                 | -1.87                 | -1.00                 | -0.26                 | -0.26                 |
| Unnoticing (34)    | 0.95     | .06       | -2.66                 | -1.77                 | -0.70                 | 0.30                  | 0.30                  |
| Unaware (19)       | 0.57     | .05       | -3.48                 | -2.29                 | -0.82                 | 0.52                  | 0.52                  |

Note. *a* = discrimination parameter. *b*<sub>1</sub>-*b*<sub>5</sub> = difficulty parameters.

benign word ratings ( $\beta = -.028, p = .56$ ). To interpret the significant interaction, we assessed the simple effects of hostile word ratings among male and female participants. We found that the relationship between hostile word ratings with trait anger was greater among men ( $\beta = .537, p < .001$ ) than women ( $\beta = .190, p < .001$ ). Thus, even though there were significant associations between hostile word ratings and trait anger for both genders, this relationship was stronger for males.

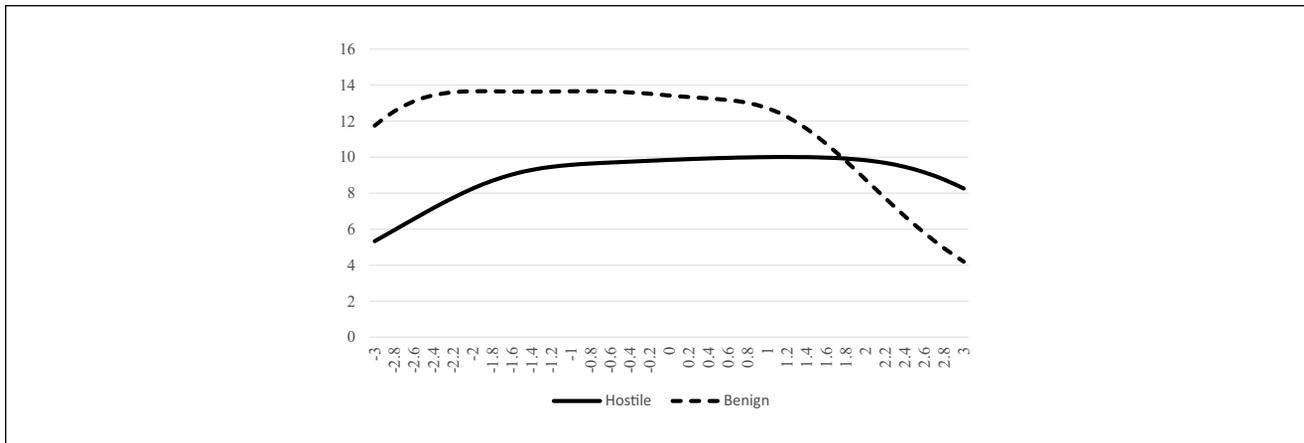
## Study 2

In this study, we sought to replicate the WSAP-Hostility and trait anger association and provide further data on convergent validity, including self-reported aggression. In

doing so, we took the precaution of controlling for anxiety and depression to ensure that the relationship between hostile interpretation bias and anger-related variables was not better explained by negative affect, as research has demonstrated that depression, anxiety, and anger are associated with higher order negative affectivity (Watson & Clark, 1992). Additionally, we tested the divergent validity of the WSAP-Hostility by examining the relative strength of the relationship between the WSAP-Hostility and trait anger as opposed to depression or anxiety.

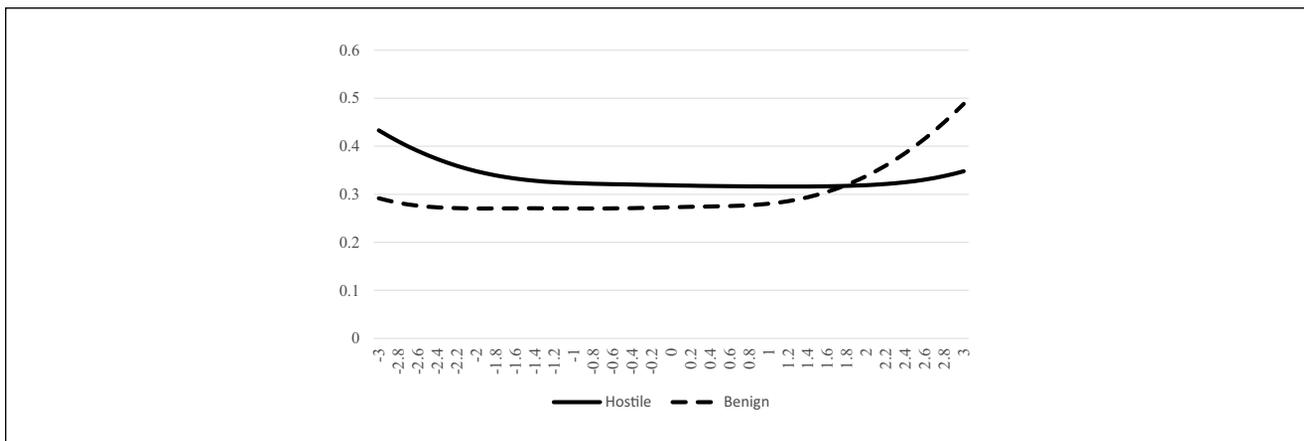
## Method

**Participants and Procedure.** Participants were recruited through introductory psychology courses at a large



**Figure 1a.** Test information function for hostile and benign scales.

Note. Theta ( $\theta$ ) is on the x-axis and test information is on the y-axis. Each scale appears to provide a broad range of information across the trait level.



**Figure 1b.** Standard errors of the test information functions for the hostile and benign scales.

Note. Theta ( $\theta$ ) is on the x-axis and the standard error of the test information function is on the y-axis. As demonstrated by standard errors below .5 across the spectrum, both scales operate with precision at the range of  $\pm 3$  standard deviations from the mean.

southeastern university and completed this study as partial fulfillment of course requirements. The sample consisted of 100 participants (68% female) ranging in age from 18 to 25 years ( $M = 18.98$ ,  $SD = 1.4$ ), and from the following ethnic groups: White (62%), Hispanic (17%), African American (6%), Asian or Pacific Islander (7%), American Indian or Alaskan Native (2%), and other (6%).

Participants completed questionnaires as part of a larger study. After giving informed consent, participants completed all self-report measures in one sitting, individually, via computer.

#### Measures

*The Word Sentence Association Paradigm for Hostility (WSAP-Hostility)*. See Study 1, for a full description of this measure. The 32-item scale derived in Study 1 was used in the present study. In the present sample, internal consistencies

were measured at  $\alpha = .88$  for the benign words and  $\alpha = .90$  for the hostility-related words.

*State-Trait Anger Expression Inventory-2 (STAXI-2; Spielberger, 1999)*. See Study 1, for a full description of this measure. In the present sample, internal consistency was  $\alpha = .86$ .

*The Buss-Perry Aggression Questionnaire (BPAQ; Buss & Perry, 1992)*. The BPAQ is a 29-item self-report measure of aggression that yields four subscales of aggressive behavior: physical aggression, verbal aggression, anger (physiological arousal), and hostility (cognitive component underlying anger and aggression). Participants were asked to rate how characteristic each item is of them on a scale of 1 (*extremely uncharacteristic of me*) to 7 (*extremely characteristic of me*). In the present sample, internal consistencies

**Table 2.** Convergent and Discriminant Correlations for Hostile and Benign Subscales of the WSAP-Hostility Across Studies 1 to 4.

| Study 1 (n = 469) |                     |               |              |              |                |                    |                 |          |
|-------------------|---------------------|---------------|--------------|--------------|----------------|--------------------|-----------------|----------|
|                   | STAXI-2 Trait Anger |               |              |              |                |                    |                 |          |
| WSAP Hostile      | .26***              |               |              |              |                |                    |                 |          |
| WSAP Benign       | -.10*               |               |              |              |                |                    |                 |          |
| Study 2 (n = 100) |                     |               |              |              |                |                    |                 |          |
|                   | STAXI-2 Trait Anger | BPAQ Physical | BPAQ Verbal  | BPAQ Anger   | BPAQ Hostility | DASS-21 Depression | DASS-21 Anxiety |          |
| WSAP Hostile      | .44***              | .16           | .15          | .30**        | .33***         | .19                | .21*            |          |
| WSAP Benign       | -.21*               | -.08          | -.08         | -.22*        | -.16           | -.19               | -.16            |          |
| Study 3 (n = 183) |                     |               |              |              |                |                    |                 |          |
|                   | CM-Hostility        | HA            | BA           | IA           | NER            | WSAO Threat        | WSAO Benign     |          |
| WSAP Hostile      | .17**               | .22**         | -.13         | .31***       | .24***         | .38***             | .26***          |          |
| WSAP Benign       | -.27***             | .23**         | .31***       | -.32***      | -.08           | .02                | .32***          |          |
| Study 4 (n = 215) |                     |               |              |              |                |                    |                 |          |
|                   | STAXI-2 Trait Anger | STAXI-2 AX-O  | STAXI-2 AX-I | STAXI-2 AC-O | STAXI-2 AC-I   | CM-Hostility       | PANAS-NA        | PANAS-PA |
| WSAP Hostile      | .39***              | .25***        | .30***       | -.35***      | -.22***        | .30***             | .30***          | .05      |
| WSAP Benign       | -.22**              | -.26***       | -.10         | .39***       | .32***         | -.37***            | -.25**          | .14      |

Note. WSAP = Word Sentence Association Paradigm; STAXI-2 = State-Trait Anger Expression Inventory-2; BPAQ = Buss-Perry Aggression Questionnaire; DASS-21 = Depression Anxiety Stress Scale-21; CM-Hostility = Cook-Medley Hostility Scale, 17 item; HA = Hostile Attribution; BA = Benign Attribution; IA = Instrumental Attribution; NER = Negative Emotional Response; WSAO = Word Sentence Association Test for OCD; AX-O = Anger Expression-Out; AX-I = Anger Expression-In; AC-O = Anger Control-Out; AC-I = Anger Control-In; PANAS-NA = Positive and Negative Affect Scale-Negative Affect Subscale; PANAS-PA = Positive and Negative Affect Scale-Positive Affect Subscale. \*\*\* $p < .001$ . \*\* $p < .01$ . \* $p < .05$ .

were as follows for each subscale, physical:  $\alpha = .86$ ; verbal:  $\alpha = .82$ ; anger:  $\alpha = .79$ ; hostility:  $\alpha = .87$ .

*Depression Anxiety Stress Scale-21 (DASS-21; Lovibond & Lovibond, 1995).* The DASS-21 is a self-report questionnaire that assesses symptoms of depression, anxiety, and stress over the past week. Participants were asked to rate how much each of 21 statements applied to them in the past week on a scale of 0 (*did not apply to me at all*) to 3 (*applied to me very much, or most of the time*). For the current study only the depression and anxiety subscales were used. Internal consistencies for these subscales in our study were  $\alpha$ 's = .86 (depression) and .76 (anxiety).

## Results and Discussion

Table S3 displays the means and standard deviations for all study variables. Zero-order correlations were computed to examine associations between average hostile word ratings, average benign word ratings, and STAXI-2 trait anger, BPAQ subscales, and DASS-21 depression and anxiety (see

Table 2). Next, partial correlations were computed between these measures using depression and anxiety as covariates (see Table S3). Trait anger and the anger and hostility scales of the BPAQ were each associated with hostile word ratings when covarying depression and anxiety. Interestingly, these scales were not related to benign word rating scores, suggesting that trait anger and hostility are driven by a tendency toward hostile interpretation rather than a lack of benign interpretation. WSAP-Hostility was not significantly correlated with self-reported physical or verbal aggression as measured by the BPAQ. However, hostile interpretation bias is more likely to be associated with reactive (anger-driven) aggression than proactive (goal-directed) aggression and the BPAQ does not differentiate between these forms of aggression. The association between WSAP-Hostility and self-reported aggression may have been stronger if we had used a measure of reactive aggression. Additional research is necessary to investigate this further.

Hierarchical regression analyses were conducted to examine the unique contribution of trait anger to WSAP-Hostility

scores (hostile and benign), when controlling for depression and anxiety. Depression and anxiety were entered as predictor variables in the first step and trait anger was entered in the second step. Two separate regressions were conducted to predict hostile word ratings and benign word ratings, respectively. For hostile word ratings, the addition of trait anger accounted for significantly more variance (15% more variance,  $F$  change = 17.81,  $p < .001$ ) than the model that only included depression and anxiety. In the regression predicting benign word ratings, the addition of trait anger did not account for significantly more variance over and above depression and anxiety ( $F$  change = 2.17,  $p = .14$ ). These findings support the divergent validity of the WSAP-Hostility hostile subscale.

**Gender Differences.** ANOVAs were performed to examine gender differences across the WSAP-Hostility subscales. We found a significant gender difference on the ratings of hostile words, such that females rated similarity of hostile words more highly,  $F_{(1, 99)} = 4.37, p < .05$ . Interestingly, this result was inconsistent with the gender differences found in Study 1 and may be an artifact of lower sample size (there were only 32 males in the current study). Differences in benign word ratings were not significantly different,  $F_{(1, 99)} = 1.49, p = .23$ . We did not find evidence of an interaction between gender and WSAP-Hostility subscales in the prediction of trait anger ( $p$  values: .79-.99).

### Study 3

Studies 1 and 2 examined the use of the WSAP-Hostility with student samples. To test the generalizability of these results, Study 3 examined the WSAP-Hostility in a community sample. Additionally, Study 3 investigated the relationship between the WSAP-Hostility and another measure of hostile interpretation bias, the SIP-AEQ (Coccaro et al., 2009). The SIP-AEQ yields several subscales (hostile attribution [HA], benign attribution [BA], instrumental attribution [IA], and negative emotional response [NER]). We were particularly interested in examining the associations between each of these two scales and trait hostility, as well as the associations between the WSAP-Hostility and the SIP-AEQ. In particular, we were interested in examining the relationship between the HA, BA, and IA subscales of the SIP-AEQ and the hostile and benign subscales of the WSAP-Hostility. Based on their conceptual similarity, we predicted that the HA and IA subscales of the SIP-AEQ would be correlated with the hostile subscale of the WSAP-Hostility and the BA subscale of the SIP-AEQ would be correlated with the benign subscale.

As a test of the divergent validity of the WSAP-Hostility, we also sought to investigate the relationship between the WSAP-Hostility and another validated scale that uses the word sentence association paradigm to assess interpretation

bias, the WSAO (Kuckertz et al., 2013). We hypothesized that the WSAO and the WSAP-Hostility would be correlated, but that the WSAP-Hostility would be more highly correlated with trait hostility than the WSAO.

### Method

**Participants and Procedure.** Participants were recruited using Mechanical Turk, an Internet service that facilitates data collection from large samples (Buhrmester, Kwang, & Gosling, 2011). Interested participants completed consent online, followed by a questionnaire battery. Next, participants were given a code to enter the Mechanical Turk website in order to receive payment for their participation. To control for order effects, participants were randomly assigned to complete either the WSAP-Hostility or the SIP-AEQ first, followed by the other measures.

The sample consisted of 183 participants (51% female;  $M_{\text{age}} = 36.77; SD = 11.33$ ). Participants were ethnically and racially diverse (47.0% Asian or Pacific Islander, 37.7% non-Hispanic White, 6.6% non-Hispanic Black, 6% Hispanic, 1.1% American Indian or Alaskan Native, 1.6% other). The sample had varying levels of education (52.5% had a bachelor's degree, 22.4% had a postgraduate degree, 17.5% had at least some college education, 7.1% had a high school diploma, and 0.5% had not graduated from high school).

### Measures

**The Word Sentence Association Paradigm for Hostility (WSAP-Hostility).** See Study 1, for a complete description of this measure. Again, the 32-item scale from Study 1 was used. In the present sample internal consistency was  $\alpha = .87$  for the benign words, and  $\alpha = .83$  for the hostility-related words.

**Social Information Processing-Attribution and Emotional Response Questionnaire (SIP-AEQ; Coccaro, Noblett, & McCloskey, 2009).** The SIP-AEQ consists of eight written vignettes that depict socially ambiguous situations in which an adverse action (e.g., physical pain or rejection) is directed at the main character. Following each vignette there are six Likert-type scaled questions that assess direct hostile intent, indirect hostile intent, instrumental nonhostile intent, benign intent, and two items assessing NER (e.g., anger) on a 0 (*not at all likely*) to 3 (*very likely*) scale. The scale yields four subscales: HA, BA, IA, and NER. Internal consistencies in the present sample were as follows:  $\alpha = .98$  for HA,  $\alpha = .96$  for BA,  $\alpha = .96$  for IA, and  $\alpha = .64$  for NER.

**The Word Sentence Association Test for OCD (WSAO; Kuckertz et al., 2013).** The WSAO is comprised of 20 ambiguous OC-related sentences. Half of these sentences are followed by an OC-related threat word and half are

followed by a benign word. Participants are then asked to rate the similarity between the word and the sentence on a scale of 1 (*not at all related*) to 7 (*very much related*). As with the WSAP-Hostility, average ratings for the threat and benign words are calculated and used to determine an interpretation bias score (subtracting benign word ratings from threat word ratings). In the present sample, internal consistency was  $\alpha = .62$  for the threat words and  $\alpha = .73$  for the benign words.

*Cook–Medley Hostility Scale, 17 item (CM-Hostility; Cook & Medley, 1954).* Trait hostility was assessed with an abbreviated 17-item version of the full Cook–Medley Hostility Scale. The scale uses a “true–false” format to assess statements reflecting interpersonal distrust, guardedness, and expectations of deceit (e.g., “Most people are honest chiefly because they are afraid of being caught”). “True” responses are summed to create a total score. This short version of the scale is highly correlated with the full scale ( $r = .93$ ) and has demonstrated reliability across subgroups (Strong, Kahler, Greene, & Schinka, 2005). In the current sample, internal consistency was  $\alpha = .83$ .

## Results and Discussion

ANOVA tests were conducted to determine whether responses to the WSAP-Hostility and SIP-AEQ differed based on the order in which the scales were presented. There were no significant differences found for any of the subscales, based on the order of administration ( $ps = .14-.84$ ). Table S4 displays the means and standard deviations for all study variables used. Zero-order correlations were performed between the WSAP-Hostility subscales, CM-Hostility, SIP-AEQ subscales, and WSAO subscales (see Table 2 for correlations with WSAP subscales and Table S5 for all intercorrelations).

We found that both WSAP-Hostility subscales were significantly correlated with CM-Hostility, which is further evidence for the scales convergent validity. All SIP-AEQ subscales, except HA, were significantly correlated with CM-Hostility. The hostile word ratings from the WSAP-Hostility were positively correlated with HA and IA, as we predicted. The correlation with BA was negative, but non-significant. Benign word ratings were modestly and positively correlated with HA, positively correlated with BA, and negatively correlated with HA. Overall, the associations between the two measures support the convergent validity of the WSAP-Hostility as a measure of hostile interpretation biases. Furthermore, the WSAP-Hostility was more strongly associated with trait hostility (measured by CM-Hostility) than the SIP-AEQ. Despite some significant associations between the WSAP-Hostility and the WSAO subscales, the correlations were modest, which suggests divergence between the scales.

*Gender Differences.* ANOVAs were performed to examine gender differences across the WSAP-Hostility subscales. We did not find evidence of significant gender differences on either of the WSAP-Hostility subscales ( $p$  values: .10-.18). We did not find evidence of an interaction between gender and WSAP-Hostility subscales in the prediction of trait anger ( $p$  values: .17-.51).

## Study 4

Study 4 also used a community sample to investigate the relationship between the WSAP-Hostility and trait anger and hostility. Additionally, we sought to examine which aspects of anger (e.g., anger expression vs. control) were related to WSAP-Hostility.

## Method

*Participants and Procedure.* As in Study 3, participants were recruited using Mechanical Turk. The sample was originally collected as part of another study in which current and former smokers were oversampled. Fifty-three percent of the sample were daily smokers, 15.9% occasional smokers, 14.9% former smokers, and 16.3% had never smoked. Interested participants completed consent online, followed by a questionnaire battery. Next, participants were given a code to enter the Mechanical Turk website in order to receive payment for their participation.

The sample comprised 215 participants (46% female;  $M_{age} = 36.21$ ;  $SD = 11.89$ ). Participants were ethnically and racially diverse (63.7% non-Hispanic White, 31.6% Asian or Pacific Islander, 0.9% non-Hispanic Black, 0.5% Hispanic, 0.5% American Indian or Alaskan Native, 1.9% other). The sample had varying levels of education (30.7% had a bachelor’s degree, 24.7% had at least some college education, 25.6% had a master’s degree, 9.3% had a high school degree or GED, 7.4% had a 2-year college degree, 0.9% had a doctoral degree, 0.9% had a professional degree, JD or MD, and 0.5% had not graduated from high school).

## Measures

*Cook–Medley Hostility Scale, 17 item (CM-Hostility; Cook & Medley, 1954).* See Study 3, for a complete description of this measure. In the current sample, internal consistency was  $\alpha = .84$ .

*The Word Sentence Association Paradigm for Hostility (WSAP-Hostility).* See Study 1, for a complete description of this measure. Again, the 32-item scale from Study 1 was used. Internal consistency in the present sample was  $\alpha = .90$  for the benign words and  $\alpha = .88$  for the hostility-related words.

*State-Trait Anger Expression Inventory-2 (STAXI-2; Spielberger, 1999)*. The STAXI-2 was used to measure trait anger as well as several aspects of anger experience. The measure assesses maladaptive ways of coping with anger, including the tendency to suppress anger expression (AX-I) and the tendency to express anger outwardly in an aggressive manner (AX-O). The anger control subscales assess adaptive coping strategies, including the tendency to calm oneself internally (AC-I) and the tendency to prevent the outward expression of anger (AC-O). In the present sample, internal consistency for the subscales ranged between  $\alpha = .80-.92$ .

*The Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988)*. This is a 20-item scale in which participants are asked to rate the extent to which they generally experience specific negative and positive emotions on a 5-point scale ranging from 1 (*very slightly or not at all*) to 5 (*very much*). The ratings of the negative and positive emotions are summed separately to form the negative and positive affect subscales (PANAS-NA and PANAS-PA, respectively). In the current sample, internal consistency for PANAS-NA was  $\alpha = .93$  and PANAS-PA was  $\alpha = .91$ .

## Results and Discussion

Table S5 displays the means and standard deviations for all study variables used. Zero-order correlations were performed among average hostile word ratings, average benign word ratings, STAXI-2 subscales, trait hostility, PANAS-NA, and PANAS-PA (see Table 2). Next, partial correlations were conducted between these measures in which PANAS-NA served as a covariate (see Table S5).

Study 4 extended the previous findings by examining the associations between the WSAP-Hostility and trait hostility and different aspects of anger, including expression and control, in a sample of participants from the community. Internal consistency for the WSAP-Hostility was again excellent. WSAP-Hostility was significantly correlated with trait anger, trait hostility, and negative affect, suggesting convergent validity. Furthermore, positive affect was not significantly correlated with WSAP-Hostility, suggesting divergent validity. All subscales except anger expression outward were associated with hostile word ratings and all subscales except trait anger and anger expression inward were associated with benign word ratings. The lack of relationship between trait anger and benign word ratings is similar to what we found in Study 2.

**Gender Differences.** ANOVAs were performed to examine gender differences across the WSAP-Hostility subscales. We found evidence of gender differences on the ratings of benign words, such that females rated similarity of benign words more highly,  $F_{(1, 214)} = 13.86, p < .001$ . Differences in

hostile word ratings were not significantly different,  $F_{(1, 214)} = 2.67, p = .10$ . These findings were similar to those of Study 1. Additionally, there was a significant interaction between gender and hostile word ratings in predicting trait anger ( $\beta = .13, p < .05$ ). To interpret this finding, we assessed the simple effects of hostile word ratings among male and female participants. We found that the relationship between hostile word ratings and trait anger was greater among women ( $\beta = .51, p < .001$ ) than men ( $\beta = .25, p < .01$ ), which was the opposite of what we had found in Study 1 and suggests that the effects of gender may be inconsistent.

## General Discussion

The present set of studies evaluated a new measure of hostile interpretation bias, the WSAP-Hostility. As hypothesized, we found that the WSAP-Hostility was consistently associated with trait anger and additional anger-relevant variables including aggression, hostility, anger expression, and anger control. In Study 3, we examined the associations between the WSAP-Hostility and another measure of hostile interpretation bias, the SIP-AEQ, and found that the WSAP-Hostility was more consistently and strongly related to trait hostility, and that this relationship remained significant when controlling for SIP-AEQ subscales. Additionally, we examined the relationship between the WSAP-Hostility and another word sentence association measure, the WSAO, and found that, though the scales were related, this correlation was moderate, which supports the divergent validity of our scale. Furthermore, in Studies 2 and 4, we were able to examine the unique relationship between the WSAP-Hostility and anger-relevant variables, by covarying symptoms of depression and anxiety and general negative affect. These results suggest that the relationship between WSAP-Hostility and anger-relevant variables is not better explained by these variables. Across the studies we found evidence of gender effects, suggesting that the relationship between WSAP-Hostility and anger-related variables may be stronger for males.

An interesting pattern emerged between the hostile and benign subscales. Generally, hostile word ratings were more consistently associated with anger-relevant variables than benign word ratings. This was especially true for trait anger, suggesting that trait anger is driven by a tendency toward hostile interpretation rather than a lack of benign interpretation.

In Study 3, we compared the WSAP-Hostility with the SIP-AEQ, an existing measure of hostile interpretation bias. Interestingly, despite being designed to measure ostensibly similar constructs, the correlations between these two measures were modest. There are several possible explanations for this divergence. Method variance is one such explanation, as the procedures for each of the assessments are quite different from each other and different ambiguous scenarios

are used. One further explanation for the difference between these measures is that, whereas the SIP-AEQ asks participants specific questions about their interpretations of the scenarios presented (e.g., Why do you think . . . happened?), the WSAP-Hostility assesses interpretations more indirectly by asking participants to rate similarities between words and sentences. In this respect, the WSAP-Hostility is more like an implicit measure of hostile interpretation bias, whereas the SIP-AEQ is an explicit measure. The modest correlation between these measures is consistent with findings of low correlations between implicit and explicit measures (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). This set of studies offers several methodological strengths. First, the use of four separate studies with consistent findings provides support for the WSAP-Hostility as a reliable measure of hostile attribution bias. Second, we examined relationships between the WSAP-Hostility and multiple measures of anger and hostility. Third, by covarying depression and anxiety in Study 2 and negative affectivity in Study 4, we were able to examine the unique relationship between WSAP-Hostility and anger-relevant variables and rule out the possibility that this relationship was better accounted for by these symptoms. Fourth, we were able to compare our measure with an existing measure of hostile interpretation bias and found evidence of its convergent validity. Fifth, we compared our measure with another word sentence association paradigm that assesses a different kind of bias (obsessive–compulsive interpretations) and found evidence of its divergent validity.

There are also several limitations in the current set of studies. In two of the four studies, undergraduate student samples were used. Future research should examine the use of the WSAP-Hostility in wider range of populations, including clinical and treatment-seeking samples. The current studies were all cross-sectional and correlational. Thus, the direction of effects between WSAP-Hostility and anger is unclear. Further studies should be conducted using longitudinal and experimental designs to examine the relationship between WSAP-Hostility and related variables over time. The current studies all relied on self-report measures, and future research may wish to examine the relationship between WSAP-Hostility and other assessments of anger and aggression (e.g., behavioral measures) to address concerns over common method variance. The Cook–Medley 17-item Hostility inventory (Cook & Medley, 1954) was one of several measures that we used to investigate the validity of the WSAP-Hostility. This measure, while possessing significant strengths, also has several limitations (see Eckhardt, Norlander, & Deffenbacher, 2004), and future research should continue to study the relationship between the WSAP-Hostility and different measures of anger and hostility.

Study 2 did not find a relationship between the WSAP-Hostility and self-reported verbal or physical aggression.

Additional research with violent and aggressive individuals (e.g., forensic populations) is necessary to further examine the relationship between WSAP-Hostility and aggressive behavior. Last, there are inherent limitations of the approach used for the measure we developed. It was our goal to develop a quick and efficient measure of hostile interpretation bias. As with any assessment method, it is important to balance its benefits against its limitations. For example, one such limitation of the WSAP is that it uses hypothetical situations, and it is certainly possible that individuals may behave or feel quite differently in real-world situations.

The WSAP-Hostility provides a means to assess and track biases that have consistently been implicated in the development of anger (Wilkowski & Robinson, 2010). These biases have important implications, both for the individuals who hold them and those who interact with them. Additionally, these biases may also be implicated in situations in which groups of people are interacting with one another (e.g., racist attitudes, political opinions) and could have implications at the international level, potentially leading to war or peace. There is evidence that hostile interpretation biases are malleable and reductions in bias may lead to lower anger reactivity (Hawkins & Cogle, 2013b). A reliable and valid measure such as the WSAP-Hostility will be helpful to accurately track these biases to determine whether their reduction mediates the effects of cognitive behavioral treatments on anger reduction. Further research is necessary to examine the psychometric properties and utility of this instrument in clinical samples (e.g., individuals presenting for anger management treatment).

In sum, the WSAP-Hostility provides an efficient, easily administered measure of hostile interpretation bias that has the potential to serve as a standard assessment in research and clinical settings. Its adoption would promote easier comparison across studies and the development of a more coherent and cumulative literature on the role of this bias in the development and treatment of anger problems.

### Declaration of Conflicting Interests

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

### Funding

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

### Notes

1. The STAXI-2 was normed on a sample of 1,644 (977 females, 667 males); the averages (*T*-score of 50) for participants aged 20 to 29 years are 19 for males and 18 for females (Spielberger, 1999).
2. Other researchers who have used the word sentence association paradigm (e.g., Kuckertz et al., 2013) have also

calculated an interpretation bias score by subtracting the benign word rating score from the negative (or threat) word rating score. In the current set of studies, this score did not yield any differences in patterns of findings, as the bias score performed similarly to the hostile word rating score across studies. Thus, we do not report these additional analyses.

- Complete data (including measures of trait anger and depression) were not available for all 517 participants. The following analyses were conducted for a subsample of 469 participants.

## References

- Amir, N., Prouvost, C., & Kuckertz, J. M. (2012). Lack of benign interpretation bias in social anxiety disorder. *Cognitive Behaviour Therapy, 41*, 119-129. doi:10.1080/16506073.2012.662655
- Baker, F. B. (2001). *The basics of item response theory. ERIC clearinghouse on assessment and evaluation*. College Park, MD: University of Maryland.
- Baron, K. G., Smith, T. W., Butner, J., Nealey-Moore, J., Hawkins, M. W., & Uchino, B. N. (2007). Hostility, anger, and marital adjustment: Concurrent and prospective associations with psychosocial vulnerability. *Journal of Behavioral Medicine, 30*, 1-10. doi:10.1007/s10865-006-9086-z
- Beard, C., & Amir, N. (2008). A multi-session interpretation modification program: Changes in interpretation and social anxiety symptoms. *Behaviour Research and Therapy, 46*, 1135-1141. doi:10.1016/j.brat.2008.05.012
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246. doi:10.1037/0033-2909.107.2.238
- Berkowitz, L. (1993). Towards a general theory of anger and emotional aggression: Implications of the cognitive-neoassociationist perspective for the analysis of anger and emotion. In R. S. Wyer & T. K. Srull (Eds.), *Perspectives on anger and emotion* (pp. 1-46). Hillsdale, NJ: Erlbaum.
- Bettencourt, B., Talley, A., Benjamin, A. J., & Valentine, J. (2006). Personality and aggressive behavior under provoking and neutral conditions: A meta-analytic review. *Psychological Bulletin, 132*, 751-777. doi:10.1037/0033-2909.132.5.751
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- 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
- Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology, 63*, 452-459. doi:10.1037//0022-3514.63.3.452
- Coccaro, E. F., Nobilett, K. L., & McCloskey, M. S. (2009). Attributional and emotional responses to socially ambiguous cues: Validation of a new assessment of social/emotional information processing in healthy adults and impulsive aggressive patients. *Journal of Psychiatric Research, 43*, 915-925. doi:10.1016/j.jpsychires.2009.01.012
- Cook, W. W., & Medley, D. M. (1954). Proposed hostility and pharisaic-virtue scales for the MMPI. *Journal of Applied Psychology, 38*, 414-418. doi:10.1037/h0060667
- Cogle, J. R., Zvolensky, M. J., & Hawkins, K. A. (2013). Delineating a relationship between problematic anger and cigarette smoking: A population-based study. *Nicotine & Tobacco Research, 15*, 297-301. doi:10.1093/ntr/nts122
- Deffenbacher, J. L. (1993). General anger: Characteristics and clinical implications. *Psicologia Conductual, 1*, 49-67.
- DiGiuseppe, R., & Tafrate, R. C. (2003). Anger treatment for adults: A meta-analytic review. *Clinical Psychology: Science and Practice, 10*, 70-84. doi:10.1093/clipsy.10.1.70
- Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development, 51*, 162-170. doi:10.2307/1129603
- Eckhardt, C., Norlander, B., & Deffenbacher, J. (2004). The assessment of anger and hostility: A critical review. *Aggression and Violent Behavior, 9*, 17-43. doi:10.1016/S1359-1789(02)00116-7
- Embretson, S. E., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Epps, J., & Kendall, P. C. (1995). Hostile attribution bias in adults. *Cognitive Therapy and Research, 19*, 159-178. doi:10.1007/BF02229692
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 291-314. doi:10.1037/1082-989X.9.4.466
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hawkins, K. A., & Cogle, J. R. (2013a). A test of the unique and interactive roles of anger experience and expression in suicidality: Findings from a population-based study. *Journal of Nervous and Mental Disease, 201*, 959-963. doi:10.1097/NMD.0000000000000041
- Hawkins, K. A., & Cogle, J. R. (2013b). Effects of interpretation training on hostile attribution bias and reactivity to interpersonal insult. *Behavior Therapy, 44*, 479-488. doi:10.1016/j.beth.2013.04.005
- Hawkins, K. A., Hames, J. L., Ribeiro, J. D., Silva, C., Joiner, T. E., & Cogle, J. R. (2014). An examination of the relationship between anger and suicide risk through the lens of the interpersonal theory of suicide. *Journal of Psychiatric Research, 50*, 59-65. doi:10.1016/j.jpsychires.2013.12.005
- Hawkins, K. A., Macatee, R. J., Guthrie, W., & Cogle, J. R. (2013). Concurrent and prospective relations between distress tolerance, life stressors, and anger. *Cognitive Therapy and Research, 37*, 434-445. doi:10.1007/s10608-012-9487-y
- Hazebroek, J. F., Howells, K., & Day, A. (2001). Cognitive appraisals associated with high trait anger. *Personality and Individual Differences, 30*, 31-45. doi:10.1016/S0191-8869(00)00007-6
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*, 1369-1385.
- Kuckertz, J. M., Amir, N., Tobin, A. C., & Najmi, S. (2013). Interpretation of ambiguity in individuals with obsessive-

- compulsive symptoms. *Cognitive Therapy and Research*, 37, 232-241. doi:10.1007/s10608-012-9478-z
- Litt, M. D., Cooney, N. L., & Morse, P. (2000). Reactivity to alcohol-related stimuli in the laboratory and in the field: Predictors of craving in treated alcoholics. *Addiction*, 95, 889-900. doi:10.1046/j.1360-0443.2000.9568896.x
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Lord, F. M., & Novick, M. R. (1968). Standard errors of measurement at different ability levels. *Journal of Educational Measurement*, 21, 239-243.
- Lovibond, P. F., & Lovibond, S. H. (1995). The structure of emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33, 335-344. doi:10.1016/0005-7967(94)00075-U
- Macatee, R. J., Capron, D. W., Schmidt, N. B., & Cugle, J. R. (2013). An examination of low distress tolerance and life stressors as factors underlying obsessions. *Journal of Psychiatric Research*, 47, 1462-1468. doi:10.1016/j.jpsy-chires.2013.06.019
- Miller, T. Q., Markides, K. S., Chiriboga, D. A., & Ray, L. A. (1995). A test of the psychosocial vulnerability and health behavior models of hostility: Results from an 11-year follow-up study of Mexican Americans. *Psychosomatic Medicine*, 57, 572-581.
- Nguyen, T. H., Han, H. R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *Patient-Centered Outcomes Research*, 7, 23-35. doi:10.1007/s40271-013-0041-0
- Novaco, R. W. (2003). *The Novaco Anger Scale and Provocation Inventory*. Los Angeles, CA: Western Psychological Services.
- Orobio de Castro, B., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development*, 73, 916-934. doi:10.1111/1467-8624.00447
- Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17). Iowa City, IA: Psychometric Society.
- Smith, T. W., Glazer, K., Ruiz, J. M., & Gallo, L. C. (2004). Hostility, anger, aggressiveness and coronary heart disease: An interpersonal perspective on personality, emotion, and health. *Journal of Personality*, 72, 1217-1270. doi:10.1111/j.1467-6494.2004.00296.x
- Spielberger, C. D. (1999). *State Trait Anger Expression Inventory-2: Professional manual*. Lutz, FL: PAR.
- Strong, D. R., Kahler, C. W., Greene, R. L., & Schinka, J. (2005). Isolating a primary dimension within the Cook-Medley hostility scale: A Rasch analysis. *Personality and Individual Differences*, 39, 21-33. doi:10.1016/j.paid.2004.08.011
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston, MA: Allyn & Bacon.
- Tafra, R. C., Kassino, H., & Dundin, L. (2002). Anger episodes in high- and low-anger community adults. *Journal of Clinical Psychology*, 58, 1573-1590. doi:10.1002/jclp.10076
- Watson, D., & Clark, L. A. (1992). Affects separable and inseparable: On the hierarchical arrangement of the negative affects. *Journal of Personality and Social Psychology*, 62, 489-505. doi:10.1037/0022-3514.62.3.489
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070. doi:10.1037//0022-3514.54.6.1063
- Wenzel, A., & Lystad, C. (2005). Interpretation biases in angry and anxious individuals. *Behaviour Research and Therapy*, 43, 1045-1054. doi:10.1016/j.brat.2004.02.009
- Wilkowski, B. M., & Robinson, M. D. (2010). The anatomy of anger: An integrative cognitive model of trait anger and reactive aggression. *Journal of Personality*, 78, 9-37. doi:10.1111/j.1467-6494.2009.00607.x
- Williams, J. E., Paton, C. C., Siegler, I. C., Eigenbrodt, M. L., Nieto, F. J., & Tyroler, H. A. (2000). Anger proneness predicts coronary heart disease risk: Prospective analysis from the atherosclerosis risk in communities (ARIC) study. *Circulation*, 101, 2034-2039. doi:10.1161/01.CIR.101.17.2034