Modern Psychological Studies

Volume 27 | Number 1 Article 7

2021

Hidden links: Trait anxiety and the hostile attribution bias

Sarah Gracia
Bridgewater State University, sgracia@umassd.edu

Ashley A. Hansen-Brown
Bridgewater State University, ahansenbrown@bridgew.edu

TRAIT ANXIETY AND THE HOSTILE ATTRIBUTION BIAS

Abstract

The hostile attribution bias (HAB) is a tendency to interpret malevolent intentions when confronted by ambiguous actions of others. This project examines the relationship between HAB and trait anxiety and whether a metacognitive manipulation reduces HAB. In Study 1, our results showed that trait anxiety and HAB had a positive correlation using both methods of detecting hostility, even when negative affect was accounted for. In Study 2, overall analyses revealed that compared to a true control condition, the metacognition manipulation reduced the link between trait anxiety and HAB. This study needs to be replicated before we can definitively draw conclusions, but still guides us to a potential new method for reducing HAB in anxious individuals.

Hidden Links: Trait Anxiety and the Hostile Attribution Bias

Trait anxiety refers to how chronically anxious a person is. Despite anxiety being a clinical diagnosis, it is also a spectrum, where some people are more anxious than others without having an anxiety disorder. Many researchers believe that worry and rumination are the two underlying factors of trait anxiety (Ugalde-Araya et al., 2020). Researchers theorize that harmful coping strategies used to deal with unpleasant situations and emotions cause excessive trait anxiety (Mennin et al., 2005). A person's guardians early in life may influence trait anxiety; individuals with high trait anxiety rated their mothers as less caring yet more overprotective than those with low trait anxiety (Bennet & Stirling, 1998). Whatever its origin, trait anxiety influences many spheres of a person's life.

Higher trait anxiety is related to psychological problems. Higher trait anxiety elevates the risk for developing anxiety disorders (Chambers et al., 2004). Likewise, trait anxiety is correlated to rejection sensitivity (Wu et al., 2020). Interestingly, those with higher trait anxiety have a slower response time to speed-based stimuli identification tasks (Xia et al., 2020). Researchers speculate this occurs because trait anxiety impedes top-down goal processes, or how one's past experiences shapes perception, which causes problems in inhibition functioning such as executive control over motor responses; this in turn may negatively impact both cognitive control and performance (Pacheco-Unguetti et al., 2010).

Many people with high trait anxiety suffer from cognitive biases that distort their perceptions, such as a cognitive pattern that selectively favors encoding threatening information over other types (Mathews & MacLeod, 1985). One of these biases is the Jumping to Conclusions bias, which is a bias where someone will stop processing a situation and "jump" to a negative conclusion (Bensi & Giusberti, 2007). Researchers think this occurs because the person

wants to reduce uncertainty, even at the cost of being inaccurate (Bensi & Giusberti, 2007). Researchers believe that these biases persist because they protect; if the Jumping to Conclusions bias says that a person is an enemy and to avoid them, and the person with the bias does so, they survive the ordeal, which reiterates the bias (Cardella & Gangemi, 2015).

Given the link between anxiety and cognitive biases, it seems plausible that people with high trait anxiety might be more likely to have a hostile attribution bias (HAB). HAB is a bias where in an ambiguous social situation, a person believes that others' intentions are malicious, even with little to no evidence (Dodge, 1980; Hawkins & Cougle, 2013). HAB is more prevalent in ambiguous social situations (Combs et al., 2007; Crick & Dodge, 1994). Just like how trait anxiety is related to various psychological issues, HAB is associated with many problems. For example, HAB is more prevalent in those with more anger (Bond et al., 2004), and in those with obsessive-compulsive disorder (OCD; Tellawi et al., 2016). People with more HAB tend to have more depression (Gasse et al., 2020). Children with more aggression (Nasby et al., 1980) or whose parents routinely "team up" against them (Coe et al., 2020) have elevated HAB.

HAB has previously been linked to social anxiety. For example, those with social anxiety are more likely to interpret events as negative (Amin et al., 1998). Social anxiety involves a fear of being judged by others (Tone et al., 2011). It appears that higher levels of the bias also increase anxiety, creating a vicious cycle (Stopa & Clark, 2000). According to the Clark and Wells Model of social phobia, people with high levels of social anxiety create a misshaped portrayal of themselves seen by others (Clark & Wells, 1995). Social anxiety has a well-documented relationship with paranoia; for example, social anxiety levels predict paranoia levels (Pisano et al, 2016). Some suggest that paranoid thinking eventually leads to social anxiety (Pisano et al, 2016), and social anxiety is a paramount component of persecutory delusions

(Freeman et al., 2005). Since HAB has already been linked to social anxiety, it may also be related to trait anxiety.

Since HAB is widespread across multiple domains, it is imperative that it is treated. HAB is treatable; for example, cognitive interventions reduce levels of HAB (see Dodge, 2006). Self-persuasion, or championing ideas that are at odds with one's own beliefs, is effective in lessening HAB in children (van Dijk et al., 2019). Children who witnessed playful fighting had lower HAB levels (Boulton, 2012). While these studies are extremely informative and important, there is a need for more literature about reducing HAB in adults.

Despite some available HAB reduction methods, biases are very pervasive. Preexisting beliefs and attitudes, like prejudice, can increase bias levels (for a review, see Chien et al., 2014), which make bias corrections challenging, as many people believe they are correct in their judgements even when they are not. Higher cognitive load promotes biased thinking, since it leads to stereotype priming (Wigboldus et al., 2004). Even subtle cues, like primes (Higgins et al., 1977) or contextual stimuli in the environment (Chien et al., 2010) can increase bias. Additionally, when ambiguous information is presented, individuals usually rely on biases to "fill in the blanks" or make an appraisal on the situation they are in (Chaiken & Maheswaran, 1994).

Many times, people need straightforward directions to reduce their biases. People must be aware of a bias before they can correct it (Gawronski, 2004). Individuals correct for perceived bias, not necessarily actual bias; in other words, people correct for biases they believe are there. Hence, they will stop correcting for bias when they think it is gone, even if the bias is present (Wegener & Petty, 1995). Another rather pernicious fact is that people still have bias even when they do their best to be impartial (for a review, see Chien et al., 2014).

Some people may not wish to remove their biases at all. In fact, people do not like correcting a bias that makes them look and feel good (McCaslin et al., 2010). Additionally, people may be less likely to want to get rid of biases that influence what stimuli they pay attention to if they are rewarded somehow (Anderson et al., 2013). Naturally, if someone does not think they are biased, they will be unmotivated to correct for bias and will stay biased (Wegener et al., 2001). Thankfully, once a person is aware they are biased, they typically do their best to eliminate the bias (McCaslin et al., 2010).

With all the challenges getting rid of biases entails, cognitive bias modification (CBM) interpretation training may help. CBM interpretation training is a therapy that attempts to reduce cognitive biases that affect how people process and interpret information by eliminating harmful processes and reducing interpretation biases (Dodd et al., 2019). Participants are given a word-sentence task, where they are given either a positive or negative word for the context of a sentence. When participants make a hostile interpretation of the sentence, they are told that they made a mistake, and to correct for positive interpretations (MacDonald et al., 2020). CBM can be improved by including goal-relevant behavioral choices to make it more personalized and meaningful to the participant, and that the person should see the consequences of their behavioral choices (Wiers et al., 2020) to see that their efforts are actually helping them.

In metacognitive therapy, a different but similar approach to CBM, the objective is changing negative perspectives about the way one thinks about their thoughts. It may also be a useful tool to treat HAB, since negative beliefs about metacognition are associated with many problems. Negative metacognitive beliefs such as the “danger and uncontrollability of thoughts” are related to emotional disfunction (Sellers et al., 2018; Wells & Matthews, 1996). In other words, individuals who consider thoughts to be dangerous and uncontrollable have more

emotional disfunction. Negative metacognitive beliefs are positively connected to psychological vulnerability (Nordahl & Wells, 2017), severe depression (Aldahadha, 2020), and may be positively related to trait anxiety (Nordahl & Wells, 2017). Metacognitive beliefs are malleable with manipulation (Sellers et al., 2018). Metacognitive therapy is used to treat many psychological issues by challenging negative beliefs such as the importance of thoughts and need to control thoughts. Metacognitive therapy treats generalized anxiety disorder (GAD; Wells & King, 2006), trait anxiety (Mathews et al., 2007), and OCD (Miegel et al., 2020).

The purpose of the current research is to see whether those with high trait anxiety have more HAB, and if so, if a metacognition manipulation will decrease HAB levels. For Study 1, we hypothesize that people with more trait anxiety will have more HAB, perceiving the intentions of others as malicious in ambiguous social situations. We also hypothesize that HAB will still be present in those with high trait anxiety when negative affect is accounted for. For Study 2, we hypothesize that we will replicate Study 1 results to find that those with more HAB will display more trait anxiety. We also hypothesize that those who receive the metacognition manipulation will have less HAB than those who were only told what the bias is. Lastly, we hypothesize that those with high trait anxiety instructed to control HAB will have less HAB than those only given the definition of HAB, but more HAB than those with low trait anxiety even after being told to control for HAB.

Study 1[1]

In Study 1, we looked at whether those with higher trait anxiety had more HAB. For this first study, we measured trait anxiety, trait anger, verbal aggression, physical aggression, and HAB. For HAB, we included both a trait-based HAB measure and a scenario-based HAB

[1] Study 1 was funded by a Bridgewater State University Adrian Tinsley Program Summer Grant.

measure. However, since HAB involves negative interpretations of others, it is possible that any link between trait anxiety and the bias are a result of negative affect, and not a true cognitive bias. Negative affect encompasses all negative or unpleasant emotions, and is correlated with health complaints (for a review, see Watson & Pennebaker, 1989). To make sure that the results would truly show that trait anxiety is connected to HAB and not negative affect interfering, we used the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) to measure both positive and negative affect.

Method

Participants

A G*Power *a priori* power analysis (Faul et al., 2009) revealed 134 participants were needed to have 95% power to detect a medium-sized effect. Before data collection, we preregistered the study on the Open Science Framework (https://osf.io/d8bu2) to eliminate hindsight bias and decrease the likelihood of Type I errors. Participants in this study were an adult sample recruited using the online data collection platform Prolific. The participants were screened to ensure they were 18 years old or above, had a task approval rating of at least 95%, and were United States citizens. Participants received $1.71 in compensation. We recruited 220 participants and did not have to exclude any based on our preregistered exclusion criteria. Females comprised 54.1% of all participants, males 43.6%, and non-binary 2.3%. Caucasian/Whites comprised 71.4% of all participants, 9.1% Asian/Asian American, 8.3% multiracial, 7.3% African American/Black, 2.7% Hispanic/Latino, 0.9% American Indian/Native American, and 0.5% did not disclose their race. The participants' ages ranged from 18 years old to 79 years old (M = 33.20, SD = 12.05).

Materials & Procedure

Participants first completed a Captcha to exclude bots, followed by a consent form. They completed a series of four scales, followed by demographic items and data quality checks. Participants also completed a question asking how anxious they were about the COVID-19 pandemic. Participants were then debriefed and compensated for their time.

Trait Anxiety. The Beck Anxiety Inventory (Beck et al, 1988) is a 21-item self-report measure to assess trait anxiety. Participants were asked to rate how much they experienced a symptom (e.g. "numbness or tingling") in the past month on a Likert scale from 1 ("Not at all") to 4 ("Severely-it bothered me a lot"). We summed all items to create a total trait anxiety score (a = .93), with higher scores indicating higher trait anxiety.

Positive and Negative Affect. The PANAS scale (Watson et al., 1988) is a 20-item self-report measure to account for positive and negative affect. Similar to the Beck Anxiety Inventory, participants were asked to rate how much they experienced a mood (e.g., "Irritable") in the past week on a Likert scale from 1 ("Very slightly or not at all") to 5 ("Extremely"). We summed all items on the negative affect subscale to create a total negative affect score (a = .91), with a higher score indicating higher negative affect. Likewise, we summed all items on the positive affect subscale to create a total positive affect score (a = .93), with a higher score indicating higher positive affect.

Trait HAB. The Aggression Questionnaire (Buss & Perry, 1992) is a 29-item self-report measure to assess hostility (e.g., "I am suspicious of overly friendly strangers"), anger (e.g., "I sometimes feel like a powder keg ready to explode"), physical aggression (e.g., "If someone hits me, I hit them back"), and verbal aggression (e.g., "I can't help getting into arguments when people disagree with me"). Participants responded to each statement on a Likert scale from 1 ("Extremely Uncharacteristic of Me") to 5 ("Extremely Characteristic of Me"). We summed all

the items on the subscales to create total scores for physical aggression (a = .82), verbal aggression (a = .76), anger (a = .83), and hostility (a = .87). These scores were our operationalizations of physical aggression, verbal aggression, and anger, and one operationalization of HAB.

Scenario-Based HAB. The Word Sentence Association Paradigm (WSAP) is a 66-item self-report measure to assess perceived hostility and perceived neutrality or non-negativity in interpretations of a variety of ambiguously negative social situations (Dillon et al., 2016). For each item, participants were presented with a sentence and a word, and were instructed to rate how similar the sentence and word are. An example question is "Sentence: Someone disagrees with your point of view. Word: Argumentative", which participants rated on a Likert scale from 1 ("Not at all similar") to 6 ("Extremely Similar"). Each sentence was presented twice, with one word being hostile and one word being neutral or non-negative. Higher scores with a negative word are associated with more perceived hostility, while higher scores with a neutral word are associated with more perceived neutrality or non-negativity. We summed all items reflecting hostile interpretations (a = .93) and neutral or non-negative interpretations (a = .89). The hostile interpretations sum score was one operationalization of HAB.

Results & Discussion

Table 1 outlines the bivariate correlations, means, and standard deviations of all variables used in Study 1.

Main Analyses

To test the hypothesis that trait anxiety is correlated with HAB, we ran Pearson correlation analyses. We first sought to discover the relationship between our two HAB operationalizations. The Aggression Questionnaire's operationalization of hostility was

moderately to strongly positively correlated with the WSAP's operationalization of hostility $r(216) = .48$, $p < .001$. Results showed that for trait anxiety and HAB, there was a significant strong positive correlation when using the Aggression Questionnaire, $r(218) = .58$, $p < .001$. When using the WSAP, it turns into a moderate positive correlation, $r(216) = .31$, $p < .001$. This supports our hypothesis that trait anxiety and HAB are positively correlated.

Next, we tested whether trait anxiety and HAB are still positively correlated when negative affect is accounted for. At the bivariate level, negative affect is strongly positively correlated with trait anxiety, $r(218) = .76$, $p < .001$, strongly positively correlated with the Aggression Questionnaire's operationalization of hostility, $r(218) = .59$, $p < .001$, and moderately positively correlated with the WSAP's operationalization of hostility, $r(216) = .31$, $p < .001$. When controlling for negative affect using a partial correlation, trait anxiety's relationship to the Aggression Questionnaire changed to a weak positive correlation but remained significant, $r(217) = .25$, $p < .001$. Trait anxiety's relationship to the WSAP changed to a non-significant correlation that was still trending in the predicted direction, $r(215) = .12$, $p < .07$.

Exploratory Analyses

Next, we explored relationships between trait anxiety and physical aggression, verbal aggression, and anger. Trait anxiety was moderately positively correlated with physical aggression, $r(218) = .30$, $p < .001$, moderately positively correlated with verbal aggression, $r(218) = .32$, $p < .001$, and moderately positively correlated to trait anger, $r(218) = .48$, $p < .001$. This means that trait anxiety is connected to physical aggression, verbal aggression, and trait anger, in agreement with previous research correlating anxiety and anger (Zinner et al., 2008), and correlating trait anxiety and both forms of aggression (Chung et al., 2019).

Then, we looked at trait anxiety's relationship with positive affect and benign interpretations. Results showed that there was a significant weak negative correlation for positive affect, $r(218) = -.15$, $p = .03$, and a nonsignificant correlation for benign interpretations, $r(216) = .08$, $p = .25$. This means that trait anxiety has a slight inverse relationship with positive affect, and not related to benign interpretations whatsoever. To our knowledge, this is the first study that examines the relationship between positive affect and trait anxiety.

Study 2[2]

Relieving anxious individuals of HAB is paramount because it can interfere with their social functioning. Previous research has separately investigated metacognitive therapy and bias reduction, but has not yet linked them. We tested whether combining the two may uncover a possible valuable therapeutic technique. In Study 2, we tested whether a metacognitive manipulation would reduce HAB in participants, and whether our Study 1 finding that trait anxiety was connected to HAB would replicate. Since we already discovered that negative affect did not explain the relationship between trait anxiety and HAB, the PANAS scale was no longer necessary. The metacognitive manipulation would work best if the measure of HAB was easier to manipulate, so we decided it would be wise to drop the trait-based Aggression Questionnaire's operationalization of HAB and retain the WSAP's scenario-based operationalization of HAB.

Method

Participants

A G*Power *a priori* power analysis (Faul et al., 2009) reported 210 participants were needed to have 95% power to detect a medium-sized effect. We again preregistered the study on the Open Science Framework (https://osf.io/vhsc3/) and recruited adult participants from

[2] Study 2 was funded by a Bridgewater State University Office of Undergraduate Research Semester Grant.

Prolific. The participants were again screened to ensure they were 18 years old or above, had a task approval rating of at least 95%, and were United States citizens. Participants were compensated $1.60. We originally recruited a total of 302 participants and had to exclude one participant based on our preregistered exclusion criteria who selected a two out of five when asked how seriously they took the study with one being not seriously at all and five being extremely serious, giving us a total of 301 participants. Females comprised 55.8% of all participants, males 43.2%, and non-binary 1%. Caucasian/Whites comprised 68.11% of all participants, 9.97% Asian/Asian Americans, 8.31% multiracial, 6.98% African American/Blacks, 4.98% Hispanic/Latinos, 0.33% American Indian/Native Americans, 0.66% participants selected "Other", and 0.33% participants did not disclose their race. The participants' ages ranged from 18 years old to 74 years old (M = 33.18, SD = 12.51).

Materials & Procedure

Participants first completed a Captcha to exclude bots, followed by a consent form. As in Study 1, participants completed the Beck Anxiety Inventory scale as a measure of trait anxiety (a = .93). Then, participants were either only given the definition of HAB, or were given the definition and told to correct for the bias. Next, as in Study 1, they completed both the hostile interpretations of the WSAP-Hostility questionnaire as a measure of HAB (a = .92), and the neutral interpretations (a = .90). Lastly, they answered demographic items and data quality checks. Participants also filled out a question that asked how anxious they were about the COVID-19 pandemic. Participants were then debriefed and compensated for their time.

We removed the PANAS scale and the Aggression Questionnaire from this study. We dropped the PANAS because controlling for negative affect was not central to our hypothesis for this study. Additionally, the Aggression Questionnaire provides a trait measure of HAB, which

we assumed would be more difficult to change than the scenario interpretation of HAB measured by the WSAP; thus, we retained the WSAP but dropped the Aggression Questionnaire in Study 2.

Metacognition manipulation. To manipulate metacognition, we provided a set of instructions to participants before they completed the WSAP. We divided participants into two groups: the definition group and the correction group. We provided solely the definition of HAB to the definition group to make them aware of the bias. Since people tend to correct for perceived bias rather than actual bias (Wegener & Petty, 1995), we explicitly brought the bias to participants' attention in both definition and correction groups. Specifically, before completing the WSAP measure of HAB, participants saw a screen which said: "The next questionnaire is measuring the hostile attribution bias, or HAB. HAB is a cognitive bias where someone in an ambiguous social situation will interpret the other person's actions as hostile." In the correction group, we provided the definition to make participants aware of the bias and instructed them to correct for the bias in their responses to the WSAP measure. Specifically, participants saw the following instructions: "The next questionnaire is measuring the hostile attribution bias, or HAB. HAB is a cognitive bias where someone in an ambiguous social situation will interpret the other person's actions as hostile. Please try to correct for this bias and respond in an unbiased way while you fill out the following questionnaires."

Manipulation check. In the demographics section of the survey, after asking participants their race and before asking how seriously they took the questionnaire, we included a manipulation check to see if the participants were paying attention to the instructions provided. Specifically, participants saw the following instructions: "What did we tell you before you completed the hostile attribution bias questionnaire?" Participants could answer either

"Definition of hostile attribution bias only" or "Definition of hostile attribution bias and instructed to correct for it". We coded Qualtrics to display whether or not they passed the manipulation check so we could note which participants passed or failed so we could exclude them from the study.

Results & Discussion

Table 2 outlines the bivariate correlations, means, and standard deviations of all variables used in Study 2.

Main Analyses

To expand on the previous study examining the hypothesis that trait anxiety is related to HAB, we calculated a Pearson Correlation analysis. Unexpectedly, there was a nonsignificant correlation between the WSAP's operationalization of hostility and trait anxiety $r(301) = .08$, $p = .18$. This goes against our hypothesis that trait anxiety and the hostile attribution bias are positively correlated, failing to replicate Study 1.

Next, we tested whether our metacognition manipulation was effective by running a *t*-test. Also unexpectedly, the results indicated that the metacognition manipulation failed, with no significant differences between the definition group ($M = 108.52$, $SD = 21.79$) and the correction group ($M = 108.58$, $SD = 23.38$) on WSAP hostility, $t(299) = .02$, $p = .98$. This goes against our hypothesis that the metacognition manipulation would work.

We then tested whether the metacognition was effective for those with high trait anxiety. In order to do this, we ran a multiple regression test using the General Linear Model. The metacognition did not reduce rates of HAB for those with high anxiety, as indicated by the non-significant interaction between the two variables, F (3, 297) = 1.49, $p = .22$. This goes against

our hypothesis that the metacognition manipulation would reduce HAB levels in participants with high trait anxiety.

Finally, we noticed that 86 out of 301 participants failed the manipulation check (75 in the correction group, 11 in the definition group). Because so many participants failed the manipulation check, we suspected that perhaps our unexpected results were derived from this phenomenon. To investigate, instead of deleting them from the study, we split the data file between participants who passed and failed the manipulation check and re-ran the Pearson correlation, *t*-test, and multiple regression using the General Linear Model. Surprisingly, there were no statistically significant differences between participants who passed the manipulation check and those who failed; thus, failure of the manipulation check does not seem to explain our findings.

Additional Analyses

The prior analyses suggest that Study 2 both failed to replicate Study 1 or show any benefit of our metacognition manipulation. However, given that previous research supports links between trait anxiety and various cognitive biases, and given that we did not have a true control condition in this study, our null findings in this study may indicate another explanation. Specifically, perhaps the surprising lack of correlation in Study 2 simply indicates that any manipulation drawing attention to HAB eliminates the link between trait anxiety and HAB. To test this possibility, we ran additional analyses combining the data from Study 1 and both conditions from Study 2. In these new analyses, we treated the data from Study 1 as a control condition, as both trait anxiety and HAB were only measured with no manipulation at all; both conditions from Study 2 became treatment conditions.

We used a multiple regression analysis to properly examine the links between the three conditions. The overall model was significant, $F(3, 516) = 7.21$, $p < .001$, $R^2 = .04$. There was no main effect of condition, $b = 6.59$, $t(516) = 1.62$, $p = .11$. There was a main effect for trait anxiety, $b = .80$, $t(516) = 3.43$, $p = .001$, replicating the bivariate correlation from Study 1. Furthermore, the interaction between condition and trait anxiety was marginally significant, $b = -.22$, $t(516) = -1.93$, $p = .054$. When we re-ran this analysis excluding participants from the Study 2 data who failed manipulation check, the interaction became significant, $b = -.29$, $t(429) = -2.18$, $p = .03$. See Figure 1 for a graph of the interaction.

Surprisingly, it appears that even the definition group being told that the bias existed was enough to reduce the link between trait anxiety and HAB, with that link even more reduced in the correction group. Interestingly, since average levels of both trait anxiety and HAB in all three conditions are roughly equal, this suggests our intervention might reduce the anxiety-HAB link but does not necessarily reduce overall HAB among everyone. Thus, it appears that perhaps both merely informing participants of HAB's existence and being told of HAB and instructed to correct for it are both effective treatments for HAB, at least temporarily. However, since the sample sizes between conditions are unequal, and particularly unequal after excluding participants who failed the manipulation check from Study 2, we must be cautious in interpreting results; more replication is required before we can definitively establish solid conclusions.

General Discussion

Across two studies, we found support for our hypotheses that trait anxiety and HAB are related. This cannot be accounted for by negative affect, which hints the variables have a true connection. Thus, people with high trait anxiety may be more likely to assume others' intentions are antagonistic, even if the opposite is true. However, a metacognition manipulation eliminated

the link between trait anxiety and HAB, which implies that activating metacognition significantly reduces HAB. These findings give researchers real insight into a new possible issue individuals with high trait anxiety suffer from, as well as a possible new treatment method.

Study 1 showed that trait anxiety was correlated with both operationalizations of HAB. This link remained even when negative affect was accounted for, supporting our hypothesis. Thus, negative affect does not explain the link, so the variables are indeed related. This means that people with more trait anxiety have more HAB in general and in specific social scenarios, and provides a fuller picture that HAB is elevated across multiple domains, not in one particular instance.

Our Study 1 exploratory analyses also revealed that trait anxiety is correlated with more trait anger, physical aggression, and verbal aggression than the general population. We speculate that this occurs because chronic anxiety takes an emotional toll on the human body and leaves less patience, which results in more aggression, both physical and verbal. This is not surprising, since this was also found in a study of rats (Beiderbeck et al., 2012) where rats with low and high trait anxiety had more aggression than controls; this finding may also apply for humans. Similarly, those with OCD have more trait anger than those without it (Radomsky et al., 2007); since OCD was classified as an anxiety disorder prior to 2013, it is logical that the same holds true for trait anxiety.

Although our analyses in Study 2 initially suggested that trait anxiety and HAB were not related and that the metacognition manipulation failed, when we compared the two manipulation conditions in Study 2 with the control group in Study 1, we found that our manipulation eliminated the correlation between trait anxiety and HAB that otherwise would have been there.

This suggests that the metacognitive manipulation actually worked, and that even informing individuals that HAB exists leads to a reduction of the bias.

These results imply that the "control" group for Study 2, which involved being told the definition of HAB, was actually a treatment. Furthermore, it suggests that both telling people about HAB and also telling them about HAB and asking to correct for it can lead to reductions of the bias, at least in the short-term. We suspect this occurred because once we defined HAB and made participants were aware of its existence, they may have naturally corrected for it without being instructed. It is also possible that when we provided the definition of HAB, that participants interpreted that as an indirect instruction to correct for the bias. Either way, this shows support for metacognitive techniques as interventions to reduce HAB, and suggests that even simple exercises will reap multiple benefits.

Limitations & Future Directions

As with all studies, this one has limitations. Only English-speaking Americans completed the survey, so the results only apply to one nation. For future studies, it is recommended that they be opened to all English speakers in order to get a more diverse subject pool. Also, translating the scales used into other languages would be beneficial so participants of different cultures with differing views of anxiety can share their input. The self-selection bias is also a risk factor for this study. Additionally, the participants were all on Prolific; we only used one source for participant gathering, which makes it harder to generalize the results to the general population. Future studies should take care to use other methods to recruit participants, such as using more than one website to increase variability in sample composition.

Future research should replicate our findings to see if the same results occur. If this happens, we can be more confident about our findings since they have not changed under

different circumstances, and the results would be more robust. This is particularly important because the interaction in Study 2 was only marginally significant with all participants included, and became significant when we excluded participants who failed the manipulation check in Study 2, leaving us with very unequal sample sizes between conditions. Furthermore, our additional analyses comparing all three conditions were not pre-registered and were exploratory. Thus, we recommend that future research implement all three conditions from our two studies (control, definition, correction) in a single study with a much larger sample size. Replication of our findings in a well-powered study like this would further solidify our findings and potentially spark further investigation of our metacognition manipulation.

Another possible avenue is to see whether these treatment effects continue over the long-term or if the results fade given enough time. If the effects only last for the short-term, then new possible interventions that have longer-lasting effects should be investigated. If the effects are long-lasting, then many therapies should integrate metacognition manipulation to better help people. It would also be prudent to investigate other methods of bias reduction, such as including long-term goals in the process to increase motivation to eliminate bias, or repeating the events slowly out loud to lead to new interpretations of events.

Implications & Conclusions

This research has many important implications. This research matters because it helps therapists be aware of a potential issue their highly anxious clients are facing, and by providing possible steps they can take to eliminate HAB. If HAB affects people with higher trait anxiety, it is most likely that these individuals will interpret others as hostile, even if the other person is friendly. This will most likely have consequences on their social lives and have fewer enriching relationships. If researchers replicate the effects of the metacognitive manipulation, therapists

can use this technique to reduce HAB in their anxious clients. With this, we offer researchers a new avenue to explore on the journey of helping the human population.

In conclusion, we discovered that both trait and scenario-based operationalizations of HAB were positively correlated with trait anxiety, even when negative affect was accounted for, indicating that those with more anxiety believe that the intentions of others are more malevolent and cruel. A metacognition manipulation can significantly reduce the correlation between HAB and trait anxiety by either informing participants of its existence or asking them to correct for it, though more research is needed to test how long this effect lasts. These studies provide an in-depth examination of the relationship between trait anxiety and HAB, showing that a metacognitive manipulation can be a key in reducing bias. This research can advance therapeutic techniques for anxiety and implemented to improve the lives of countless individuals.

References

Aldahadha, B. (2020). Metacognition, mindfulness attention awareness, and their relationships with depression and anxiety. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*. Advance online publication. https://doi-org.libserv-prd.bridgew.edu/10.1007/s10942-020-00367-y.

Anderson, B.A., Laurent, P.A., & Yantis, S. (2013). Reward predictions bias attentional selection. *Frontiers in Human Neuroscience, 7*. https://doi.org/10.3389/fnhum .2013.00262

Amin, N., Foa, E.B., & Coles, M.E. (1998). Negative interpretation bias in social phobia. *Behaviour Research and Therapy, 36*, 945-957. https://doi.org/10.1016/s0005-7967(98)00060-6

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*, 893–897. https://doi.org/10.1037/0022-006X.56.6.893

Beiderbeck, D.I., Reber, S.O., Havasi, A., Bredewold, R., Veenema, A.H., & Neumann, I.D. (2012). High and abnormal forms of aggression in rats with extremes in trait anxiety—Involvement of the dopamine system in the nucleus accumbens. *Psychoneuroendocrinology, 37*, 1969-1980. https://doi.org/10.1016/j.psyneuen.2012.04.011

Bennet, A. & Stirling, J. (1998). Vulnerability factors in the anxiety disorders. *British Journal of Medical Psychology, 71*, 311-321. https://doi.org/10.1111/j.2044-8341.1998.tb00994.x

Bensi, L. & Giusberti, F. (2007). Trait anxiety and reasoning under uncertainty. *Personality and Individual Differences, 43*, 827–838. https://doi.org/10.1016/j.paid.2007.02.007

Bond, A.J., Verheyden, S.L., Wingrove, J., & Curran, H.V. (2004). Angry cognitive bias, trait aggression and impulsivity in substance users. *Psychopharmacology, 171,* 331–339. http://dx.doi.org/10.1007/s00213-003-1585-9

Boulton, M.J. (2012). Children's hostile attribution bias is reduced after watching realistic playful fighting, and the effect is mediated by prosocial thoughts. *Journal of Experimental Child Psychology, 113,* 36-48. https://doi.org/10.1016/j.jecp.2012.02.011

Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology, 63*(3), 452–459. https://doi.org/10.1037/0022-3514.63.3.452

Cardella, V., & Gangemi, A. (2015). Reasoning in schizophrenia. Review and analysis from the cognitive perspective. *Clinical Neuropsychiatry, 1*(2), 3-8.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology, 66,* 460-473. https://doi.org/10.1037/0 022-3514.66.3.460

Chambers, J. A., Power, K. G., & Durham, R. C. (2004). The relationship between trait vulnerability and anxiety and depressive diagnoses at long-term follow-up of generalized anxiety disorder. *Journal of Anxiety Disorders, 18,* 587– 607. http://dx.doi.org/10.1016/j.janxdis.2003.09 .001

Chien, Y.-W., Wegener, D. T., Hsiao, C.-C., & Petty, R. E. (2010). Dimensional range overlap and context effects in consumer judgments. *Journal of Consumer Research, 37,* 530–542. https://doi.org/10.1086/652415

Chien, Y-W., Wegener, D.T., Petty, R.E., & Hsiao, C-C. (2014). The flexible correction model: Bias correction guided by naïve theories of bias. *Social and Personality Psychology Compass, 8,* 275-286. https://doi.org/10.1111/spc3.12105

Clark, D. M. and Wells, A. (1995). A cognitive model of social phobia. In R. G. Heimberg, M. R. Liebowitz, D. A. Hope and F. R. Schneier (Eds.), *Social phobia: Diagnosis, assessment and treatment*. New York: Guilford Press.

Coe, J.L., Davies, P.T., Hentges, R.F., & Sturge-Apple, M.L. (2020). Detouring in the family system as an antecedent of children's adjustment problems. *Journal of Family Psychology, 34,* 814-824. http://dx.doi.org/10.1037/fam0000727

Combs, D.R., Penn, D.L., Wicher, M., & Waldheter, E. (2007). The Ambiguous Intentions Hostility Questionnaire (AIHQ): A new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive Neuropsychiatry, 12,* 128-143. https://doi.org/ 10.1080/13546800600787854

Crick, N. R. & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115,* 74-101. https://doi.org/10.1037/0033-2909.115.1.74

Dillon, K.H., Allan, N.P., Cougle, J.R., & Fincham, F.D. (2016). Measuring hostile attribution bias: The WSAP-Hostility scale. *Assessment, 23,* 1-13. https://doi.org/10.1177/1073191115599052

Dodd, D.R., Parsons, E.M., Clerkin, E.M., Forrest, L.N., Velkoff, E.A., Kunstman, J.W., & Smith, A.R. (2019). Perfectly imperfect: The use of cognitive bias modification to reduce perfectionism. *Journal of Behavior Therapy and Experimental Psychiatry, 64,* 167–174.

Dodge, K.A. (1980). Social cognition and children's aggressive behavior. *Child Development, 51*, 62–70. http://dx.doi.org/10.2307/1129603

Dodge, K.A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology*, *18*, 791-814. https://doi.org/10.1017/S0954579406060391

Faul, F., Erdfelder, E., Buchner, A., & Lang, A-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Freeman, D., Garety, P. A., Bebbington, P. E., Smith, B., Rollinson, R. and Fowler, D. (2005). Psychological investigation of the structure of paranoia in a non-clinical population. *British Journal of Psychiatry, 186*, 427–435. https://doi.org/10.1192/bjp.186.5.427

Gasse, A., Kim, W.S., & Gagnon, J. (2020). Association between depression and hostile attribution bias in hostile and non-hostile individuals: An ERP study. *Journal of Affective Disorders, 276*, 1077-1083. https://doi.org/10.1016/j.jad.2020.06.025

Gawronski, B. (2004). Theory-based bias correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias. *European Review of Social Psychology, 15*, 183-217. https://doi.org/10.1080/10463280440000026

Hawkins, K.A., & Cougle, J.R. (2013). Effects of interpretation training on hostile attribution bias and reactivity to interpersonal insult. *Behavior Therapy*, *44*, 479-488. https://doi.org/10.1016/j.beth.2013.04.005

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13*, 141–154. https://doi.org/10.1016/S0022-1031(77)80007-3

MacDonald, E.M., Koerner, N., Antony, M.M., Vickers, K., Mastorakos, T., & Kuo, J. (2020). Investigating the therapeutic potential of cognitive bias modification for high anxiety sensitivity. *Journal of Behavior Therapy and Experimental Psychiatry*, *68,* Article 101521.

Mathews, A., & MacLeod, C. (1985). Selective processing of threat cues in anxiety states. *Behaviour Research and Therapy*, *23*(5), 563–569. https://doi.org/10.1016/0005-7967(85)90104-4

Mathews, A., Ridgeway, V., Cook, E. & Yiend, J. (2007). Inducing a benign interpretational bias reduces trait anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *38,* 225–236. https://doi.org/10.1016/j.jbtep.2006.10.011

McCaslin, M.J., Petty, R.E., & Wegener, D.T. (2010). Self-enhancement and theory-based correction processes. *Journal of Experimental Social Psychology*, *46,* 830-835. https://doi.org/10.1016/j.jesp.2010.05.002

Miegel, F., Demiralay, C., Sure, A., Moritz, S., Hottenrott, B., Cludius, B., & Jelinek, L. (2020). The metacognitive training for obsessive-compulsive disorder: A pilot study. *Current Psychology*. Advance online publication, https://doi-org.libserv-prd.bridgew.edu/10.1007/s12144-020-01158-5

Mennin, D.S., Heimberg, R.G., Turk, C.L., Fresco, D.M. (2005). Preliminary evidence for an emotion dysregulation model of generalized anxiety disorder. *Behaviour Research and Therapy*, *43,* 1281–1310. https://doi.org/10.1016/j.brat.2004.08.008.

Nasby, W., Hayden, B., & DePaulo, B. M. (1980). Attributional bias among aggressive boys to interpret unambiguous social stimuli as displays of hostility. *Journal of Abnormal Psychology*, *89*, 459–468. https://doi.org/10.1037/0021-843X.89.3.459

Nordahl, H., & Wells, A. (2017). Individual differences in metacognitive knowledge contribute to psychological vulnerability more than the presence of a mental disorder does. *Mental Health & Prevention, 7,* 17-20. https://doi.org/10.1016/j.mhp.2017.07.003

Pacheco-Unguetti, A. P., Acosta, A., Callejas, A., & Lupiáñez, J. (2010). Attention and anxiety: Different attentional functioning under state and trait anxiety. *Psychological Science, 21,* 298–304.

Pisano, S., Catone, G., Pascotto, A., Iuliano, R., Tiano, C., Milone, A., Masi, G., & Gritti, A. (2016). Paranoid thoughts in adolescents with social anxiety disorder. *Child Psychiatry and Human Development, 47,* 792-798. https://doi.org/10.1007/s10578-015-0612-5

Radomsky, A.S., Ashbaugh, A.R., & Gelfand, L.A. (2007). Relationships between anger, symptoms, and cognitive factors in OCD checkers. *Behaviour Research and Therapy, 45,* 2712–2725. https://doi.org/10.1016/j.brat.2007.07.009

Sellers, R., Wells, A., & Morrison, A.P. (2018). An experimental manipulation of negative metacognitive beliefs in non-clinical paranoia: Effects on intrusions and state anxiety. *Journal of Experimental Psychopathology, 9,* 1-11. https://doi.org/10.5127/jep.062117

Stopa, L. & Clark, D.M. (2000). Social phobia and interpretation of social events. *Behaviour Research and Therapy, 38,* 273-283. https://doi.org/10.1016/s0005-7967(99)00043-1

Tellawi, G., Williams, M.T. & Chasson, G.S. (2016). Interpersonal hostility and suspicious thinking in obsessive-compulsive disorder. *Psychiatry Research, 243,* 295-302. https://doi.org/10.1016/j.psychres.2016.06.038

Tone, E.B., Goulding, S.M., & Compton, M.T. (2011). Associations among perceptual anomalies, social anxiety, and paranoia in a college student sample. *Psychiatric Research, 188,* 258-263. https://doi.org/10.1016/j.psychres.2011.03.023

Ugalde-Araya, D., Coto-Vílchez, C., Ávila-Aguirre, A., Chavarría-Soley, G., Raventós, H., & Contreras Rojas, J. (2020). Underlying domains of anxiety trait in a Costa Rican sample: Preliminary results. *Neurology, Psychiatry, & Brain Research, 35*, 38-41. https://doi.org/10.1016/j.npbr.2020.01.001

van Dijk, A., Thomaes, S., Poorthuis, A.M.G., & Orobio de Castro, B. (2019). Can self-persuasion reduce hostile attribution bias in young children? *Journal of Abnormal Child Psychology, 47,* 989-1000. https://doi.org/10.1007/s10802-018-0499-2

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. https://doi.org/10.1037/0022-3514.54.6.1063

Watson, D., & Pennebaker, J. W. (1989). Health complaints, stress, and distress: Exploring the central role of negative affectivity. Psychological Review, 96(2), 234–254. https://doi.org/10.1037/0033-295X.96.2.234

Wegener, D. T., Dunn, M., & Tokusato, D. (2001). The flexible correction model: Phenomenology and the use of naive theories in avoiding or removing bias. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton Symposium on the Legacy and Future of Social Cognition* (p. 277–290). Lawrence Erlbaum Associates Publishers.

Wegener, D.T., & Petty, R.E. (1995). Flexible correction processes in social judgment: The role of naïve theories in corrections for perceived bias. *Journal of Personality and Social Psychology*, *68,* 36-51. https://doi.org/10.1037/0022-3514.68.1.36

Wells, A., & King, P. (2006). Metacognitive therapy for generalized anxiety disorder: An open trial. *Journal of Behavior Therapy and Experimental Psychiatry*, *37,* 206–212. https://doi.org/10.1016/j.jbtep.2005.07.002

Wells, A., & Matthews, G. (1996). Modelling cognitive in emotional disorder: The S-REF model. *Behaviour Research & Therapy*, *34*, 881–888. http://dx.doi.org/10.1016/S0005-7967(96)00050-2

Wiers, R.W., Van Dessel, P., & Köpetz, C. (2020). ABC training: A new theory-based form of cognitive-bias modification to foster automatization of alternative choices in the treatment of addiction and related disorders. *Current Directions in Psychological Science*, 29, 499–505. https://doi.org/10.1177/0963721420949500

Wigboldus, D. H. J., Sherman, J. W., Franzese, H. L., & van Knippenberg, A. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition, 22*, 292–309. https://doi.org/10.1521/soco.22.3.292.35967

Wu, Q., Ran, G., Zhang, Q. (2020). Rejection sensitivity and trait anxiety: The indirect effects of regulatory emotional self-efficacy and shyness. *Current Psychology*. Advance online publication. https://doi.org/10.1007/s12144-020-01070-y

Xia, L., Mo, L., Wang, J., Zhang, W., & Zhang, D. (2020). Trait anxiety attenuates response inhibition: Evidence from an ERP study using the Go/NoGo task. *Frontiers in Behavioral Neuroscience*, *14,* 1-9. https://doi.org/10.3389/fnbeh.2020.00028

Zinner, L. R., Brodish, A. B., Devine, P. G., & Harmon-Jones, E. (2008). Anger and asymmetrical frontal cortical activity: Evidence for an anger-withdrawal relationship. *Cognition and Emotion*, *22*(6), 1081–1093. https://doi.org/10.1080/02699930701622961

Tables

Table 1

Correlations, means, and standard deviations of Study 1 variables

Variable	1	2	3	4	5	6	7	8	Mean	SD
1. Trait Anxiety									34.17	10.95
2. WSAP Hostility	.31**								110.65	24.03
3. WSAP Benign	.08	.07							132.70	19.18
4. AQ Hostility	.58**	.48**	-.01						19.08	6.87
5. AQ Anger	.48**	.27**	-.07	.57**					14.44	4.95
6. AQ Physical Aggression	.30**	.27**	.01	.38**	.48**				17.82	6.09
7. AQ Verbal Aggression	.32**	.23**	.07	.41**	.50**	.48**			12.04	3.66
8. Positive Affect	-.15*	-.04	.29**	-.24**	-.20**	.004	.14*		28.62	9.34
9. Negative Affect	.76**	.31**	.12	.59**	.46**	.17*	.22**	-.16*	20.16	7.88

Note. WSAP: Word Sentence Association Paradigm. AQ: Aggression Questionnaire. **$p \leq .01$; *$p \leq .05$

Table 2

Correlations, means, and standard deviations of Study 2 variables

Variable	1	2	Mean	SD
1. Trait Anxiety			33.16	10.65
2. WSAP Hostility	.08		108.55	22.56
3. WSAP Benign	.06	-.06	131.18	19.74

Note. WSAP: Word Sentence Association Paradigm. **$p \leq .01$; *$p \leq .05$

Figure Captions

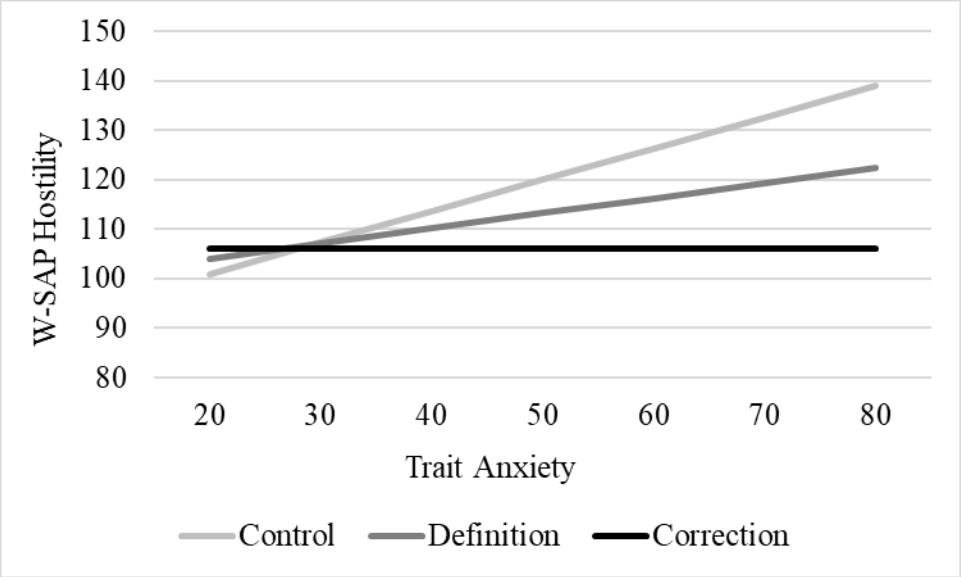

Figure 1. Interaction between trait anxiety and condition on W-SAP Hostility, comparing the true control condition from Study 1 against the two manipulation conditions from Study 2 and excluding data from Study 2 participants who failed the manipulation check.