

# Implicit Theories of Personality and Attributions of Hostile Intent: A Meta-Analysis, an Experiment, and a Longitudinal Intervention

David S. Yeager  
*University of Texas at Austin*

Adriana S. Miu  
*Emory University*

Joseph Powers and Carol S. Dweck  
*Stanford University*

Past research has shown that hostile schemas and adverse experiences predict the hostile attributional bias. This research proposes that seemingly nonhostile beliefs (implicit theories about the malleability of personality) may also play a role in shaping it. Study 1 meta-analytically summarized 11 original tests of this hypothesis ( $N = 1,659$ ), and showed that among diverse adolescents aged 13–16 a fixed or *entity* theory about personality traits predicted greater hostile attributional biases, which mediated an effect on aggressive desires. Study 2 experimentally changed adolescents' implicit theories toward a malleable or *incremental* view and showed a reduction in hostile intent attributions. Study 3 delivered an incremental theory intervention that reduced hostile intent attributions and aggressive desires over an 8-month period.

The *hostile attributional bias* is the tendency to interpret ambiguous provocation as intentional, that is, to view others' negative actions toward you as purposeful and hostile when their intention is unclear (Dodge, 1980, 2006). For example, imagine an adolescent who is walking down the hallway in his school. A peer runs down the hallway, knocking him over and spilling his books on the floor, causing other peers to laugh (Dodge, 2006, p. 791). Was this done on purpose or was it an accident? What would the student do in response?

Over a hundred studies have demonstrated that a hostile attributional bias is a predictor of the desire to enact reactive aggression. This has been shown in response to hypothetical scenarios, laboratory provocations, and provocations in real-world settings (Dodge, 2006; Dodge, Coie, & Lynam, 2006). Research has also shown that changing this hostile bias toward more of a benign attribution of intent can reduce children's and adolescents' reactive aggression (e.g., Hudley & Graham, 1993; for a

meta-analysis, see Wilson & Lipsey, 2006). Combined, these constitute some of the most robust, generative, and important findings in all of developmental psychology. Indeed, this research has provided leverage for understanding and addressing peer aggression, one of the most pressing problems facing today's youth.

But where does the hostile attributional bias come from? Dodge (2006) has theorized that past hostile experiences, such as abuse or long-term exposure to violent contexts, can lead to hostile schemas that produce a heightened vigilance to peer hostility and therefore promote hostile intent attributions. Although such hostile experiences and schemas are undoubtedly influential, we propose that a hostile attributional bias can also emerge from more seemingly nonhostile sources, such as an implicit theory about whether people's traits are fixed and unchangeable. More specifically, we propose that adolescents who hold an *entity* theory of personality—the idea that people's traits cannot change (Dweck, Chiu, & Hong, 1995; Erdley & Dweck, 1993; Yeager, Trzesniewski, Tirri, Nokelainen, & Dweck, 2011)—will be more likely to attribute ambiguous provocations to a peer's hostile intent and, because of this, to express a greater

---

Support for this research was provided by the Spencer Foundation and the Thrive Foundation for Youth. The authors are grateful to the students and teachers who participated in this research and to April House, April Scott, Kerry Morrissey, Shannon Morrissey, Ahmad Saleh, Shannon Brady, Cary Catching, Liz Chamberlain, Brian Spitzer, Kali Trzesniewski, Michelle Harris, Jessica Reed, and Rebecca Johnson for their assistance conducting this research.

Correspondence concerning this article should be addressed to David Yeager, 108 E. Dean Keeton Stop A8000, Austin, TX 78712-1043. Electronic mail may be sent to yeager@psy.utexas.edu.

© 2013 The Authors

Child Development © 2013 Society for Research in Child Development, Inc. All rights reserved. 0009-3920/2013/xxxx-xxxx

DOI: 10.1111/cdev.12062

desire for aggressive revenge. This is because those with more of an entity theory see people's good and bad behaviors as emanating from enduring traits rather than being a product of circumstances (Erdley & Dweck, 1993; see also Chiu, Hong, & Dweck, 1997; Levy & Dweck, 1999), and so they may then be more likely to conclude that a peer who upset them is a bad person who upset them on purpose. If this were true, then one method for reducing attributions of hostile intent might be to change adolescents' implicit theories toward more of a malleable or *incremental* view of personality traits.

To test these proposals, we conducted the present research. Study 1 was a meta-analysis of 11 original correlational tests of the hypothesis that an entity theory would predict greater attributions of hostile intent following both imagined and experienced provocations. Study 1 also explored whether these increased hostile attributions might statistically mediate the effect of an entity theory on the desire to respond aggressively. Study 2 addressed the causal role of implicit theories. It experimentally changed adolescents' theories toward an incremental view and observed differences in attributions of hostile intent. Study 3 extended this by conducting a longitudinal experiment that measured the effects of an incremental theory intervention on hostile attributions and a desire for vengeance over an 8-month period.

#### *Schemas That Shape Hostile Attributional Biases*

As noted, Dodge (2006) has argued that attributions of hostile intent are a function of hostile schemas that result in part from negative life events and that hostile attributional tendencies mediate the impact of these negative schemas on behavior. For instance, research has documented that socialization from parents who, themselves, have hostile attributional tendencies can predict children's increased hostile attributional styles (MacBrayer, Milich, & Hundley, 2003), as can frequent experiences of peer victimization (Yeung & Leadbeater, 2007), or the experience of abuse during childhood (Dodge, Bates, & Pettit, 1990; see Dodge, 2006, for a review). Relatedly, other research has shown that hostile "schemas"—such as the chronic accessibility of hostile thoughts—can predict attributions of hostile intent and because of this predict aggressive desires or behavior (Burks, Laird, Dodge, Pettit, & Bates, 1999; Dodge, Laird, Lochman, & Zelli, 2002; see Dodge, 2006, for a review). Overall, this is important evidence that hostile experiences and schemas

might produce hostile attributional tendencies. However, how complete is this explanation of the origins of the hostile attributional bias?

#### *Implicit Theories*

We propose that other, seemingly nonhostile schemas—ones that may or may not arise from hostile experiences and are distinct from overall negative views about the social world—could also contribute to hostile intent attributions. Specifically, we argue that implicit theories of personality, which involve a theory about the fixedness versus malleability of personality traits, can promote hostile intent attributions when adolescents are confronted with peer provocations of ambiguous intent.

Implicit theories are core beliefs about the malleability of people's traits, and they frame people's interpretations of events in their social worlds. As a result, they play a role in shaping judgments and reactions to other people's behaviors (Chiu et al., 1997; Erdley & Dweck, 1993; Yeager et al., 2011). For instance, adults with more of an entity theory often interpret minor behaviors as having a stronger positive or negative valence (Hong, Chiu, Dweck, & Sacks, 1997) and as more indicative of underlying moral character (Chiu et al., 1997), compared to adults with more of an incremental theory. Past research is consistent in showing that an entity theory about personality creates a psychological world in which people's global character can be judged as good or bad from even thin slices of behavior.

Building on this research, we propose that although an entity theory of personality itself does not necessarily arise from hostile influences, it can nevertheless create a psychological preparedness toward hostile judgments of others. Consistent with this prediction, Erdley and Dweck (1993, Study 2) found that fourth- and fifth-grade students who held more of an entity theory viewed a peer's antisocial behavior as arising from the peer's underlying, stable, and deficient traits. They then displayed less empathy for the peer and prescribed more punishment for him or her. Moreover, when later confronted with positive behaviors performed by the peer—thus providing participants with an opportunity to view the peer's previous antisocial behavior as due to circumstances rather than traits—those with more of an entity theory maintained their global negative trait judgments of the antisocial peer.

More recently, Yeager et al. (2011) examined the effect of implicit theories of personality on high

school students' responses to personal experiences of peer conflict. They showed that those with more of an entity theory were more likely to attribute instances of victimization directed at them to a peer's personal qualities. This characterological attribution mediated an effect on a greater desire to "get back at," "hurt," or "punish" the transgressing peer. Next, Yeager et al. (2011) used a brief experiment to change adolescents' implicit theories toward more of an incremental view, and showed that the incremental theory reduced the belief that a hypothetical peer who bullied them had negative personal qualities. This, in turn, reduced the desire for revenge.

Although instructive, this past research did not investigate the role of implicit theories in shaping the interpretation of negative events in which the intent of an anonymous peer was ambiguous—which is the critical test of an attributional *bias* (cf. Dodge, 1980). The Yeager et al. (2011) scenarios involved peers who explicitly and unambiguously were bullying the participants repeatedly on purpose. Thus, it is currently unknown whether implicit theories of personality would predict the hostile attributional bias. Moreover, both the Erdley and Dweck (1993) and the Yeager et al. (2011) research relied on hypothetical scenarios and did not measure attributions following actual experiences of peer conflict. Therefore, the present studies extended past research by testing whether implicit theories might predict attributions of hostile intent following ambiguous provocations from unknown peers in hypothetical scenarios and in controlled behavioral experiences of peer exclusion.

#### *The Origins of Implicit Theories*

Although the developmental origins of an entity theory are not fully known, some past research supports the notion that entity theories can result from nonhostile or even positive influences, and yet nonetheless predict negative attributions and reactions. In analogous past experiments on implicit theories of *intelligence*, positive, well-intentioned praise for being "smart" could induce an entity theory of intelligence (Mueller & Dweck, 1998). Moreover, Dweck et al. (1995) reported that a fixed or entity theory about intelligence is not simply the residue of poor academic performance. Many high-achieving students hold an entity theory about their academic ability, yet those with more of an entity theory of academic ability, regardless of their actual levels of ability, tend to respond more negatively in the face of academic challenges, for example, by

blaming their failures on themselves, lying about low grades, considering cheating, or giving up (Blackwell, Trzesniewski, & Dweck, 2007; Mueller & Dweck, 1998; for an overview, see Dweck, 2006). More directly relevant to the present research, Chiu et al. (1997, Study 5) changed participants' implicit theories of *personality* with a brief scientific article summarizing longitudinal studies that either showed that people's traits become set, like plaster (in an entity theory condition) or that they can be molded, like clay, throughout life (in an incremental theory condition). This brief article led participants to adopt entity versus incremental frameworks when judging others. Hence, past research is suggestive that implicit theories can arise even from positive or nonhostile influences (such as scientific information) and are not only the product of hostile experiences. Building on this past research, we propose that adolescents who have learned more of an entity theory of personality, regardless of the frequency of past negative events or their overall negative schemas about people, will respond to ambiguous peer conflicts with a different pattern of attributions and behavioral desires than those who have learned more of an incremental theory.

#### *The Present Research*

Study 1 reports 11 original correlations that were collected from eight independent samples of adolescents. Five of these samples included one test of the relation between implicit theories and the hostile attributional bias, and three included two tests employing measures collected at different times. These 11 tests of our hypothesis employed various questions, stimuli (e.g., scenarios and behavioral provocations), and lengths of time between measurements, and they were aggregated into a single effect size estimate using meta-analytic methods, to facilitate more general conclusions. In addition, to test for whether implicit theories would predict attributions among adolescents from relatively hostile as well as relatively nonhostile contexts, Study 1 included samples from both lower violence and higher violence neighborhoods. Last, Study 1 meta-analytically summarized tests of whether attributions of hostile intent mediated the effect of implicit theories on aggressive desires.

Study 2 was an experiment designed to test the causal effect of implicit theories on hostile attributional biases. In it we changed implicit theories toward more of a malleable or incremental view of personality and then measured short-term changes in attributions of hostile intent. As in Study 1, this

experiment was conducted in two contexts with highly different levels of neighborhood violence, to demonstrate the generality of this process. Finally, Study 3 was a longitudinal experiment that administered a brief (two class session) incremental theory intervention and measured attributions of hostile intent and vengeful desires 8 months later.

## Study 1

### *Method*

#### *Participants*

Participants were students in Grades 8, 9, and 10 from eight independent samples coming from six different schools in the United States (samples collected from the same schools included different students surveyed in different years;  $N = 1,128$  unique participants provided a total of  $N = 1,659$  observations). See Table 1 for a summary of sample characteristics. Some of these studies were conducted expressly to test the present hypotheses, and others were designed to test additional hypotheses that will be included in future research articles. However, critically, all the samples included here were administered measures of implicit theories and attributions of hostile intent, and none of the relations between these variables have been reported in another research article.

In the aggregate, 49% of participants were female. Twelve percent self-identified as Black or African American, 17% as White, 26% as Asian, 42% as Hispanic or Latino, and the rest as another race or ethnicity. Schools were located in New York City or the San Francisco Bay Area. In six of the eight samples, the majority of students at the school received free or reduced-price lunch, an indicator of

low socioeconomic status, and came from neighborhoods with violence rates above the national average, according to federal statistics (United States Department of Justice, Federal Bureau of Investigation, 2011). In the remaining two samples the majority of students came from middle- or high-income families and lived in neighborhoods with violence rates below the national average. Response rates in all studies were over 75%. Parental consent and student assent were obtained for all participants.

#### *Procedure*

In all studies, students completed activities and questions as a class on computers in a school computer lab during school hours. The session was administered by a trained research assistant who led students to work quietly and privately for the roughly 30 min required. Four of the studies included an ambiguous provocation that students believed was a real interaction with another student (described below). In each case, students were thoroughly debriefed after the experience. At the end of all of the studies, students were thanked for their participation and were told how their responses would be used to advance science and help future students like themselves.

#### *Measures*

##### *Entity Theory*

Participants in all studies answered five questions measuring an entity theory of personality traits relevant to peer conflicts in high school. These questions were developed and used in past research (Yeager et al., 2011). Participants were asked to

Table 1  
*Samples Summarized Meta-Analytically in Study 1*

Sample	<i>N</i>	Grade level	Location	% female	% Black	% White	% Asian	% Hispanic or Latino
1	39	9th	New York, NY	63	36	8	28	28
2	128	9th–10th	Bay Area, CA	49	28	4	30	38
3	179	9th–10th	Bay Area, CA	53	9	13	9	58
4	204	9th	New York, NY	51	22	12	26	35
5	36	9th	Bay Area, CA	46	3	50	41	6
6	305	9th	Bay Area, CA	45	2	35	46	15
7	211	9th	Bay Area, CA	49	5	4	6	80
8	26	8th	Bay Area, CA	51	3	7	10	80

*Note.* Participants could indicate as many races or ethnicities as they wished, or none at all, and therefore row percentages need not add up to 100%. Samples 1, 2, 3, 4, 7, and 8 were collected from schools located in neighborhoods with higher levels of poverty and violence, whereas samples 5 and 6 were located in upper-middle-class neighborhoods with lower levels of poverty and violence.

agree or disagree with statements such as “Bullies and victims are types of people that really can’t be changed” (1 = *strongly disagree*, 6 = *strongly agree*; below we report analyses showing that implicit theories are distinct from overall negative schemas about people). These items were internally consistent across studies:  $\alpha$ s ranged from .66 to .83.

#### *Attributions of Hostile Intent*

In each of the tests of our hypothesis, one of two different ambiguous provocations was used to evoke attributions of hostile intent. One was a hypothetical scenario and one was a behavioral experience of social exclusion. Following the design employed in past demonstrations of an attributional bias (e.g., Dodge, 1980), in each of these provocations the identity of the peer was unknown to the participant. Indeed, a crucial precondition for a test of an attributional *bias* is to examine person perceptions in the absence of any previous knowledge about the character of the provoking peer.

*Hypothetical scenario.* Seven tests of our hypothesis employed an ambiguous scenario. This was a paradigmatic scenario used to illustrate the hostile attributional bias and was based on an example described by Dodge (2006, p. 791):

Imagine that you were walking in a crowded hallway at school and everybody was rushing to get to the next class so they wouldn’t be late. While you were looking the other way, you and another student bumped into each other (pretty hard), so it hurt your shoulder and you dropped the books that you were carrying. The other student paused briefly, looked at you quickly, and then turned away and hurried to class.

A pilot experiment showed that this stimulus was, in fact, ambiguous with regard to hostile intent (see online Supporting Information).

*Cyberball exclusion.* In four tests of our hypothesis, students experienced a standardized (but short-lived and harmless) instance of social exclusion during an online game of catch (“Cyberball”) with two “peers” purportedly from their secondary school but located in another room (Williams, 2009; Williams, Yeager, Cheung, & Choi, 2012). In reality, there were no other human players and the game was programmed so that the computer players threw the ball to the participant twice at the outset and then threw it exclusively to each other. From the perspective of the participant, the hostility of the other player’s intentions was highly ambiguous.

Participants did not communicate with the excluders before, during, or after the game and were aware that the other players did not know their identity. In this way, participants were provided no clues as to the identities or intentions of the peers who excluded them, producing a great deal of latitude for them to fill in the details (cf. Heider & Simmel, 1944). Note that a strength of this method is that it more narrowly isolates attributions of *hostile* intent. Indeed, even when students believed that the Cyberball exclusion was done *on purpose*, not all of these students believed that it was done on purpose *in order to be mean*. In a pilot study, we interviewed participants about their impression of the other players’ motives following Cyberball. In this pilot, roughly half of students interpreted Cyberball exclusion as arising from hostile intent, saying “they were trying to make me feel bad by leaving me out.” Yet other students thought the exclusion arose from a joke that was not intended to be malicious, from a computer malfunction, or because the other players did not understand the rules of the game. Hence, even when students believed that the peers meant to exclude them, there was variation in whether those intentions were attributed to hostility. In sum, exclusion via Cyberball allowed for important variety in our tests of the relation between implicit theories and attributions of hostile intent.

*Measures of intent following the provocations.* After reading the scenario or experiencing peer exclusion, participants rated attributions of hostile intent. These questions varied across studies to reduce the possibility that the results were an artifact of a certain question type (see online Supporting Information Table S1 for more detail).

In two of the tests of the hypothesis that employed the ambiguous scenario, participants rated how much they thought “this person bumped into you on purpose or on accident” on a 5-point scale (1 = *a complete accident*, 5 = *completely on purpose*). In the five remaining tests that employed the scenario, participants rated on a 7-point scale how much they thought “They were being mean to me on purpose” and “They were trying to be mean to me” (1 = *not at all*, 7 = *an extreme amount*), and these two questions were combined by taking their unweighted average ( $\alpha$ s > .8 across studies). These items were modeled after commonly used measures of the hostile attributional bias (Pettit, Lansford, Malone, Dodge, & Bates, 2010).

In the four tests of our hypothesis that employed a behavioral experience of Cyberball exclusion, participants answered two items such as “They were

being mean to me on purpose" (1 = *Not at all*, 7 = *An extreme amount*). These two questions were combined into a single measure ( $\alpha s > .8$  across studies).

### *Aggressive Desires*

Seven tests of our hypothesis measured a type of aggressive desire following the ambiguous provocation.

*Desire for vengeance.* In five tests of the hypothesis, a desire for vengeance was measured following the provocation. Participants rated how much they felt like "hurting," "punishing," or "getting back at" the peer in the ambiguous provocation scenario (1 = *not at all*, 7 = *an extreme amount*; items were averaged,  $\alpha s > .8$  across studies). In past research, the desire for vengeance was strongly correlated ( $r = .50$ ) with self-report measures of real-world reactive aggressive behavior (Yeager et al., 2011, Study 1).

*Desire to use vengeance for emotion regulation.* Recent theories of aggressive behavior have noted the importance of anticipated positive emotional states following aggression as a precursor for aggressive behaviors (for a review, see Baumeister, Vohs, DeWall, & Zhang, 2007; see also Yeager et al., 2011, Study 3). Therefore, to broaden the scope of our measures of aggressive desires, in two tests of our hypothesis a desire to use vengeance for emotion regulation was measured (note that the "desire for vengeance" was not measured in these two samples). Participants rated five items such as "It would make me feel better to think about hurting them back" and "All I would have to do is imagine myself taking revenge and I would feel better" (1 = *strongly disagree*, 7 = *strongly agree*). These five items were averaged, with higher values corresponding to a greater desire for vengeance ( $\alpha s > .8$  across studies).

### *Peer Victimization*

Students in nine of our tests were asked to report levels of peer victimization by answering four items (including how often they have "been hit, slapped, or pushed" or "been excluded or left out by other students" in the "past month or so"), each on a 5-point scale from 0 = *never* to 4 = *very often*. These items were averaged,  $\alpha = .79$ , with higher values corresponding to more victimization.

### *Analysis Plan*

This study utilizes a method employed in recent investigations that have conducted a set of original

studies and then summarized them using the statistical method of meta-analysis to reach more general conclusions about the phenomenon under study (e.g., Yeager & Krosnick, 2011, 2012). Our analyses were performed in the following manner. First, ordinary least squares regressions were conducted for each of the key hypotheses (e.g., an entity theory predicting attributions of hostile intent). (Individual unstandardized regression coefficients from these models are presented in the online Supporting Information Table S1.) All variables in the study were recoded to range continuously from 0 to 1. The indirect or mediated effects and their 95% confidence intervals were estimated using bootstrapping methods in the *Mediation* package in R (Imai, Keele, & Tingley, 2010; Imai, Keele, Tingley, & Yamamoto, 2011; R Development Core Team, 2008). The variance statistics for these indirect effects was approximated by dividing the range of the 95% confidence interval by 4 and then squaring. Next, effect sizes ( $r$ ) and variances were calculated for each of these tests using the corresponding  $t$  statistics (for the  $a$ ,  $b$ , and  $c$  effects) or variances (for the  $ab$ , or indirect, effects). Finally, these effect sizes were aggregated in separate random-effects meta-analyses, estimated by the *metafor* package in R (Viechtbauer, 2010).

This meta-analytic procedure is desirable for several reasons. First, it guards against publication bias—or the tendency to report only studies that had significant effects consistent with hypotheses (Schooler, 2011). Second, it guards against "researcher degrees of freedom"—or idiosyncratic decisions about how to conduct an analysis so as to maximize the significance in a given study, such as which items to drop from aggregate scales or which control variables to include (Simmons, Nelson, & Simonsohn, 2011). Third, it guards against artifactual findings by allowing for tests of homogeneity of effect sizes across different measures, methods, and populations. Given that the hostile attributional bias is one of the most frequently studied constructs in the psychology of aggression, we believed that any claim about the potential origins of this highly important construct should overcome each of these potential criticisms. Therefore, our meta-analyses include all tests that we have conducted of the key hypotheses, even ones that did not reach statistical significance. No study we have run was excluded, and, as noted, these studies employed different methods to allow for more general conclusions. Relatedly, we present statistics supporting the idea that no studies were withheld (i.e., we test for "publication bias"). Next, in every case we made

the same decision about control variables across analyses (e.g., we do not include any, except in supplemental analyses testing a specific hypothesis about a potential confounding variable, peer victimization; note that conducting analyses without control variables that reduce error variance in the key dependent variable most likely makes our effect size estimates more conservative). Finally, we made a priori and consistent decisions about the composition of each of the scales employed in our research.

### Results and Discussion

#### *Does an Entity Theory Predict Attributions of Hostile Intent?*

An entity theory significantly predicted attributions of hostile intent across 11 tests ( $N = 1,659$ ), meta-analytic  $r = .18$ ,  $Z = 7.80$ ,  $p < .0001$ . The funnel plot of effect sizes (a test for publication bias) was not asymmetric, Kendall's  $\tau = .38$ ,  $p > .10$ , adding support to our disclosure that no test was withheld from our meta-analysis. Differences in samples or procedures did not affect our key result: test of homogeneity,  $Q(10) = 15.54$ ,  $p > .10$ . For instance, implicit theories significantly predicted hostile intent attributions when both were measured on the same day, when attributions were measured after 3 weeks, and when attributions were measured after 1.5 years (see online). The finding of nonsignificant heterogeneity was supported by planned tests of moderation by ambiguous provocation stimulus (hypothetical scenario vs. Cyberball),  $Q(1) = .002$ ,  $p = .97$ , and by sample composition (samples with higher neighborhood violence vs. samples with lower neighborhood violence),  $Q(1) = .001$ ,  $p = .96$ .

Next, classic investigations of the hostile attributional bias have focused on boys only (Dodge, 1980; Hudley & Graham, 1993). We therefore meta-analyzed Sex  $\times$  Implicit Theories interactions from each of the 11 regressions predicting hostile intent and found no significant moderation by sex, meta-analytic  $r = .01$ ,  $Z = .31$ ,  $p = .75$ .

Some research has found that peer victimization predicts greater hostile attributional biases (e.g., Yeung & Leadbeater, 2007). Yet when we reconducted our meta-analysis controlling for adolescents' reported frequency of victimization, our results were unchanged,  $r = .17$ ,  $Z = 6.36$ ,  $p < .0001$ .

We have claimed that the seemingly nonhostile implicit theories of personality traits are distinct from explicitly hostile schemas; that is, believing that bad traits cannot change is not the same as

believing that people in general are bad. However, we have not yet presented evidence testing this proposition. Therefore, we conducted an additional study with Grade 9 students at the same school that provided Sample 6 in the present meta-analysis (see Table 1),  $N = 212$ . We measured overall negative schemas about people (three items, e.g., "Overall, how mean and hostile are people in school," "Overall, how much are people in your school out to get you"; 1 = *not at all*, 5 = *an extreme amount*;  $\alpha > .8$ ), in addition to an entity theory of personality, attributions of intent, and frequency of peer victimization. An entity theory predicted greater attributions of hostile intent when controlling for these hostile schemas and for frequency of peer victimization,  $r = .19$ ,  $p < .01$ . Moreover, an entity theory predicted hostile intent attributions even among the subset of 97 students who reported largely nonhostile schemas (i.e., average of *not at all* on the 5-point hostile schema measure),  $r = .20$ ,  $p < .05$ , again controlling for victimization. Thus, implicit theories are distinct from hostile schemas.

Our studies revealed that the size of the effect of an entity theory on attributions of hostile intent was consistent but modest, roughly  $r = .18$ . How does this result compare to extant theories of the origins of the hostile attributional bias? To address this, we compared our study's effect size with the effects highlighted in Dodge's (2006) authoritative summary of hostile schemas and experiences that predict a hostile attributional bias (Burks et al., 1999; Dodge et al., 1990; Dodge et al., 2002; MacBrayer et al., 2003; Yeung & Leadbeater, 2007; —the latter was substituted for Graham & Juvonen, 1998, which was cited by Dodge, 2006, but did not report the necessary effect size). This set of studies was chosen by Dodge (2006) as representative of the documented precursors of an attributional bias. Our calculation showed that the effect of hostile influences on the attributional bias in past research ranged from  $r = .08$  to  $r = .34$ , with a meta-analytic average of  $r = .189$ ,  $Z = 5.05$ ,  $N = 1,362$  (for hostile schemas:  $r = .20$ ; for hostile experiences:  $r = .17$ , moderator  $Q(1) = .12$ , *ns*). Notably, this average effect was not significantly different from the effect of implicit theories found in our studies, moderation test  $p > .50$ . Also note that the past studies summarized by Dodge only included published studies with significant effects and likely constitutes an upward bound on the average size of the effect. If we restricted our meta-analysis of the studies we conducted to only those that yielded a significant effect, we would obtain an effect of  $r = .25$ , which is slightly larger than past studies of hostile schemas

and experiences summarized by Dodge (2006). Hence, the relation between seemingly nonhostile implicit theories and the hostile attributional bias is comparable to prominent examples of the relation between this bias and hostile schemas and experiences.

*Do an Entity Theory and Attributions of Hostile Intent Predict More Aggressive Reactions?*

In 7 of our 11 tests ( $N = 1,102$ ), we measured aggressive desires following an ambiguous provocation. In these 7 tests, there was a significant meta-analytic average effect of an entity theory on hostile intent (*a* paths),  $r = .20$ ,  $Z = 6.79$ ,  $p < .0001$ , and on aggressive desires (*c* paths),  $r = .23$ ,  $Z = 6.61$ ,  $p < .0001$  (effects shown in Figure 1). In addition, conceptually replicating much past research, there was a significant effect of hostile intent on aggressive desires (*b* paths),  $r = .51$ ,  $Z = 11.63$ ,  $p < .0001$ .

*Did Hostile Intent Attributions Mediate the Relation Between an Entity Theory and Aggressive Desires?*

On average, there was a significant indirect (*ab*) effect leading from an entity theory to aggressive desires through hostile intent,  $r = .09$ ,  $Z = 4.29$ ,  $p < .0001$  (see Figure 1), and this was homogeneous across the studies,  $Q(6) = 6.98$ ,  $p = .32$ . For five of the seven tests the effect of implicit theories was reduced to nonsignificance when controlling for intent. The average size of this indirect effect corresponds to a weighted average of 41% of the direct (*c*) effect of an entity theory on aggressive desires. This shows that a significant and substantial proportion of the effect of an entity theory on aggressive

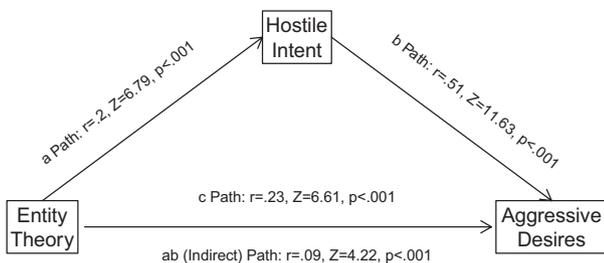


Figure 1. Meta-analytic summary of direct and indirect correlational effects of an entity (fixed) theory of personality on aggressive desires, mediated by attributions of hostile intent (Study 1;  $N = 1,102$ ).  $r$  = meta-analytic correlation coefficient. The meta-analysis of *a* paths depicted in Figure 1 includes only the subset of studies that also measured aggressive desires and so it differs from the meta-analytic *a* path reported in the text that included all studies measuring implicit theories and hostile intent.

desires was accounted for by the relation between an entity theory and hostile intent attributions.

In sum, our findings are consistent with Dodge's (2006) claim that schemas will predict aggression mediated through attributions of hostile intent. However, it extends this past theory by showing that a seemingly nonhostile schema about the fixedness of traits can produce hostile intent attributions at, statistically, the same level that more explicitly hostile schemas and experiences have in past research. As a side note, our study was a methodological advance because it was, to our knowledge, the first or one of the first to meta-analyze mediation effects.

## Study 2

Study 1 showed that adolescents who held more of an entity theory of personality were more likely to interpret ambiguous provocations from unknown peers as arising from hostile intent. In turn, they exhibited more aggressive desires. Study 1 was limited, however, in that it did not address the causal direction of the relation between implicit theories and attributions of hostile intent. Therefore, in Study 2 we conducted an experiment that changed adolescents' implicit theories and then assessed the effect of this change on their attributions.

As in Study 1, we intentionally included samples of adolescents from different environments, each with dramatically different levels of neighborhood hostility. The first sample lived and attended school in an urban section of Oakland, California. According FBI official statistics (United States Department of Justice, Federal Bureau of Investigation, 2011), in 2010, the year of our study, Oakland had the fourth highest per capita violent crime rate in the nation, with one violent crime for every 65 residents. The second sample lived and attended school in a medium-sized city in Georgia that had a much lower rate, at one violent crime for every 250 residents, which is almost exactly the national average. We predicted that changes in implicit theories could affect attributions in both hostile contexts and in more nonhostile contexts.

## Method

### Participants

Sample 1 ( $N = 20$ ) was a racially diverse sample of ninth-grade students from primarily low-income households in Oakland, California. Seven percent self-identified as White, 32% as Hispanic or Latino, 2% as Asian, and 38% as African American, while

the rest marked "Other." Sixty-seven percent were female. Sample 2 ( $N = 43$ ) consisted of primarily White, non-Hispanic seventh-grade students from mid- to high-income families in a medium-sized city in Georgia. Sixty-five percent self-identified as White, 3% as Hispanic or Latino, 3% as Asian, 5% as African American, and the rest marked "Other." Sixty-seven percent were female. Participants came from studies that were conducted only for the present purposes.

### Procedure

*Experimental manipulation.* The incremental theory of personality manipulation was adapted from previous investigations that have changed implicit theories in this age group (Yeager et al., 2011; Yeager, Trzesniewski, & Dweck, in press). The treatment activity, which lasted roughly 15 min on average, had three parts: (a) a brief and factually accurate scientific article that described neuroscientific studies showing that people's behaviors are controlled by thoughts and feelings in their brains, which have the constant potential for plasticity; (b) notes from upperclassmen who described how they endorsed a malleable view of people's characteristics; and (c) an exercise in which students wrote notes to future students drawing on scientific information to describe the malleability of people's traits (i.e., "self-persuasion"; Aronson, 1999; see Walton & Cohen, 2011; Yeager & Walton, 2011). Importantly, nothing in the manipulation addressed attributions of hostile intent, and no emotions or behaviors were explicitly endorsed. Moreover, the intervention did not say that people *would* change, only that there is the *potential* for change (indeed, a content analysis of students' writing samples found almost no instances of saying that peers *will* change, and numerous instances of saying that they *might or might not* change).

A control group completed a parallel writing activity that emphasized the malleability of academic skills such as study skills. This comparison group was used to control for the possibility that simply receiving a message about malleability or growth could have led to our findings. Therefore, the present positive, hopeful control group constituted a rigorous test.

### Measures

#### *Tests of Equivalence of Manipulations*

To show that the manipulations were parallel, we compared the experimental and control manipulations in terms of features unrelated to the experimental

message. That is, after completing the activities, participants were asked to rate how much they "enjoyed," "understood," "tried their hardest to read," and "agreed with" the manipulation on a 5-point scale (1 = *not at all*, 5 = *a great deal*).

#### *Implicit Theories of Personality*

To check the predicted effect of the manipulation on changes in implicit theories, we measured an entity theory as in Study 1.

#### *Attributions of Hostile Intent*

Participants read the same ambiguous provocation scenario described in Study 1. As in Study 1, after reading the scenario participants rated on a 7-point scale how much they thought, "They were being mean to me on purpose" and "They were trying to be mean to me" (1 = *not at all*, 7 = *an extreme amount*), and these two questions were combined by averaging them ( $\alpha s > .8$  in both samples).

#### *Reactions to Ambiguous Provocation*

To provide a richer picture of the different psychologies evidenced by those who learned the incremental theory of personality—one that was not constrained or contaminated by the limited, negative choices offered by the desire for vengeance rating-scale measure used in Study 1—we asked participants to write a free-response description of their reactions to the scenario (answering the question: "How would you feel?"). Their responses were reliably coded for positive, neutral, or negative reactions toward the offending peer. Specifically, each student's entire open-ended response was coded for the presence of statements that fell into the positive or negative categories (percent agreement for each category > 92%; Krippendorff's alphas for each category = .86, .90) by two independent coders who were blind to hypotheses and experimental condition. A detailed codebook and procedure description is available on request. Some participants' responses were coded as *negative* (24%) because they included at least one statement that expressed negative emotions or desired behaviors toward the transgressor, such as anger or a desire for revenge (e.g., they felt "like punching them") and contained no positive statements. Next, some participants' responses were coded as *positive* (35%) because they involved positive reactions toward the transgressor, such as trying to understand, empathize with, or prosocially confront the

transgressor (e.g., “At least they thought about helping me, but I guess they didn’t want to [be] late”), and contained no negative statements. Some participants’ responses did not include any statements that could be classified as positive or negative with regard to the transgressor and so they were classified as *neutral* (35%; i.e., “I would ignore it”; this also included negative responses that were not directed toward the transgressor, e.g., “I would be frustrated”). Last, some participants’ responses were classified as *ambivalent* (6%) because their responses had been coded as including both positive statements and negative statements. For analyses, participants’ responses were given ordered categories such that higher values corresponded to more negative statements (1 = *negative*: participant wrote at least one negative statement and no positive statements, 0 = participant’s response was *neutral* or *ambivalent*, -1 = *positive*: participant wrote at least one positive statement and no negative statements; see the Method section of Study 3 for validity evidence).

### Results and Discussion

We conducted analyses using least squares and ordered logistic regressions. No result was significantly moderated by sample (all Sample  $\times$  Condition interaction,  $t_s < 1$ ,  $p_s > .35$ ) and so the subsamples were combined for all analyses. As in Study 1, no Sex  $\times$  Condition interactions were found ( $t_s < 1$ ,  $p_s > .35$ ). Regression models included a dummy variable indicating sample, however, to account for variance due to sample (models that did not control for sample produced identical statistical significance). Note that all analyses were also rerun when controlling for reported levels of peer victimization (see Study 1 for measures). As in Study 1, this had no effect on any of our conclusions. Therefore, we report the more parsimonious models that did not control for victimization.

#### Preliminary Analyses

The experimental and control groups were equivalent on preexperimental variables (sex, race, and peer victimization; all  $p_s > .30$ ). Participants rated the treatment and control group activities equally in terms of enjoyment, ability to understand the message, effort put forth in reading the message, and agreement with the message, all  $t_s < 1.1$ ,  $p_s > .30$ . Last, and most critically, the experimental manipulation led to a lower belief in an entity

theory about people’s traits compared to the control group, (control:  $M = 0.43$ ,  $SD = 0.21$ ; treatment:  $M = 0.31$ ,  $SD = 0.19$ ),  $t(62) = -2.33$ ,  $p = .02$ ,  $d = .59$ .

#### Did the Incremental Theory of Personality Reduce Attributions of Hostile Intent?

The experimental group, which learned a theory that personality traits have the potential for change, exhibited fewer attributions of hostile intent following the ambiguous provocation scenario (control:  $M = 0.37$ ,  $SD = 0.24$ ; treatment:  $M = 0.22$ ,  $SD = 0.18$ ),  $t(62) = -2.81$ ,  $p = .007$ ,  $d = .71$ . As noted, the treatment effect was statistically the same size across the two samples: Sample 1,  $t(19) = 1.85$ ,  $p = .08$ ,  $d = .84$ ; Sample 2,  $t(42) = 2.15$ ,  $p = .04$ ,  $d = .66$ ; Sample  $\times$  Condition interaction,  $t(59) = -0.87$ ,  $p = 0.39$ .

#### Does an Incremental Theory of Personality Change Responses to Ambiguous Provocations?

In an ordered logistic regression predicting negative versus positive reactions to the ambiguous provocation scenario, the experimental condition showed a lower probability of negative reactions and a greater probability of positive reactions, unstandardized logistic  $b = -1.05$ ,  $SE = 0.49$ ,  $Z = -2.14$ ,  $p = .03$  (see Figure 2a). More concretely, the experimental group was half as likely to spontaneously describe retaliatory aggression such as “I would be [angry] and say foul stuff and probably hit the person.” Instead of anger or aggression, the experimental group was more likely to say they would understand the peer’s actions, such as “Everyone’s trying to get to class and so am I, no one wants to be late, I would of [sic] done the same.”

#### Test of Mediation

We next tested whether the changes in reactions to the ambiguous provocation were mediated by changes in attributions of hostile intent. In an ordered logistic regression, attributions of hostile intent predicted more negative and fewer positive reactions to the provocation,  $b = 3.22$ ,  $SE = 1.30$ ,  $Z = 2.49$ ,  $p = .01$ . Next, in a model with both attributions of intent and experimental condition, intent remained significant,  $b = 2.70$ ,  $SE = 1.34$ ,  $Z = 2.01$ ,  $p = .04$ , but the experimental condition dropped to nonsignificance,  $b = -.75$ ,  $SE = 0.52$ ,  $Z = -1.45$ ,  $p = .15$ . In a causal mediation analysis (Imai et al., 2010; Imai et al., 2011), the indirect effect of condition on the polytomous outcome mediated through

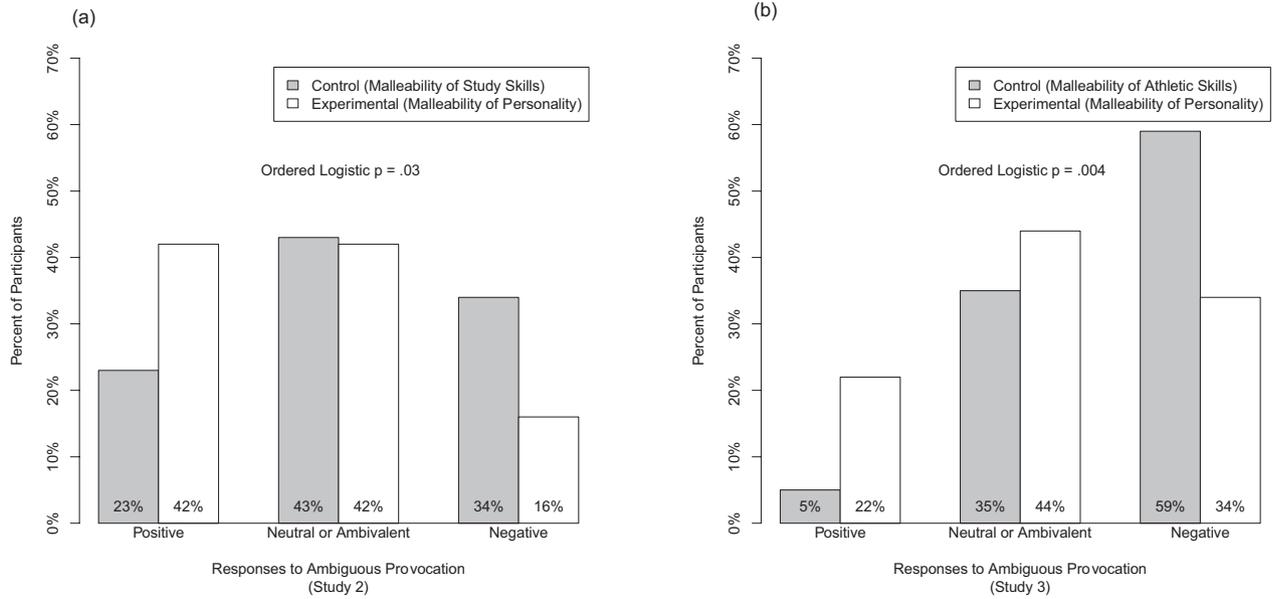


Figure 2. Effect of incremental (malleable) theory of personality manipulation on open-ended responses to an ambiguous peer provocation. Positive = understanding or prosocial responses to the peer’s provocation; Negative = angry or vengeful responses to the peer’s provocation; Neutral or ambivalent = other responses or participants who mentioned both a positive and a negative response. Values are raw, unadjusted percentages. (a) Study 2,  $N = 63$ , a short-term (one class-session) experiment; (b) Study 3,  $N = 78$ , a longitudinal (8-month) intervention.

attributions of hostile intent was significantly different from zero, indirect effect =  $-.06$  (Quasi-Bayesian 95% CI  $[-1.65, -.001]$ ). Hence, it was in part by reducing attributions of hostile intent that the incremental theory of personality manipulation reduced negative reactions to the ambiguous provocation scenario.

### Study 3

Study 2 showed that changing implicit theories toward more of an incremental view could reduce attributions of hostile intent assessed a short time later in an urban context and in a suburban context. However, an important question is whether an incremental theory intervention can change attributions of hostile intent and aggressive desires over longer periods of time. As a first test of this possibility, we took advantage of an ongoing longitudinal experiment that was examining the impact of an implicit theories intervention (nearly identical to Study 2’s manipulation) on several outcomes over the 1st year of high school. On the final survey for this study, we were able to administer the ambiguous provocation scenario employed in Study 2 and we assessed the effect of an incremental theory on hostile intent and vengeful desires.

### Method

#### Participants

The experiment was conducted with all ninth-grade students enrolled in Algebra 1 at a secondary school in the San Francisco Bay Area ( $N = 82$ ; all students assented, and all but 4 participants provided complete data, leaving  $N = 78$  for analysis). Parental consent and student assent were obtained for all participants. The student body was largely White or Asian and relatively affluent (44% White, 44% Asian American, 12% Latino). As noted, this sample will provide data to test other hypotheses, but, critically, the data presented were collected expressly for this study and have not appeared in any other manuscript. Next, although this suburban sample may have been predicted to experience less of an attributional bias based on past theory (Dodge, 2006), recall that Studies 1 and 2 found the effect of implicit theories on hostile intent attributions did not vary in strength across urban and suburban samples, making this an informative sample in which to conduct a first test of whether an implicit theories manipulation can produce lasting effects in at least one setting.

#### Procedure

Surveys were administered at baseline, during the 2nd week of school and again 2 days postintervention

(to obtain manipulation checks), as well as 8 months later, in the last month of school (to measure the key dependent variables). The intervention procedure was as follows: During the 1st week of school, Algebra 1 teachers gave an overview to both the treatment and control groups of how the brain changes and learns, and this provided background for both the treatment and control messages. About 2 weeks later, students received envelopes (randomly assigned within classes) containing either a treatment activity or a control activity, which they completed silently during class time. The treatment activity was identical to the treatment activity described in Study 2 (i.e., it involved reading a scientific article that summarized neuroscientific information, reading quotes from upperclassmen, and writing a persuasive message to future students endorsing the incremental theory). Recall that this activity made no mention of intent in general and more specifically it did not address accidental or ambiguous actions. Instead, it focused on the function and malleability of the brain and the implications of this for understanding people's character traits.

The control activity was nearly identical to the control activity employed in Study 2, except that it endorsed a different malleable theory: one about athletic skills, rather than study skills, which was used in Study 2. This was done because an additional objective of this study (to be reported in another manuscript) was to investigate the effect of the intervention on academic performance, and so a nonacademic control group was required.

### Measures

#### *Implicit Theories of Personality*

The same measure to assess students' beliefs in an entity theory of personality (described in Study 1) was administered 1 week preintervention, 2 days postintervention, and on the year-end survey (for analyses of the latter, see online Supporting Information).

#### *Attributions of Hostile Intent and Desire for Vengeance*

Eight months after the intervention, we measured hostile intent attributions following the same hypothetical scenario employed in Studies 1 and 2 (i.e., depicting an ambiguously intended collision with a peer in the hallway). We note that although the present sample was a relatively affluent one, the percent of students in the control group who

said the peer was *not at all* being mean on purpose and hence showed a completely nonhostile attribution—62%—is comparable with and not significantly different from answers provided on an identical measure by Study 1's participants (e.g., the largely low-income, racial minority Sample 7 reported in Study 1, had a nonhostile attribution rate of 54%). In this study, when the control and treatment groups were combined, the posttreatment attributional measure was highly skewed, with 71% of students choosing the lowest point on the scale, indicating that they believed the peer's negative action was completely accidental. Due to skew, we dichotomized the hostile intent postexperimental outcome such that a score of 0 indicated that a student perceived no hostile intent whatsoever (i.e., average of *not at all* mean on purpose) and a score of 1 indicated a student perceived at least some hostile intent (average greater than *not at all*; note that Study 1's and 2's results were unchanged when we reanalyzed them with a dichotomized hostile intent variable).

Next, so as to test the potential causal relation suggested by the correlational tests in Study 1, we administered the Study 1 rating scale measure of a desire for vengeance following the ambiguous hypothetical provocation (asking, e.g., how much they would feel like "hurting," "getting back at," or "punishing" the transgressor; see Yeager et al., 2011, for validity data). This continuous measure had high internal consistency reliability ( $\alpha = .81$ ) and was not skewed.

At the same time, we sought to replicate longitudinally Study 2's effect of an incremental theory manipulation on open-ended reactions to the provocation scenario. Therefore, we again asked participants to write free responses about how they would feel after the provocation, and two independent coders (blind to condition and hypotheses) followed our extensive coding process to reliably categorize the free responses as positive, negative, or neutral or ambivalent (percent agreement for each > 92%; Krippendorff's  $\alpha$ s = .84, .84). In support of the validity of this measure, it was significantly correlated with ratings of a desire for vengeance,  $r = .30, p = .009$ .

### Results and Discussion

To reduce error variance, regressions in Study 3 included dummy variables for participant race, participant gender, and classroom, while also controlling for baseline entity theories. These reduced the standard errors associated with treatment effects, although models that did not control for any covariates produced a virtually identical overall pattern of results.

### *Preliminary Analyses*

Randomization was effective. There were no baseline differences between the groups in terms of gender, race, mother's education, primary language at home, eighth-grade math and English achievement, eighth-grade absences, belief in an entity theory of personality, or peer exclusion or victimization (statistical tests available on request). Also, as predicted, the incremental theory intervention led students to express less of a belief in an entity theory of personal traits (control  $M$  [covariate adjusted] = 2.86,  $SD$  = 1.20; treatment  $M$  = 2.39,  $SD$  = 1.09),  $t(76) = -2.05$ ,  $p = .04$ ,  $d = -.39$ , controlling for baseline levels. Thus, our experiment constituted a fair test of whether implicit theories could change attributions of hostile intent.

### *Did the Intervention Affect Reactions to an Ambiguous Provocation 8 Months Postintervention?*

Experimental group participants, who learned during the 1st month of school that people have the potential to change, were significantly less likely to attribute a hypothetical peer's negative action to hostile intent at the end of the school year, 8 months later (control = 38%; experimental = 22%; percentages are raw, unadjusted values), unstandardized logistic  $b = -1.13$ ,  $SE = 0.56$ ,  $Z = -1.99$ ,  $p = .047$ . Next, and perhaps more importantly, experimental group participants also showed greatly lower ratings of a desire for revenge 8 months later compared to the control group, unstandardized  $b = -.09$ ,  $SE = 0.03$ ,  $t(76) = -2.72$ ,  $p = .008$ ,  $d = .62$ .

We predicted that the intervention would have these effects because it led participants to hold less of an entity theory compared to the control group, and this is what a mediational analysis showed. An entity theory of personality measured 1–2 days postintervention predicted greater attributions of hostile intent and desire for vengeance 8 months postintervention,  $r = .25$ ,  $p = .03$ , and  $r = .35$ ,  $p = .001$ , respectively. Crucially, the treatment effect on immediate changes in an entity theory (1–2 days postintervention) statistically mediated the longitudinal treatment effect on both attributions and desire for vengeance 8 months postintervention, indirect  $b = -.03$  (95% CI [-.09, -.001]), indirect  $b = -.02$  (95% CI [-.05, -.001]), respectively, and in both cases the treatment effect on these dependent variables was reduced to nonsignificance when accounting for immediate change in entity theories. This mediational result supports our claim that it was the immediate difference in implicit the-

ories produced by the intervention that led to long-term effects on hostile intent attributions and a desire for vengeance.

### *Did Attributions of Hostile Intent Mediate the Intervention's Effects on the Desire for Vengeance?*

Next, we sought to replicate Study 1's finding that the effect of implicit theories on attributions of hostile intent would mediate the relation with the desire for vengeance. As expected, attributions of intent were highly significantly correlated with higher ratings of a desire for vengeance,  $r = .41$ ,  $p < .001$ . In a formal test of mediation, there was a significant indirect effect of the intervention on ratings of vengeance mediated by attributions of intent  $b = -.03$  (95% CI [-.08, -.005]), and this amounted to roughly 32% of the effect of the intervention on vengeance. In this model, the intervention effect on vengeance was reduced to marginal significance,  $b = -.06$ ,  $t(76) = -1.931$ ,  $p = .058$ . Hence, a significant proportion of the relation between the incremental theory intervention and ratings of a desire for vengeance was accounted for by the reduction in attributions of hostile intent.

### *Replication of Effect on Open-Ended Comments*

Finally, we sought to replicate Study 2's effect on open-ended-positive and -negative reactions to the provocation scenario. This analysis showed that the experimental group students wrote more positive and fewer negative free responses than control students at the end of the year, ordered logistic  $b = -1.47$ ,  $SE = 0.51$ ,  $t(76) = -2.93$ ,  $p = .004$ . More concretely, the experimental group was 43% less likely to write negative, angry responses to the scenario 8 months postintervention (control = 59%; experimental = 34%; see Figure 2b), such as "I would be extremely angry" or "I would call him names." And the intervention group was more than 4 times more likely to write positive, understanding responses 8 months postintervention (control = 5%; experimental = 22%; see Figure 2b), such as "I would first expect them to say sorry and I myself would say sorry to them. I would not be that mad because they were in a hurry to get to class and we were in a hurry."

### *Summary*

In sum, Study 3 provided causal evidence for Study 1's correlational relation between implicit theories and ratings of a desire for vengeance, mediated by the hostile attributional bias. In addition,

Study 3 replicated the experimental findings in Study 2 over an entire school year and in doing so provided early evidence for a new method to supplement existing strategies to reduce attributions of hostile intent in real-world settings.

### General Discussion

As we noted at the outset, the hostile attributional bias (Dodge, 1980, 2006) has been one of the most generative and influential constructs in the study of reactive aggression, and its discovery has led to many successful interventions that reduce youth aggression in real-world settings (Wilson & Lipsey, 2006). Yet to more effectively prevent the development of this attributional bias—and to more effectively reduce it throughout the life span—we require a fuller picture of its origins. Although past research has shown that negative experiences (such as abuse or victimization) and hostile schemas (such as the chronic accessibility of aggressive thoughts) can predict the hostile attributional bias, we sought to expand the theoretical account of its development in adolescence to include beliefs that need not have arisen from negative or hostile influences. Specifically, we proposed that implicit theories of personality—underlying beliefs about whether personality traits are fixed or are malleable—could shape adolescents' patterns of attributions following ambiguous provocations and predict their aggressive reactions.

In a meta-analytic summary of 11 original correlational tests of the hypothesis, in an experiment replicated in two independent samples, and in a longitudinal intervention, we found support for this proposal. Study 1's tests showed that an entity theory of personality predicted attributions of hostile intent nearly as strongly as (and not significantly different from) the more hostile schemas and experiences that have been highlighted in summaries of past research (Dodge, 2006). An entity theory predicted hostile intent attributions even when controlling for adolescents' past experiences with peer victimization, suggesting that this link was not due to a shared relation with peer victimization. Study 1 also showed that implicit theories predicted hostile attributions equally for males and females, for students from communities with higher levels of violence and from communities with lower levels of violence, and even for students who reported a complete absence of hostile schemas about people in general.

In Study 2 we addressed the causal question more directly by showing that experimentally changing implicit theories toward more of an

incremental view could substantially reduce attributions of hostile intent during the same testing session. This was true both in a suburban neighborhood with low levels of violence and in an urban neighborhood with higher levels of violence. This suggests that even in generally nonhostile environments an incremental theory of personality can ameliorate an attributional precursor to aggressive behavior, and it shows that even in a context with higher levels of violence adolescents' hostile intent attributions could be changed. Study 3 extended this research by showing that an incremental theory, learned in only two class sessions at the beginning of the school year, could produce more benign intent attributions 8 months postintervention, and, because of this, greatly reduce the desire for revenge at the end of the school year. Note that Study 3 was only an initial test of this intervention in a high-income suburban school, and it will be important to continue to examine its effects on actual aggressive behavior among students from more hostile neighborhoods in the future (although see Yeager et al., in press). Nevertheless, as Figure 2b shows, even the relatively affluent sample in Study 3 showed substantial negative responses to the provocation scenario, suggesting that interventions may be beneficial for such samples.

Because entity beliefs about traits can increase the probability of aggressive responses to provocations, it is possible that even adolescents who live in largely nonhostile environments, but nevertheless learn that traits are fixed, can have a latent vulnerability that leads them to desire violent revenge toward peers who transgress against them. Of course, this desire for revenge may remain latent as long as adolescents successfully employ self-regulation skills to restrain their desires (DeWall & Anderson, 2011). Yet this hidden propensity for aggressive fantasies is alarming because it may go unaddressed. The present findings might lead us to think more broadly about the factors that contribute to youth aggression and about methods to prevent it, even in populations not typically thought of as "at risk" for hostile attributional biases.

We also hope our findings will lead the field to think carefully about the messages that we as a society convey to students about the fixedness of people's character traits. New research might examine what messages parents, media, teachers, and caregivers give to children and adolescents about the nature of traits, and might test whether changing those messages might promote greater resilience over time. For instance, in an effort to boost self-esteem, parents or teachers may use fixed labels to

comfort socially excluded teens (e.g., “don’t worry, you’re a good person”) and to describe aggressive peers (e.g., “they’re bad people” or “they’re evil”) so as to ward off victims’ self-blaming attributions. These messages, although designed to be helpful, may actually be fostering more vengeful responses to future social conflicts by leading students to live in an entity theory framework (cf. Mueller & Dweck, 1998; Rattan, Good, & Dweck, 2012, for analogous research in the academic domain). Hence, by directing future research toward understanding the seemingly nonhostile or purportedly positive messages from parents or educators, we may uncover unexpected psychological causes of aggression.

What might be some ways of comforting children and adolescents without implying that the aggressor is someone who is inherently bad and cannot change? Research has not conclusively answered this question. However, based on the present findings we propose that it could be helpful to emphasize that people do not bully others because they are unchangeably “bad,” but because of their thoughts and feelings—thoughts and feelings that live in the brain and that have the potential to change. For instance, a child could be reminded that some peers bully others because they are insecure about their social status and see bullying as a way to climb the social hierarchy. In fact, compelling research demonstrates that these insecurities produce bullying (Cohen & Prinstein, 2006; Faris & Felmler, 2011). More research is needed to arrive at effective methods for deflecting self-blaming attributions and acknowledging the injustice of peer aggression without incurring the cost of promoting an entity theory of traits.

Next, although we have focused on the pathway from implicit theories to aggression through hostile intent attributions, we do not believe that these hostile attributions are the only mediator between implicit theories and aggression. Our past research and theory suggests that those with more of an entity theory, in addition to the transgressor-focused attributions and emotions, also experience more negative self-focused emotions and attributions. For instance, in the face of social adversity, those with more of an entity theory are more likely to think that perhaps they are not likable people (Erdley et al., 1997), to say they felt bad about themselves (Yeager et al., 2011, Study 1), and to report feeling more ashamed (Yeager et al., 2011, Study 3). Yeager et al. (2011) showed that these negative feelings about the self, in turn, led to a

desire for vengeance, an effect that was independent of negative attributions about the transgressor. Thus, we view implicit theories as a broad construct that can have multiple pathways to aggression.

The present research may have implications for practice. Although past research has consistently shown that a hostile attributional bias can *predict* vengeful responses among older adolescents in middle school and high school (a conclusion supported by the present studies as well), past interventions designed to *change* this bias have been less consistently successful in reducing aggression in older age groups than among younger, elementary school-aged children (e.g., Metropolitan Area Child Study Research Group, 2002; for a meta-analysis, see Wilson & Lipsey, 2006). Future research can determine whether our implicit theories intervention can serve as a helpful addition to social-information processing interventions for adolescents.

#### *Limitations*

A notable limitation of this research, like much of the research on the hostile attributional bias, is that it relied on self-report measures of attributions and aggressive desires. Because these attributions are a subjective interpretation of an ambiguous stimulus, self-report is the preferred method to elicit them (e.g., Dodge, 1980, Study 2), and it often yields useful predictors of aggressive behavior (e.g., Pettit et al., 2010). Nevertheless, it will be important to continue to document the impact of implicit theories of personality on the behavioral consequences of these attributions. Interestingly, some research is beginning to find that changes in implicit theories can result in differences in behavioral aggression among low-income students attending an urban public high school (Yeager et al., in press). Hence, we think the results shown here for self-reported desire for vengeance are likely to be found in future analyses of behavioral aggression.

Next, the conclusions of the present research are limited to judgments of peers about whom an adolescent has no prior information. As noted, this was done to replicate the procedures used in classic studies (e.g., Dodge, 1980). However, attributions of intent can differ when considering peers about whom one already feels positively or negatively (Peets, Hodges, Kikas, & Salmivalli, 2007). Therefore, future studies might vary the participants’ prior knowledge of the provoking peer and assess differential effects of implicit theories.

### Conclusion

Social-cognitive theories of aggression have typically assumed that aggressive behavior comes from aggressive inputs. This assumption dates from the earliest social learning theories (Bandura, Ross, & Ross, 1961), which emphasized viewing and internalizing aggressive behavior, to the contemporary social information processing and general aggression models (Anderson & Bushman, 2001; Bushman & Huesmann, 2010; DeWall and Anderson, 2011; Dodge et al., 2006), which have highlighted aggressive inputs such as hostile environments, aggressive scripts, and violent media. As noted, we do not debate these findings, but we also believe it is critical to understand that aggressive behaviors come from aggressive interpretations of social stimuli. These interpretations can arise from a variety of inputs, perhaps even well-intentioned, seemingly helpful messages from valued caregivers who convey an entity theory of traits. When we shift our perspective to focus on overall mindsets that give rise to aggressive interpretations and on the messages that reinforce them, then we may be able to more fully map out the range of experiences and messages that can prepare adolescents to respond resiliently to life's social adversities.

### References

- Anderson, C. A., & Bushman, B. J. (2001). Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: A meta-analytic review of the scientific literature. *Psychological Science, 12*, 353–359.
- Aronson, E. (1999). The power of self-persuasion. *American Psychologist, 54*, 875–884.
- Aronson, J., Fried, C., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*, 113–125.
- Bandura, A., Ross, D., & Ross, A. (1961). Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology, 63*, 575–582.
- Baumeister, R. F., Vohs, K. D., DeWall, C. N., & Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review, 11*, 167–203.
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*, 246–263.
- Burks, S. V., Laird, R. D., Dodge, K. A., Pettit, G. S., & Bates, J. E. (1999). Knowledge structures, social information processing, and children's aggressive behavior. *Social Development, 8*, 220–236.
- Bushman, B. J., & Huesmann, L. R. (2010). Aggression. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 833–863). New York: Wiley.
- Chiu, C., Hong, Y., & Dweck, C. S. (1997). Lay dispositionism and implicit theories of personality. *Journal of Personality and Social Psychology, 73*, 19–30.
- Cohen, G. L., & Prinstein, M. J. (2006). Peer contagion of aggression and health-risk behavior among adolescent males: An experimental investigation of effects on public conduct and private attitudes. *Child Development, 77*, 967–983.
- DeWall, C. N., & Anderson, C. A. (2011). The general aggression model. In M. Mikulincer & P. R. Shaver (Eds.), *Understanding and reducing aggression, violence, and their consequences* (pp. 15–33). Washington, DC: American Psychological Association.
- Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development, 51*, 162–170.
- Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology, 18*, 791–814.
- Dodge, K. A., Bates, J. E., & Pettit, G. S. (1990). Mechanisms in the cycle of violence. *Science, 250*, 1678–1683.
- Dodge, K. A., Coie, J. D., & Lynam, D. (2006). Aggression and antisocial behavior in youth. In W. Damon, (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 719–788). New York: Wiley.
- Dodge, K. A., Laird, R., Lochman, J. E., & Zelli, A. (2002). Multidimensional latent-construct analysis of children's social information processing patterns: Correlations with aggressive behavior problems. *Psychological Assessment, 14*, 60–73.
- Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York: Random House.
- Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry, 6*, 267–285.
- Erdley, C., Cain, K., Loomis, C., Dumas-Hines, F., & Dweck, C. S. (1997). The relations among children's social goals, implicit personality theories and response to social failure. *Developmental Psychology, 33*, 263–272.
- Erdley, C. A., & Dweck, C. S. (1993). Children's implicit personality theories as predictors of their social judgments. *Child Development, 64*, 863–878.
- Faris, R., & Felmlee, D. (2011). Status struggles: Network centrality and gender segregation in same- and cross-gender aggression. *American Sociological Review, 76*, 48–73.
- Graham, S., & Juvonen, J. (1998). Self-blame and peer victimization in middle school: An attributional analysis. *Developmental Psychology, 34*, 587–599.
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57*, 243–259.
- Hong, Y., Chiu, C., Dweck, C. S., & Sacks, R. (1997). Implicit theories and evaluative processes in person cognition. *Journal of Experimental Social Psychology, 33*, 296–323.

- Hudley, C., & Graham, S. (1993). An attributional intervention to reduce peer-directed aggression among African-American boys. *Child Development, 64*, 124–138.
- Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15*, 309–334.
- Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review, 105*, 765–789.
- Levy, S. R., & Dweck, C. S. (1999). The impact of children's static versus dynamic conceptions of people on stereotype formation. *Child Development, 70*, 1163–1180.
- MacBrayer, E. K., Milich, R., & Hundley, M. (2003). Attributional biases in aggressive children and their mothers. *Journal of Abnormal Psychology, 112*, 698–708.
- Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology, 75*, 33–52.
- Peets, K., Hodges, E. V. E., Kikas, E., & Salmivalli, C. (2007). Hostile attributions and behavioral strategies in children: Does relationship type matter? *Developmental Psychology, 43*, 889–900.
- Pettit, G. S., Lansford, J. E., Malone, P. S., Dodge, K. A., & Bates, J. E. (2010). Domain specificity in relationship history, social-information processing, and violent behavior in early adulthood. *Journal of Personality and Social Psychology, 98*, 190–200.
- R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Rattan, A., Good, C., & Dweck, C. S. (2012). "It's ok—Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology, 48*, 731–737.
- Schooler, J. (2011). Unpublished results hide the decline effect. *Nature, 470*, 437.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology. *Psychological Science, 22*, 1359–1366.
- United States Department of Justice, Federal Bureau of Investigation. (2011). *Crime in the United States, 2010*. Retrieved January 22, 2012, from <http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/violent-crime/violent-crime>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48.
- Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science, 331*, 1447–1451.
- Williams, K. D. (2009). Ostracism: A temporal need-threat model. *Advances in Experimental Social Psychology, 41*, 275–314.
- Williams, K. S., Yeager, D. S., Cheung, C. K. T., & Choi, W. (2012). Cyberball (version 4.0) [Software]. Retrieved January 10, 2012, from <https://cyberball.wikispaces.com>
- Wilson, S. J., & Lipsey, M. W. (2006). *The effects of school-based social information processing interventions on aggressive behavior, Part I: Universal programs*. A Campbell Collaboration Systematic Review. Retrieved January 10, 2012, from [www.campbellcollaboration.org/lib/download/73/](http://www.campbellcollaboration.org/lib/download/73/)
- Yeager, D. S., & Krosnick, J. (2011). Does mentioning "some people" and "other people" in a survey question increase the accuracy of adolescents' self-reports? *Developmental Psychology, 47*, 1674–1679.
- Yeager, D. S., & Krosnick, J. (2012). Does mentioning "some people" and "other people" in an attitude question improve measurement quality? *Public Opinion Quarterly, 76*, 131–141.
- Yeager, D. S., Trzesniewski, K. H., & Dweck, C. S. (in press). An implicit theories of personality intervention reduces adolescent aggression in response to victimization and exclusion. *Child Development*.
- Yeager, D. S., Trzesniewski, K. H., Tirri, K., Nokelainen, P., & Dweck, C. S. (2011). Adolescents' implicit theories predict desire for vengeance after peer conflicts: Correlational and experimental evidence. *Developmental Psychology, 47*, 1090–1107.
- Yeager, D. S., & Walton, G. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research, 81*, 267–301.
- Yeung, R. S., & Leadbeater, B. J. (2007). Does hostile attributional bias for relational provocations mediate the short-term association between relational victimization and aggression in preadolescence? *Journal of Youth and Adolescence, 36*, 973.

### Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

**Table S1.** Summary of Correlational Relationships Between Implicit Theories, Attributions of Hostile Intent, and Aggressive Desires Across Eight Samples and Eleven Studies (Study 1).