Empirical Article

Clinical Psychological Science
1–21

DOI: 10.1177/2167702620966294
www.psychologicalscience.org/CPS

Physical Aggression Is Associated With More Effective Postdecisional Processing of Social Threat

Grace M. Brennan and Arielle Baskin-Sommers
Department of Psychology, Yale University

Abstract

Physically aggressive individuals are more likely to decide that others are threatening. Yet no research has examined how physically aggressive individuals' social decisions unfold in real time. Seventy-five incarcerated men completed a task in which they identified the emotions in faces displaying anger (i.e., threat) and happiness (i.e., nonthreat) at low, moderate, or high ambiguity. Participants then rated their confidence in their decisions either immediately or after a delay, and changes in confidence provided an index of postdecisional processing. Physical aggression was associated with stronger differentiation of threatening and nonthreatening faces under moderate ambiguity. Moreover, physical aggression was associated with steeper decreases in confidence over time following decisions that threatening faces were nonthreatening, indicating more extensive postdecisional processing. This pattern of postdecisional processing mediated the association between physical aggression and angry rumination. Findings suggest a role for postdecisional processing in the maintenance of threat-based social decisions in physical aggression.

Keywords

aggressive behavior, decision-making, facial expressions, social cognition, threat

Received 4/10/20; Revision accepted 8/16/20

Physical aggression, defined as behavior directed toward another person that results in physical harm or has the potential to cause physical harm, represents a transdiagnostic marker of social dysfunction. Engaging in physical aggression is associated with an elevated likelihood of mood, anxiety, personality, and substance use disorders (Okuda et al., 2015). Moreover, physical aggression is a hallmark symptom of several psychiatric diagnoses (e.g., antisocial personality disorder, borderline personality disorder, intermittent explosive disorder; American Psychiatric Association, 2013) and represents a primary feature of the externalizing spectrum of psychopathology (Krueger, Markon, Patrick, Benning, & Kramer, 2007). Elevated engagement in physical aggression has devastating intrapersonal and interpersonal consequences, including increasing the risk for criminal justice system involvement, damaging relationships, and promoting social rejection and isolation (Bierman & Wargo, 1995; Huesmann, Dubow, & Boxer, 2009; Poulin & Boivin, 1999).

Decades of research findings suggest that physical aggression is rooted in pervasive aberrations in social decision-making (Crick & Dodge, 1994). Social decision-making can be conceptualized as proceeding through different stages. First, at the formation stage, evidence is accumulated to inform an initial decision about a stimulus (e.g., whether someone poses a potential threat; Ratcliff & McKoon, 2008). Next, at the maintenance stage, which begins after an initial decision has been made, evidence about the stimulus continues to be accumulated. Depending on the incoming evidence, the initial decision may gain or lose strength, or it may be revised (Pleskac & Busemeyer, 2010). There is strong evidence that physical aggression is associated with aberrations in social decision-making that span these two stages.

Corresponding Author:
Grace M. Brennan, Department of Psychology, Yale University
E-mail: grace.m.brennan@yale.edu

At the formation stage, physically aggressive individuals are more likely to interpret social stimuli as threatening. They are more likely to identify ambiguous faces as angry (Brennan & Baskin-Sommers, 2020; Mellentin, Dervisevic, Stenager, Pilegaard, & Kirk, 2015; Schönenberg & Jusyte, 2014; Wilkowski & Robinson, 2012). In addition, they are more likely to interpret others' ambiguous actions as being carried out with hostile intent (De Castro, Veerman, Koops, Bosch, & Monshouwer, 2002; Dodge, 1980). Moreover, a careful examination of the research on the formation of social decisions in physical aggression highlights the role of ambiguity. The tendency among physically aggressive individuals to decide that others are threatening is amplified under more ambiguous conditions (Brennan & Baskin-Sommers, 2020; Dodge, 1980; Schönenberg & Jusyte, 2014; Wilkowski & Robinson, 2012; Zimmer-Gembeck & Nesdale, 2013). Taken together, the formation stage of social decision-making in physical aggression is characterized by a greater likelihood of deciding that others are threatening, particularly under greater ambiguity.

In contrast with the sizable body of research on the formation stage of social decision-making in physical aggression, very few studies have focused on the maintenance of social decisions in physical aggression. According to the existing evidence, it appears that once physically aggressive individuals form decisions that others are threatening, these decisions are more likely to persist over time. When deciding about others' traits, more physically aggressive individuals are more certain about their decisions that others are hostile, which suggests that these decisions are less flexible and are more likely to endure (Brennan & Baskin-Sommers, 2019). Moreover, physical aggression is robustly linked to angry rumination, a pattern of repetitive and unintentional thinking that persists after an anger-provoking experience (Anestis, Anestis, Selby, & Joiner, 2009; Bushman, 2002; Denson, 2013; Peled & Moretti, 2007; Sukhodolsky, Golub, & Cromwell, 2001; Wilkowski & Robinson, 2008). The content of angry rumination typically involves replaying the transgression, thinking about why it happened, and imagining revenge against the supposedly hostile transgressor (Sukhodolsky et al., 2001). Overall, physical aggression appears to be characterized by a lower likelihood of disengaging from decisions that others are threatening at the maintenance stage of social decision-making.

Although research suggests that physically aggressive individuals show aberrations at the maintenance stage of social decision-making, previous research has relied on self-report measures assessing the extent to which people engage in angry rumination in general. No research has examined directly how aggressive individuals' social decisions unfold in real time. The absence of research on this topic represents a major knowledge gap that may be hindering the improvement of clinical interventions. Interventions that focus on social decision-making in aggressive individuals generally aim to alter social decisions at the formation stage while neglecting the maintenance stage (e.g., AlMoghrabi, Huijding, & Franken, 2018; Penton-Voak et al., 2013). The focus of interventions on the formation stage of social decision-making may constrain their effectiveness given that aberrations at both stages increase risk for aggression (Dodge, 2006; McLaughlin, Aldao, Wisco, & Hilt, 2014). Important questions remain regarding the mechanisms through which physically aggressive individuals maintain social decisions over time.

Recent advances in the cognitive and decision sciences provide appealing possibilities for addressing these questions. One influential theory of decision-making (Pleskac & Busemeyer, 2010) suggests that when individuals make decisions, they engage in a process of evidence accumulation both before (i.e., predecisional processing) and after (i.e., postdecisional processing) the decision is made. This evidence accumulation process informs both their initial decision as well as confidence in that decision (Pleskac & Busemeyer, 2010). Crucially, the existence of this evidence accumulation process implies that confidence levels continue to shift even after decisions are made, and these shifts in confidence may bring about reversals of the initial decision (Murphy, Robertson, Harty, & O'Connell, 2015; van den Berg et al., 2016; Van Zandt & Maldonado-Molina, 2004).

In a key demonstration of this theory, Yu, Pleskac, and Zeigenfuse (2015) developed a double interrogation paradigm in which participants made perceptual decisions (e.g., whether the majority of dots in a cloud of moving dots were moving left or right). After the decision, participants rated their confidence in their decision either immediately (i.e., after a short interjudgment time [IJT]) or following a delay (i.e., after a long IJT). Across three studies, participants showed decreases in confidence from the short IJT to the long IJT. These decreases in confidence were driven by decreases in confidence for incongruent decisions, or decisions that were at odds with the evidence in the stimuli (e.g., responding left when the majority of dots were moving right). When participants made congruent decisions, in contrast, confidence levels remained relatively stable over time. Thus, declines in confidence from one time point to another reflected ongoing evidence accumulation that continued after the decision was made—that is, declines in confidence reflected postdecisional processing.

Postdecisional processing represents a mechanism that might help account for the aberrant maintenance

of social decisions in physical aggression. For example, more physically aggressive individuals might show more stable levels of confidence over time in decisions that others are threatening (e.g., identifying ambiguous faces as angry). Alternatively, they might show sharper decreases in confidence over time in social decisions that others are not threatening (e.g., identifying ambiguous faces as happy). Adopting an experimental approach to identifying whether physically aggressive individuals show aberrant patterns of postdecisional processing of threatening compared with nonthreatening social information could provide novel insights into how physically aggressive individuals maintain their beliefs that others are threatening.

Present Study and Hypotheses

To examine processes related to the formation and maintenance of social decisions in physical aggression, we developed a novel adaptation of the double interrogation paradigm. Our adaptation replaced the nonsocial stimuli from Yu and colleagues' (2015) paradigm (e.g., dots, lines) with social stimuli. The social stimuli were ambiguous emotional faces that displayed varying degrees of anger and happiness corresponding to low, moderate, or high ambiguity. Within each face, one emotion, either anger or happiness, was the dominant emotion. We used a sample of incarcerated adult male offenders with varying levels of physical aggressiveness. Because physical aggression is more pronounced in men than in women and more than half of state inmates in the United States are currently serving sentences for violent crimes (Bronson & Carson, 2019), incarcerated men represent an ideal population for studying physical aggression. Moreover, because the cognitive mechanisms influencing social-threat processing in physically aggressive individuals are shaped through repeated adverse experiences (e.g., violent victimization) over the course of development, these mechanisms are likely to be more strongly present in a sample of adults than in younger samples. The primary dependent variables derived from the experimental task were (a) emotion decisions, operationalized as the proportion of trials within each condition on which participants identified faces as angry (our measure related to social-decision formation), and (b) confidence in emotion decisions, operationalized as the average of all confidence ratings across trials within each condition. Changes in confidence over time (i.e., after the long IJT vs. the short IJT) served as an index of postdecisional processing of social information (our measure related to social-decision maintenance).

First, with regard to emotion decisions, we examined the association between physical aggression and anger identification and how this association varied as a function of facial characteristics (i.e., dominant emotion, ambiguity). We sought to provide a conceptual replication of previous work indicating that physical aggression was associated with greater sensitivity to subtle cues of social threat and more efficient processing of anger under heightened ambiguity (Brennan & Baskin-Sommers, 2020; Teige-Mocigemba, Hölzenbein, & Klauer, 2016; Wilkowski & Robinson, 2012). To this end, we hypothesized that physical aggression would be associated with a higher rate of anger identification, but only under greater ambiguity (Hypothesis 1).

Second, with regard to confidence in emotion decisions, we examined the association between physical aggression and confidence not only as a function of facial characteristics (i.e., dominant emotion, ambiguity) and time (i.e., IJT) but also as a function of which emotion decision (i.e., angry or happy) participants made. We were particularly interested in examining change in confidence over time as an index of postdecisional processing because this construct is most directly relevant to social-decision maintenance in physical aggression. For confidence as a function of facial characteristics, we hypothesized that physical aggression would be associated with less modulation of confidence as a function of ambiguity (Hypothesis 2) on the basis of previous research that suggested a failure of physically aggressive individuals to appropriately calibrate their confidence to match the level of ambiguity in the decision-making context (Brennan & Baskin-Sommers, 2019). For confidence as a function of emotion decisions, we hypothesized that physical aggression would be associated with heightened confidence in angry decisions (i.e., decisions that faces were angry; Hypothesis 3) on the basis of previous research that indicated heightened confidence in threat-based decisions among more physically aggressive individuals (Brennan & Baskin-Sommers, 2019).

For confidence as a function of time, we envisioned two main possibilities given the absence of previous research on this topic in physical aggression. On the one hand, physical aggression could be associated with smaller decreases in confidence over time for angry decisions even if the decisions are incongruent with the evidence displayed in the face (i.e., dominant emotion; Hypothesis 4a). This hypothesis is consistent with an inflexible style of postdecisional processing (i.e., confidence ratings change less over time, denoting reduced postdecisional evidence accumulation). On the other hand, physical aggression could be associated with larger decreases in confidence over time for happy decisions (Hypothesis 4b). This hypothesis is consistent with a pattern of more extensive postdecisional processing (i.e., confidence ratings change more over time,

denoting heightened postdecisional evidence accumulation). Essentially, both hypotheses represent different ways in which decisions that others are threatening (i.e., threat-based decisions) might be maintained over time.

Finally, we were interested in examining postdecisional processing as a potential mechanism involved in the maintenance of threat-based social decisions. That is, because angry rumination is an example of aberrant maintenance of threat-based social decisions in physical aggression, we wanted to know whether postdecisional processing helps to account for the link between physical aggression and angry rumination. We hypothesized that postdecisional processing on the task would mediate the association between physical aggression and angry rumination (Hypothesis 5).

Method

Participants

Participants were 78 men from a high-security correctional institution in Connecticut who ranged in age from 20 to 59 years ($M = 33.58$, $SD = 8.76$).[1] In terms of race, 65.4% of participants identified as Black, 32.1% identified as White, 1.3% identified as Asian, and 1.3% identified as multiracial. In terms of ethnicity, 16.7% of participants identified as Hispanic. In terms of educational attainment, 10.3% of the sample completed middle school or below, 47.4% completed some high school, 38.5% completed high school, and 3.8% completed some college. Almost all participants (97.4%) had been charged with a violent crime in their lifetime, and almost half (47.4%) had been charged with a violent institutional infraction while incarcerated (i.e., violations against persons, including fighting and assault on correctional staff). We used a prescreen of institutional files to exclude individuals who had documentation of a history of psychosis or bipolar disorder, current psychotropic medication, a family history of psychosis, certain medical problems that could impede comprehension of or performance on the task (e.g., uncorrectable auditory or visual deficits, three or more serious head injuries), IQ below 70, or reading level below fourth grade. These exclusion criteria were used primarily to reduce the influence of extraneous factors on task performance.

An a priori power analysis based on published studies on related topics (i.e., individual differences in facial-emotion identification and confidence in these decisions; Thome et al., 2016; Wilkowski & Robinson, 2012) indicated that a sample size of approximately 75 participants would be sufficient to detect small to medium effects with 80% power. To ensure sufficient power to account for the normative loss of data because of invalid task performance, we collected data from 78 participants.

Measures

Buss-Perry Aggression Questionnaire. The Buss-Perry Aggression Questionnaire (AQ; Buss & Perry, 1992) is a 29-item self-report measure of aggression. Participants rate each item on a 5-point Likert scale (1 = *extremely uncharacteristic of me*, 5 = *extremely characteristic of me*). The four widely used subscales of the questionnaire, established through factor analysis, are Physical Aggression (nine items), Verbal Aggression (five items), Anger (seven items), and Hostility (eight items). The AQ is a reliable, valid, and widely used measure of aggression (Harris, 1997; Tremblay & Ewart, 2005), with evidence for adequate reliability and validity in incarcerated samples (Archer & Haigh, 1997; Ireland & Archer, 2004). On the basis of previous research that demonstrated specificity of effects to physical aggression (e.g., Brennan & Baskin-Sommers, 2019, 2020; Wilkowski & Robinson, 2012), the hypotheses in the present study centered on physical aggression. Scores for the Physical Aggression subscale can range from 5 to 45; higher scores indicated individuals' greater endorsement that certain physically aggressive behaviors were characteristic of themselves. Unlike other aggression measures, which directly measure the frequency of aggressive behavior by prompting the individual to provide a count of aggressive behaviors within a specified time frame, the Physical Aggression subscale reflects an individual's self-characterization. The mean Physical Aggression score in the present sample ($M = 25.19$; see Table 1) was only slightly higher than that reported for male college students in Buss and Perry's (1992) original AQ validation study ($M = 24.3$). However, the mean Physical Aggression score in the present sample was comparable with mean scores reported in other studies that used samples of incarcerated male offenders (e.g., $M = 25.73$, Archer & Haigh, 1997; Sample 1: $M = 24.1$, Sample 2: $M = 24.4$, Ireland & Archer, 2004), and we observed a wider range of scores than studies that used college/community samples (e.g., Burt, Mikolajewski, & Larson, 2009). Internal consistency for the Physical Aggression subscale in the present sample (Cronbach's $\alpha = .82$) was good.

Anger Rumination Scale. The Anger Rumination Scale (ARS; Sukhodolsky et al., 2001) is a 19-item self-report measure of angry rumination. Participants rate each item on a 4-point Likert scale (1 = *almost never*, 4 = *almost always*). Total scores can range from 19 to 76; higher scores reflect higher levels of angry rumination. Internal consistency for the ARS in the present sample (Cronbach's $\alpha = .92$) was excellent.

Table 1. Sample Characteristics for Final Sample and Correlations Among Task Variables

Variable	1	2	3	4	5	6	7	8	9	*M*	*SD*	Range
1. Age	—									33.19	8.35	20–57
2. Race[a]	.11	—										
3. Ethnicity[b]	−.17	−.27*	—									
4. Educational attainment[c]	−.02	−.11	−.11	—								
5. AQ Physical Aggression	−.36*	.04	.13	.01	—					25.19	7.30	9–41
6. ARS total[d]	−.24*	−.13	.00	.03	.55*	—				37.46	10.97	20–68
7. Overall task accuracy	−.20	−.20	.16	.02	.27*	.18	—			74.18%	5.23%	57.50%–82.92%
8. Accuracy: low ambiguity	−.14	−.15	.05	−.02	.21	.12	.94*	—		86.93%	6.57%	63.75%–96.25%
9. Accuracy: moderate ambiguity	−.22	−.15	.11	.04	.36*	.23	.94*	.84*	—	76.58%	6.74%	55.63%–87.50%
10. Accuracy: high ambiguity	−.18	−.13	.24*	.02	.13	.14	.80*	.66*	.65*	59.02%	3.91%	48.75%–70.00%

Note: $N = 75$ (except as noted). Correlations for race, ethnicity, and educational attainment are reported as Spearman's ρ; all other correlations are reported as Pearson's *r*. AQ = Buss Perry Aggression Questionnaire; ARS = Anger Rumination Scale.

[a]There were 50 Black participants, 23 White participants, 1 Asian participant, and 1 multiracial participant. [b]There were 63 non-Hispanic participants and 12 Hispanic participants. [c]Eight participants had completed middle school or below, 35 had completed some high school, 29 had completed high school, and 3 had completed some college. [d]$N = 74$.

*$p < .05$.

Facial-emotion postdecisional-processing task. Participants completed a novel adaptation of the double interrogation paradigm developed by Yu and colleagues (2015). The task was a two-alternative, forced-choice task in which participants decided which of two emotions was displayed in a series of ambiguous emotional faces and then rated their confidence in their emotion decisions after one of two IJTs.

Stimuli. Stimuli consisted of emotional face images generated using the software package FaceGen Modeller Core (Version 3.18; Singular Inversions, Vancouver, Canada). This software uses a large database of scanned face images to generate avatars that appear realistic. Numerous studies on a range of topics, including physical aggression, have used these faces as stimuli and established that they are perceived similarly to images of posed facial expressions (Freeman & Ambady, 2009; Schulte-Rüther, Markowitsch, Fink, & Piefke, 2007; Todorov, Baron, & Oosterhof, 2008; Wilkowski & Robinson, 2012). Images of 40 unique male avatars of two racial backgrounds (Black and White) were used as stimuli. The racial composition of the face stimuli (i.e., 60% Black, 40% White) roughly mirrored that found in our sample. All participants viewed the same set of stimuli.

The intensity of various emotional expressions can be manipulated using the FaceGen Modeller software, allowing for the creation of faces displaying emotions from 0% intensity (i.e., fully ambiguous) to 100% intensity (i.e., nonambiguous). We manipulated the intensity of both anger and happiness simultaneously to generate faces displaying varying degrees of these two emotions. We chose anger and happiness because we wanted to examine the processing of social threat (i.e., anger) and nonthreat (i.e., happiness) in a manner most consistent with previous studies that investigated individual differences in social-threat perception (Maoz et al., 2016; Penton-Voak et al., 2013; Schönenberg & Jusyte, 2014; Thome et al., 2016; Wilkowski & Robinson, 2012). Through this process, stimuli representing three different ambiguity levels were created: 75% one emotion/25% other emotion (low ambiguity), 65% one emotion/35% other emotion (moderate ambiguity), and 55% one emotion/45% other emotion (high ambiguity). Within each ambiguity level, either anger or happiness served as the dominant emotion. Thus, within mostly angry faces, higher ambiguity corresponded to lower levels of anger and higher levels of happiness, and within mostly happy faces, higher ambiguity corresponded to lower levels of happiness and higher levels of anger. In total, six image types per avatar were created (three ambiguity levels for each of two dominant emotion types; see Fig. 1a). The process of generating six different image types for each avatar resulted in 240 unique images.

Task procedure. Participants were seated in front of a 27-in. high-performance LED gaming monitor (BenQ America, Costa Mesa, CA). Participants were told they would be playing a game that would involve making decisions about faces. Before starting, participants completed a three-part practice in which they practiced identifying the emotion displayed in a series of faces (10 trials), practiced using a rating bar (10 trials), and practiced playing the actual task (10 trials, with the possibility of an additional 10 trials of practice depending on performance; more details below). In the first and second parts of the practice, participants received accuracy feedback. In the third part of the practice, participants received timing feedback (i.e., about whether they made their response within the 1,500-ms limit). If participants did not respond quickly enough on at least 80% of responses in the third part of the practice, they completed an additional set of 10 practice trials to reinforce quick responding because timing is crucial for one of the key manipulations of the task (i.e., varying IJTs).

During the task, participants made two responses for each face they saw: First, they identified the emotion displayed in each face as quickly and accurately as possible (emotion-decision phase); second, they rated how confident they were about their emotion decision—for example, if participants identified a face as angry on a given trial, they would then rate how confident they were that the face was angry (confidence-rating phase). Participants identified the emotion displayed in the faces by moving the mouse left and right and then clicking to lock in their response. When participants moved the mouse to the left, the left response option (e.g., angry) was outlined in green. Conversely, when participants moved the mouse to the right, the right response option (e.g., happy) was outlined in green. When the response option of their choice was outlined in green, participants clicked to lock in that option as their response (see Fig. 1b). The emotion options (i.e., angry and happy) appeared on a predetermined, pseudorandomly ordered side of the screen on each trial; on half of the trials, "angry" appeared on the left side, and on half of the trials, "happy" appeared on the left side. After making their emotion decision, participants saw a blank screen for the duration of the IJT (either 50 or 1,500 ms; the selection of these IJTs followed the methodology of Yu et al., 2015). Finally, participants rated their confidence in their emotion decision using a rating bar, which ranged from 0% (*not at all confident*) to 100% (*extremely confident*), marked at intervals of 10%. Participants moved the mouse left and right to move a marker along the rating bar, then clicked to lock in their confidence rating at the location of the marker.

If participants took more than 1,500 ms to respond during either the emotion-decision phase or the

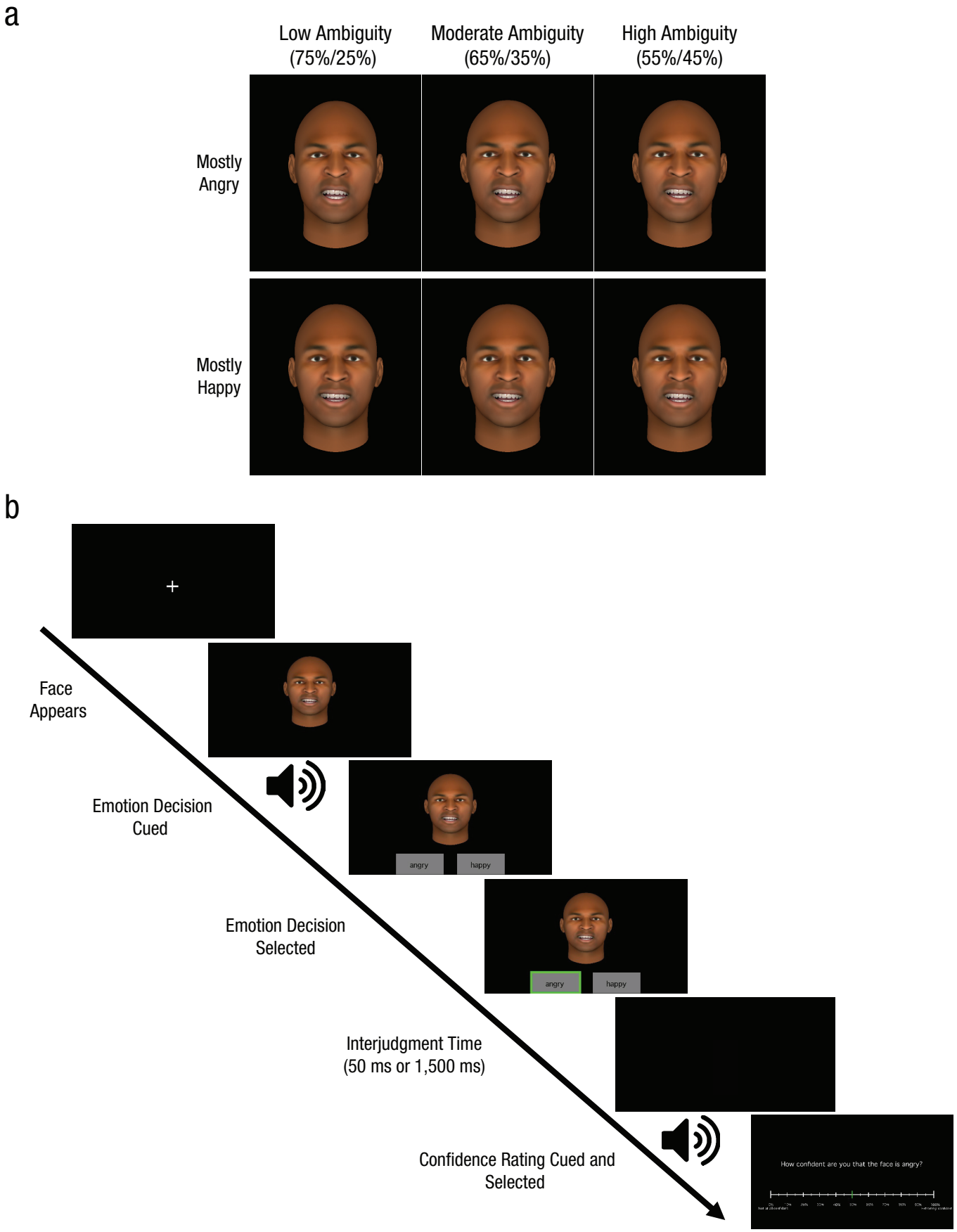

Fig. 1. *(continued on next page)*

Fig. 1. Sample task stimuli (a) and schematic representation of trial layout and timing in the facial-emotion postdecisional processing task (b). Stimuli displayed three different ambiguity levels (represented by the three columns). Within each ambiguity level, either anger (top row) or happiness (bottom row) was the dominant emotion. Each trial began with a fixation cross for 500 ms (b). Then a face image appeared, and after 800 ms, a tone sounded, cuing participants to provide an emotion decision for the face by using the mouse to select one of two response options (i.e., angry or happy). Then participants encountered an interjudgment time (IJT) of either 50 ms or 1,500 ms, during which a blank screen was displayed. Finally, after the IJT, a second tone sounded, cuing participants to rate their confidence in their emotion decision by moving the mouse to slide the green marker along the rating bar and clicking to lock in their rating.

confidence-rating phase, the words "too slow" appeared on the screen. Participants also were instructed that they would earn points for responding accurately and with sufficient speed. This procedure was designed to motivate participants to respond quickly given the importance of timing in this paradigm.

Stimulus presentation and response collection were controlled using the Psychtoolbox–3 extension (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) in MATLAB 2017b (The MathWorks, Natick, MA). Ordering of trials was pseudorandomized such that stimuli appeared in a random order, but the same avatar did not appear two trials in a row. First, each trial began with a fixation cross (500 ms), after which a face was displayed on the screen. After the face was on the screen for 800 ms, a tone sounded, and the response options appeared on the screen, prompting the participant to select one of the response options using the mouse. After participants identified the emotion in the face, there was an IJT of either 50 ms or 1,500 ms, after which a second tone sounded and the confidence-rating bar appeared on the screen, prompting participants to rate their confidence in their emotion decision (see Fig. 1b). The intertrial interval varied randomly between 1,000 ms and 2,000 ms (average 1,500 ms). The task consisted of 480 trials total, broken up into four separate blocks consisting of 120 trials each, allowing for short breaks in between each block. During the interblock breaks, participants were able to view the number of points they earned during the previous block (points were not visible to participants at any other time during the task).

General procedure

Before recruitment, study personnel received an institutional roster of inmates. Study personnel used this roster to review medical files for exclusion criteria (see Participants subsection above). This prescreening process was sanctioned by a Health Insurance Portability and Accountability Act (HIPAA) waiver and was designed to minimize the burden on individuals and the facility (i.e., to avoid moving people to the research room who would ultimately be excluded). Then, individuals were selected randomly from the list of eligible inmates and invited to participate. Invited individuals were provided with information about study procedures and informed that any information collected during the study would remain confidential and would not affect their legal status in any way. They were informed that they could withdraw from the study at any time. All participants provided written informed consent. In keeping with the Connecticut Department of Correction regulations, participants did not receive financial compensation. After providing consent, participants completed an initial session that involved a brief clinical overview, interview-based measures of personality traits and disorders particularly relevant to antisocial behavior (e.g., psychopathy, substance use disorders), and a series of neuropsychological assessments. Participants who did not meet eligibility thresholds on any of these assessments (see Participants subsection above) were excluded from further participation. Participants who screened positive for a current mood or anxiety disorder during the brief clinical overview were also excluded because of the potential for severe mood and anxiety symptoms to interfere with task performance. Eligible participants returned for a second session in which they completed the task and then completed the AQ and ARS. Both in-person sessions took place in a private testing space within the prison. The study protocol was approved by the Yale University Human Investigation Committee and was carried out per the provisions of the World Medical Association Declaration of Helsinki.

Data processing and analysis

Data quality control. Participants were excluded from analyses if their task data were invalid. Data were considered invalid if at least one of the following a priori criteria was met: (a) untimely responses (i.e., reaction times > 1,500 ms) on more than 20% of emotion decisions or confidence ratings, (b) emotion-decision accuracy at or below chance (i.e., ≤ 50%), (c) insensitivity to experimental manipulation of ambiguity (i.e., no differences in emotion decisions across levels of ambiguity), (d) no difference in observed IJT for the short IJT compared with the long IJT conditions, or (e) insufficient variability in confidence ratings across the entire task (i.e., limited to a range of 10% or less across all trials). Three participants were excluded from analyses because of these criteria (two for too many untimely responses and one for insufficient variability in confidence ratings). The final sample consisted of 75 participants (for sample characteristics

and correlations among task variables, see Table 1). Excluded participants did not differ from included participants in terms of physical aggression ($p = .729$).

Data analytic plan. Repeated measures general linear model (GLM) analysis was conducted to examine patterns of emotion decisions and confidence, separately, as a function of task manipulations and physical aggression. First, to examine patterns of emotion decisions and to provide a test of Hypothesis 1, we conducted a 2 (dominant emotion: anger, happiness) × 3 (ambiguity: low, moderate, high) repeated measures GLM, with AQ Physical Aggression (z-scored) as a continuous between-subjects independent variable, age as a covariate,[2] and the proportion of trials on which participants identified faces as angry (i.e., angry decisions) as a dependent variable. Follow-up repeated interaction contrasts were used to yield the following comparisons: low ambiguity compared with moderate ambiguity and moderate ambiguity compared with high ambiguity.

Second, to examine confidence as a function of facial characteristics and timing and to provide a test of Hypothesis 2, we conducted a 2 (dominant emotion: anger, happiness) × 3 (ambiguity: low, moderate, high) × 2 (IJT: short, long) repeated measures GLM, with AQ Physical Aggression (z-scored) as a continuous between-subjects independent variable and confidence as a dependent variable. Follow-up repeated interaction contrasts were used to yield the following comparisons: low ambiguity compared with moderate ambiguity and moderate ambiguity compared with high ambiguity.

Third, to examine confidence as a function of the variables listed in the preceding paragraph plus emotion decisions (i.e., whether faces were identified as angry or happy) and to provide a test of Hypotheses 3, 4a, and 4b, we initially planned to conduct a 2 (dominant emotion: anger, happiness) × 3 (ambiguity: low, moderate, high) × 2 (emotion decision: angry, happy) × 2 (IJT: short, long) repeated measures GLM, with AQ Physical Aggression (z-scored) as a continuous between-subjects independent variable and confidence as a dependent variable. However, we did not anticipate that a considerable number of participants would not exhibit variability in terms of emotion decisions within certain task conditions. More specifically, 25 out of 75 participants (33.3%) showed no emotion-decision variability (e.g., identified faces as happy on all trials) within at least one condition. For example, the condition under which the greatest number of participants exhibited no response variability was the mostly happy, low ambiguity, short IJT condition, in which 14 participants made congruent decisions (i.e., identified the faces as happy) on all trials. Participants with no response variability within at least one condition had no confidence values for one type of emotion decision (i.e., either angry or happy) within those conditions, creating a problem of empty cells that prevented participants with missing confidence values from being included in an analysis involving both emotion decision and all of the other independent variables.

Conducting the analysis without the participants who had empty cells was undesirable for two reasons: First, we would be excluding participants in a nonrandom fashion because participants with empty cells had superior task performance in at least one task condition; second, excluding such a large number of participants would significantly reduce our power to detect hypothesized effects. Therefore, to examine confidence as a function of emotion decision, the alternative was to collapse across one of the other task conditions. We considered collapsing across IJT, dominant emotion, or ambiguity. Following the approach of avoiding empty cells so that we could analyze all participants' data, we could not collapse across IJT because doing so would still result in participants with empty cells. Collapsing across dominant emotion would avoid empty cells; however, from a logical standpoint, it made little sense to collapse across dominant emotion because then we would lose all context for knowing whether emotion decisions (i.e., angry/happy) were congruent or incongruent. A primary reason for examining confidence in angry decisions compared with happy decisions was to have the ability to characterize confidence in congruent emotion decisions compared with incongruent emotion decisions. On the basis of these considerations, we chose to collapse across ambiguity, which allowed us to analyze all participants' data and examine confidence as a function of emotion decisions that were either congruent or incongruent with the dominant emotion displayed in the faces. Therefore, our revised model was a 2 (dominant emotion: anger, happiness) × 2 (IJT: short, long) × 2 (emotion decision: angry, happy) repeated measures GLM.

Finally, to examine potential mechanisms supporting the link between physical aggression and angry rumination and to provide a test of Hypothesis 5, a mediation analysis was conducted, with AQ Physical Aggression as the independent variable, ARS total score as the dependent variable, and postdecisional processing as the mediator. The analysis was performed using the PROCESS macro Model 4 (Hayes, 2018) for IBM SPSS (Version 22). We used a nonparametric resampling procedure (bootstrapping) with 10,000 samples to estimate the indirect effect.

Task validation. We relied in part on previous research to inform our manipulation checks, particularly the task effects demonstrated using the original double-interrogation

paradigm (Yu et al., 2015). With regard to emotion decisions, we expected that angry decisions (i.e., the proportion of trials on which participants identified faces as angry) would be higher for mostly angry faces compared with mostly happy faces (i.e., a main effect of dominant emotion on emotion decisions; Manipulation Check 1). Furthermore, we expected that angry decisions would decrease as ambiguity increased for mostly angry faces, tracking the decreasing level of anger in these faces; conversely, we expected that angry decisions would increase as ambiguity increased for mostly happy faces, tracking the increasing level of anger in these faces (i.e., a Dominant Emotion × Ambiguity interaction in the analysis of emotion decisions; Manipulation Check 2).

With regard to confidence, we expected that confidence would decrease as ambiguity increased (i.e., a main effect of ambiguity on confidence; Manipulation Check 3). Furthermore, we expected that confidence would be lower after the long IJT compared with the short IJT (i.e., a main effect of IJT on confidence; Manipulation Check 4). We also expected that confidence would be lower for incongruent decisions compared with congruent decisions (i.e., a Dominant Emotion × Emotion Decision interaction; Manipulation Check 5). Finally, we expected that confidence would be lower after the long IJT compared with the short IJT but that this effect would depend on dominant emotion as well as emotion decision. More specifically, we expected that confidence would be lower after the long IJT, but only when participants made incongruent decisions—that is, identified mostly happy faces as angry or identified mostly angry faces as happy (i.e., a Dominant Emotion × IJT × Emotion Decision interaction; Manipulation Check 6).

Results

Emotion decisions

The repeated measures GLM involving emotion decisions revealed both task effects and effects related to physical aggression. In terms of task effects, we detected a main effect of dominant emotion on emotion decisions, $F(1, 73) = 1{,}657.01$, $p < .001$, $\eta_p^2 = .96$, 90% CI = [.94, .97],[3] such that mostly angry faces were more likely to be identified as angry ($M = 64.3\%$, 95% CI = [61.0%, 67.6%]) compared with mostly happy faces ($M = 16.0\%$, 95% CI = [13.8%, 18.1%]). This main effect provides a key demonstration of task validity by indicating that participants were able to differentiate between the two types of faces and identify the dominant emotion (i.e., Manipulation Check 1 was successful).

We also detected a main effect of ambiguity on emotion decisions, $F(2, 146) = 79.20$, $p < .001$, $\eta_p^2 = .52$, 90% CI = [.42, .59]. Examination of the repeated contrasts indicated that both the contrast between low ambiguity and moderate ambiguity, $F(1, 73) = 82.11$, $p < .001$, $\eta_p^2 = .53$, 90% CI = [.39, .62], and the contrast between moderate ambiguity and high ambiguity, $F(1, 73) = 47.14$, $p < .001$, $\eta_p^2 = .39$, 90% CI = [.25, .50], were significant. Examination of the means indicated that as ambiguity increased, faces were less likely to be identified as angry (low ambiguity: $M = 44.9\%$, 95% CI = [43.0%, 46.8%]; moderate ambiguity: $M = 39.9\%$, 95% CI = [37.2%, 42.5%]; high ambiguity: $M = 35.6\%$, 95% CI = [32.4%, 38.9%]). This finding that more ambiguous faces were more likely to be identified as happy may reflect the fact that happiness is the most easily recognized facial emotion (Sauter, 2010), and therefore happiness cues may have had a greater impact on participants' emotion decisions when relative levels of anger and happiness were more equivalent (i.e., at higher levels of ambiguity).

Finally, we detected a Dominant Emotion × Ambiguity interaction, $F(2, 146) = 1{,}453.74$, $p < .001$, $\eta_p^2 = .95$, 90% CI = [.94, .96], qualifying the main effect of ambiguity reported above. More specifically, within mostly angry faces, faces were less likely to be identified as angry as ambiguity increased (low ambiguity: $M = 81.9\%$, 95% CI = [78.9%, 84.8%]; moderate ambiguity: $M = 66.5\%$, 95% CI = [62.7%, 70.2%]; high ambiguity: $M = 44.6\%$, 95% CI = [41.0%, 48.3%]). Within mostly happy faces, however, faces were more likely to be identified as angry as ambiguity increased (low ambiguity: $M = 8.0\%$, 95% CI = [6.4%, 9.6%]; moderate ambiguity: $M = 13.3\%$, 95% CI = [11.2%, 15.4%]; high ambiguity: $M = 26.6\%$, 95% CI = [23.5%, 29.7%]). This interaction provides further evidence of task validity by indicating that participants' ability to differentiate between mostly angry faces and mostly happy faces decreased as ambiguity increased (i.e., Manipulation Check 2 was successful).

In terms of effects related to physical aggression-, we detected a Dominant Emotion × Ambiguity × Physical Aggression interaction, $F(2, 146) = 5.42$, $p = .007$, $\eta_p^2 = .07$, 90% CI = [.01, .14]. Examination of the interaction contrasts indicated that the difference in the proportion of angry decisions for mostly angry faces compared with mostly happy faces varied as a function of ambiguity level and physical aggression. More specifically, the contrast between low ambiguity and moderate ambiguity for the difference between mostly angry faces and mostly happy faces as a function of physical aggression was significant, $F(1, 73) = 4.13$, $p = .046$, $\eta_p^2 = .05$, 90% CI = [.001, .15]. The contrast between moderate ambiguity and high ambiguity for the difference between mostly angry faces and mostly happy faces as a function of physical aggression was significant as well, $F(1, 73) = 9.39$, $p = .003$, $\eta_p^2 = .11$, 90% CI = [.02, .23].

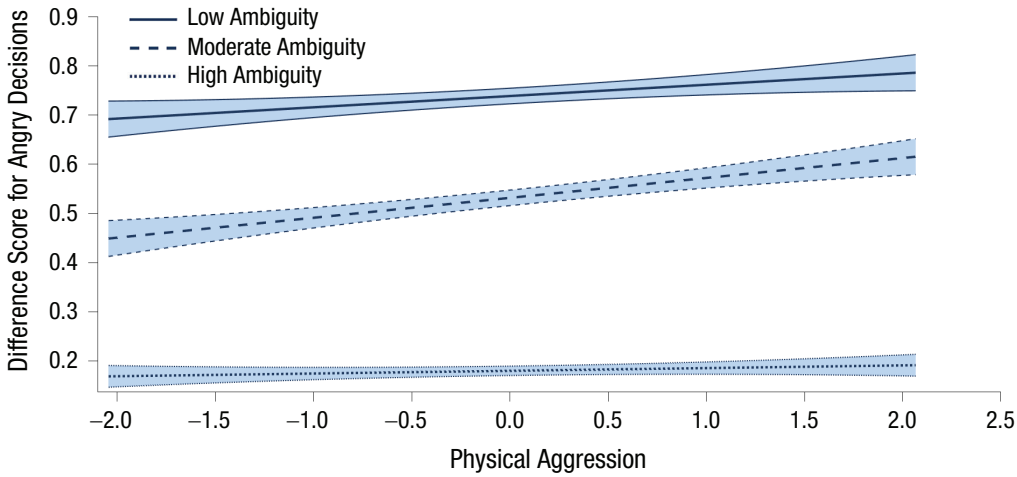

Fig. 2. The relationship between physical aggression (*z*-scored) and difference score for the proportion of angry decisions at low, moderate, and high ambiguity. Difference scores were calculated by subtracting the proportion of angry decisions for mostly happy faces from the proportion of angry decisions for mostly angry faces and thus represent how well participants were able to differentiate mostly angry faces from mostly happy faces (higher difference scores represent better differentiation). Participants with higher levels of physical aggression were better at differentiating mostly angry faces from mostly happy faces, but only at moderate ambiguity. Error bands represent ±1 *SE*.

To represent and interpret this interaction, we calculated a difference score by subtracting the proportion of angry decisions for mostly happy faces from the proportion of angry decisions for mostly angry faces. Thus, higher scores represent better differentiation between mostly angry faces and mostly happy faces. We detected a significant effect of physical aggression on the difference score for moderate ambiguity, b = 0.04, SE = 0.02, p = .009, η_p^2 = .09, 90% CI = [.01, .20], but not for low ambiguity, b = 0.02, SE = 0.02, p = .137, η_p^2 = .03, 90% CI = [.00, .12], or high ambiguity, b = 0.01, SE = 0.01, p = .551, η_p^2 = .01, 90% CI = [.00, .06] (see Fig. 2). Thus, largely consistent with Hypothesis 1, higher levels of physical aggression were associated with better differentiation between mostly angry faces and mostly happy faces under moderate ambiguity.

Confidence

Confidence as a function of facial characteristics and IJT. The analysis revealed both task effects and effects related to physical aggression. In terms of task effects, we detected a main effect of dominant emotion on confidence, F(1, 73) = 39.84, p < .001, η_p^2 = .35, 90% CI = [.21, .47], such that participants were more confident in their decisions about mostly happy faces (M = 79.7, 95% CI = [76.9, 82.4]) compared with mostly angry faces (M = 75.9, 95% CI = [73.1, 78.7]). Furthermore, we detected a main effect of ambiguity on confidence, F(2, 146) = 103.79, p < .001, η_p^2 = .59, 90% CI = [.50, .65]. Examination of the repeated contrasts indicated that both the contrast between low ambiguity and moderate ambiguity, F(1, 73) = 124.00, p < .001, η_p^2 = .63, 90% CI = [.51, .70], and the contrast between moderate ambiguity and high ambiguity, F(1, 73) = 43.65, p < .001, η_p^2 = .37, 90% CI = [.23, .49], were significant. Examination of the means indicated that as ambiguity increased, confidence decreased (low ambiguity: M = 80.7, 95% CI = [78.2, 83.2]; moderate ambiguity: M = 77.3, 95% CI = [74.5, 80.1]; high ambiguity: M = 75.3, 95% CI = [72.4, 78.2]). This main effect provides a key demonstration of task validity in general and the success of the ambiguity manipulation in particular because it indicates that participants showed the expected pattern of lower confidence under greater ambiguity (i.e., Manipulation Check 3 was successful). We also detected a main effect of IJT on confidence, F(1, 73) = 54.71, p < .001, η_p^2 = .43, 90% CI = [.28, .54], such that participants were less confident in their decisions after the long IJT (M = 77.1, 95% CI = [74.4, 79.8]) compared with the short IJT (M = 78.5, 95% CI = [75.7, 81.2]). This finding is consistent with previous research that indicated that confidence tends to decrease over time (Yu et al., 2015) and provided evidence for task validity (i.e., Manipulation Check 4 was successful).

In addition, we detected several interactions. First, we detected a Dominant Emotion × Ambiguity interaction, F(2, 146) = 20.79, p < .001, η_p^2 = .22, 90% CI = [.12, .31]. Second, we detected an Ambiguity × IJT interaction, F(2, 146) = 5.86, p = .004, η_p^2 = .07, 90% CI = [.02,

.14]. Finally, we detected a Dominant Emotion × Ambiguity × IJT interaction, $F(2, 146) = 3.61$, $p = .032$, $\eta_p^2 = .05$, 90% CI = [.003, .11], which qualified the two-way interactions reported above. Examination of the repeated interaction contrasts indicated that the moderate-ambiguity difference in confidence compared with the high-ambiguity difference in confidence from the short IJT to the long IJT varied as a function of dominant emotion, $F(1, 73) = 8.55$, $p = .005$, $\eta_p^2 = .11$, 90% CI = [.02, .22], whereas the low-ambiguity difference in confidence compared with the moderate-ambiguity difference in confidence from short IJT to long IJT did not, $F(1, 73) = 0.23$, $p = .636$, $\eta_p^2 = .003$, 90% CI = [.00, .05]. Examination of the means indicated that the decrease in confidence over time (i.e., from short IJT to long IJT) for high-ambiguity faces compared with moderate-ambiguity faces was greater for mostly happy faces (moderate ambiguity, short IJT: $M = 80.3$, 95% CI = [77.4, 83.2]; moderate ambiguity, long IJT: $M = 79.4$, 95% CI = [76.5, 82.3]; high ambiguity, short IJT: $M = 77.3$, 95% CI = [74.5, 80.2]; high ambiguity, long IJT: $M = 74.8$, 95% CI = [71.8, 77.8]) compared with mostly angry faces (moderate ambiguity, short IJT: $M = 75.8$, 95% CI = [72.8, 78.7]; moderate ambiguity, long IJT: $M = 73.8$, 95% CI = [70.9, 76.7]; high ambiguity, short IJT: $M = 75.3$, 95% CI = [72.3, 78.3]; high ambiguity, long IJT: $M = 73.7$, 95% CI = [70.7, 76.6]).

In terms of effects related to physical aggression, we detected a Dominant Emotion × IJT × Physical Aggression interaction, $F(1, 73) = 8.68$, $p = .004$, $\eta_p^2 = .11$, 90% CI = [.02, .22]. To decompose this interaction, we examined the IJT × Physical Aggression interaction within each of the two dominant emotions (mostly angry and mostly happy). Whereas the IJT × Physical Aggression interaction was significant for mostly angry faces, $F(1, 73) = 10.67$, $p = .002$, $\eta_p^2 = .13$, 90% CI = [.03, .25], it was not significant for mostly happy faces, $F(1, 73) = 0.17$, $p = .680$, $\eta_p^2 = .002$, 90% CI = [.00, .05] (see Fig. 3). Thus, higher levels of physical aggression were associated with steeper decreases in confidence over time for mostly angry faces but not mostly happy faces. This three-way interaction remained significant after controlling for overall task accuracy. Because we did not detect an Ambiguity × Physical Aggression interaction, we did not find support for Hypothesis 2.

Confidence as a function of facial characteristics, IJT, and emotion decisions. The analysis revealed both task effects and effects related to physical aggression. Task effects already reported above will not be repeated. The only additional effects were those involving emotion decisions.

In terms of task effects involving emotion decisions, we detected a main effect of emotion decision on confidence, $F(1, 73) = 29.93$, $p < .001$, $\eta_p^2 = .29$, 90% CI = [.15, .41], such that participants were more confident when they identified faces as happy ($M = 76.6$, 95% CI = [73.7, 79.5]) compared with when they identified faces as angry ($M = 71.6$, 95% CI = [68.7, 74.6]). This finding mirrors the main effect of dominant emotion on confidence reported above (i.e., that participants were more confident in their decisions about mostly happy faces compared with mostly angry faces). Furthermore, this finding is consistent with previous research indicating that people are more confident when they make benign compared with hostile judgments of others (Brennan & Baskin-Sommers, 2019; Rand, Ohtsuki, & Nowak, 2009; Siegel, Mathys, Rutledge, & Crockett, 2018).

We also detected a Dominant Emotion × Emotion Decision interaction, $F(1, 73) = 228.02$, $p < .001$, $\eta_p^2 = .76$, 90% CI = [.67, .81]. Examination of the means indicated that participants were more confident when they identified mostly angry faces as angry ($M = 77.2$, 95% CI = [74.4, 80.0]) compared with happy ($M = 71.6$, 95% CI = [68.4, 74.8]); conversely, participants were more confident when they identified mostly happy faces as happy ($M = 81.7$, 95% CI = [79.0, 84.4]) compared with angry ($M = 66.0$, 95% CI = [62.6, 69.4]). This finding indicates that participants adjusted their confidence appropriately to the congruence, or lack thereof, between their decision and the dominant emotion displayed in the face (and thus provided valid confidence ratings; i.e., Manipulation Check 5 was successful).

Finally, we detected a Dominant Emotion × IJT × Emotion Decision interaction, $F(1, 73) = 5.53$, $p = .021$, $\eta_p^2 = .07$, 90% CI = [.01, .18]. Examination of the means indicated that the Dominant Emotion × Emotion Decision interaction reported in the preceding paragraph was qualified by IJT in the following way: Participants were more confident when they made decisions congruent with the dominant emotion displayed in the face, particularly after the long IJT (compared with the short IJT). More specifically, the difference in confidence between the congruent decision (i.e., angry) and incongruent decision (i.e., happy) for mostly angry faces was larger after the long IJT (angry decision: $M = 76.5$, 95% CI = [73.6, 79.3]; happy decision: $M = 70.5$, 95% CI = [67.3, 73.6]) compared with the short IJT (angry decision: $M = 77.9$, 95% CI = [75.2, 80.7]; happy decision: $M = 72.7$, 95% CI = [69.3, 76.0]), and the same pattern was seen for mostly happy faces (happy decision, long IJT: $M = 81.3$, 95% CI = [78.5, 84.0]; angry decision, long IJT: $M = 64.6$, 95% CI = [61.2, 68.1]; happy decision, short IJT: $M = 82.1$, 95% CI = [79.4, 84.8]; angry decision, short IJT: $M = 67.4$, 95% CI = [63.8, 71.0]). This finding shows that resolution (i.e., the difference in confidence for the congruent decision vs. the incongruent decision) increased from the short IJT to the long

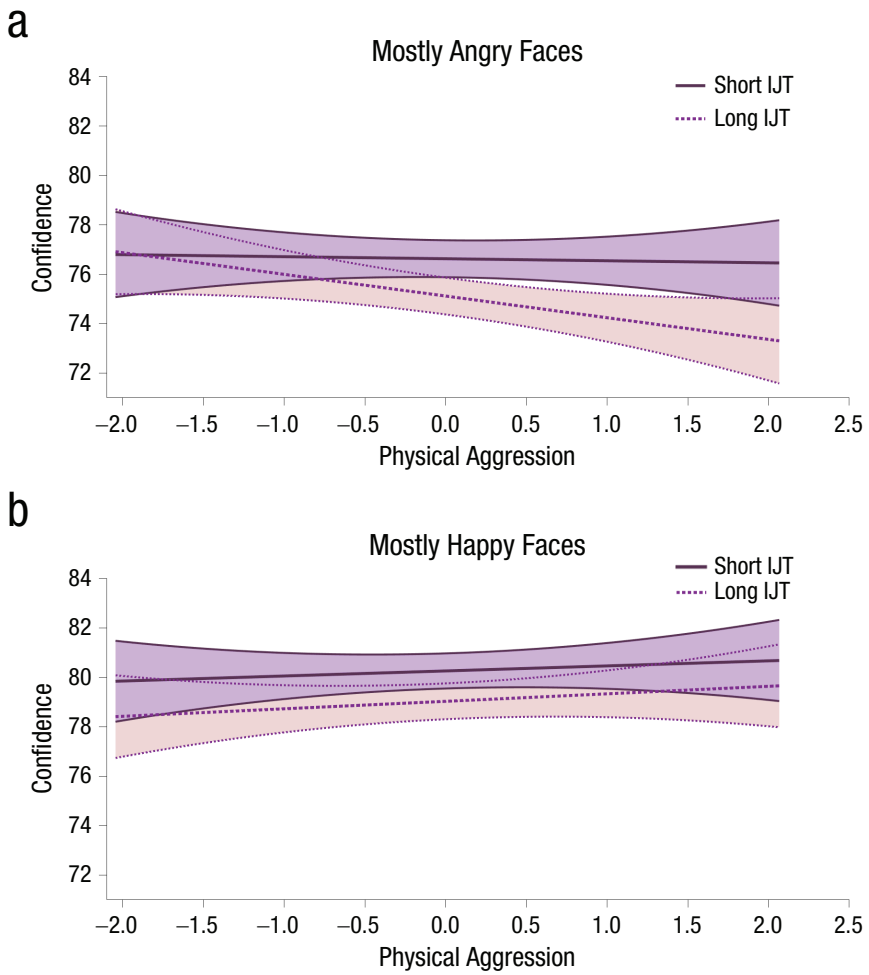

Fig. 3. The relationship between physical aggression (*z*-scored) and confidence as a function of interjudgment time (IJT) for mostly angry faces (a) and mostly happy faces (b). Participants with higher levels of physical aggression showed steeper decreases in confidence from the short IJT to the long IJT for mostly angry faces but not for mostly happy faces. Error bands represent ±1 *SE*.

IJT and provides a key demonstration that the IJT manipulation was successful. In other words, consistent with previous research (Yu et al., 2015), when participants had more time to process their emotion decision and continue to accumulate evidence, they were better able to adjust their confidence according to whether the decision was congruent with the dominant emotion in the face (i.e., Manipulation Check 6 was successful).

In terms of effects related to physical aggression, we detected an IJT × Physical Aggression interaction, $F(1, 73) = 4.90$, $p = .030$, $\eta_p^2 = .06$, 90% CI = [.003, .17]. This interaction was qualified by a Dominant Emotion × IJT × Emotion Decision × Physical Aggression interaction, $F(1, 73) = 4.99$, $p = .029$, $\eta_p^2 = .06$, 90% CI = [.004, .17]. We examined the IJT × Physical Aggression interaction within each of the conditions (dominant emotion angry, decision angry; dominant emotion angry, decision happy; dominant emotion happy, decision angry; dominant emotion happy, decision happy). The IJT × Physical Aggression interaction was not significant for mostly angry faces, angry decision, $F(1, 73) = 0.08$, $p = .779$, $\eta_p^2 = .001$, 90% CI = [.00, .04]; mostly happy faces, angry decision, $F(1, 73) = 1.16$, $p = .284$, $\eta_p^2 = .02$, 90% CI = [.00, .09]; or mostly happy faces, happy decision, $F(1, 73) = 0.01$, $p = .906$, $\eta_p^2 = .00$, 90% CI = [.00, .01]. However, the IJT × Physical Aggression interaction was significant for mostly angry faces, happy decision, $F(1, 73) = 10.71$, $p = .002$, $\eta_p^2 = .13$, 90% CI = [.03, .25], which suggests that this two-way interaction was driving the four-way interaction (see Fig. 4). This finding suggests that more physically aggressive individuals showed steeper decreases in confidence from the short IJT to the long IJT when they made incongruent decisions about mostly angry faces (i.e., when they misidentified mostly angry faces as happy).

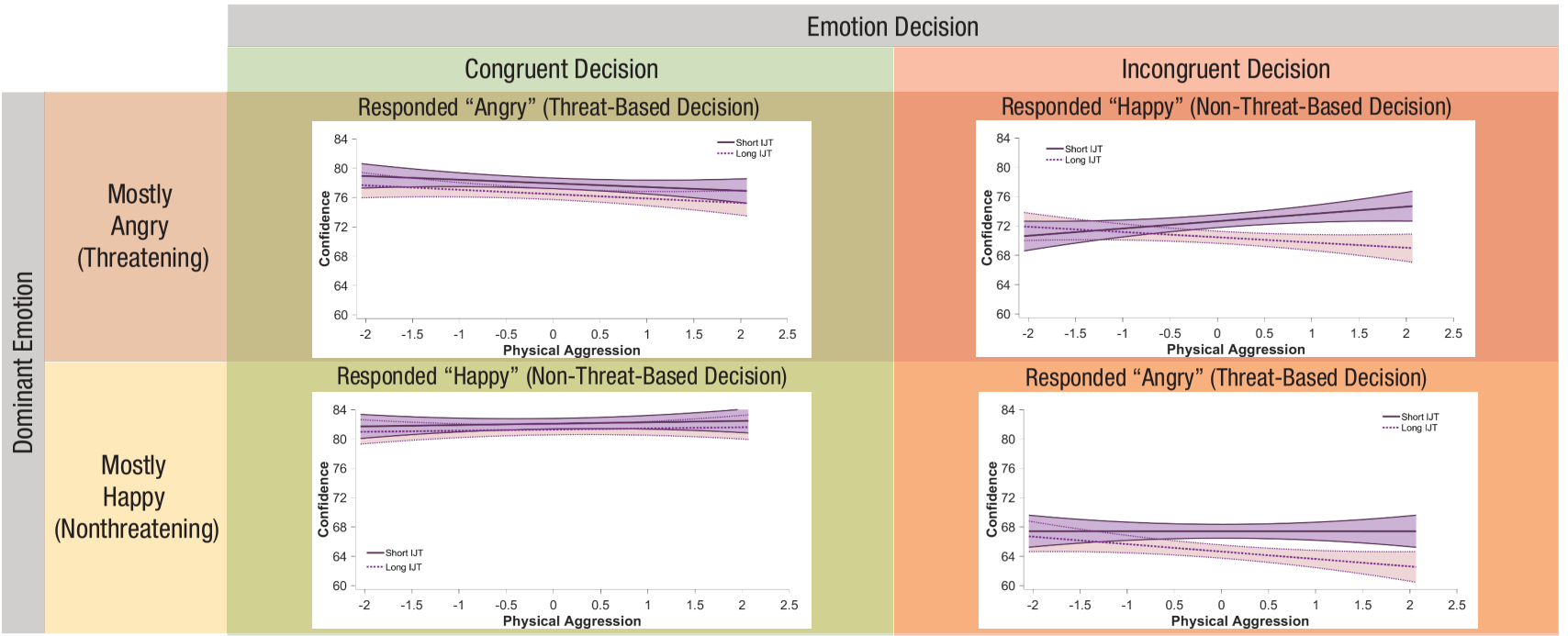

Fig. 4. The relationship between physical aggression (*z*-scored) and confidence as a function of interjudgment time (IJT) for faces that were mostly angry compared with mostly happy and for emotion decisions that were congruent (i.e., matched the dominant emotion displayed in the face) compared with incongruent (i.e., did not match the dominant emotion displayed in the face). Participants with higher levels of physical aggression showed steeper decreases in confidence from the short IJT to the long IJT for incongruent decisions about mostly angry faces (top right), but not for any of the other three dominant emotion/decision combinations. Error bands represent ±1 *SE*.

This pattern of findings is consistent with Hypothesis 4b (i.e., that more physically aggressive individuals would show larger decreases in confidence over time for happy decisions), but we failed to find support for Hypothesis 4a (i.e., that more physically aggressive individuals would show smaller decreases in confidence over time for angry decisions). This finding also lends nuance to the Dominant Emotion × IJT × Physical Aggression interaction detected in the previous model, which suggested that more physically aggressive individuals showed steeper decreases in confidence over time for mostly angry faces. The four-way interaction suggests that physical aggression was not related to steeper decreases in confidence over time for angry faces in general; rather, the effect was moderated by emotion decision such that steeper decreases in confidence were found only when angry faces were misidentified as happy. Because we did not detect an Emotion Decision × Physical Aggression interaction, we did not find support for Hypothesis 3. All effects related to physical aggression reported within this subsection remained significant after controlling for overall task accuracy.

Supplemental analysis related to slow responding

Given that differences in rates of slow responding could lead to differences in emotion decisions or confidence ratings, we wanted to rule out slow responding as a potential confounding variable in the relationship between physical aggression and dependent variables derived from the task. Therefore, we needed to establish that physical aggression was not associated with rates of slow responding. Correlation analyses indicated that physical aggression was not significantly associated with slow responding (i.e., the number of trials with reaction times > 1,500 ms) for emotion decisions, $r(73) = -.06$, $p = .604$, or confidence ratings, $r(73) = -.10$, $p = .405$. For additional analyses examining the robustness of results, see Supplemental Results and Table S1 in the Supplemental Material available online.

Linking physical aggression to angry rumination via postdecisional processing

For the mediation analysis, we adopted a data-driven approach to derive the postdecisional-processing variable. In the above analysis of confidence, we found that more physically aggressive individuals showed steeper decreases in confidence over time when they misidentified mostly angry faces as happy. We computed a difference score to represent this significant two-way interaction by subtracting confidence in happy decisions for mostly angry faces after the long IJT from confidence in happy decisions for mostly angry faces after the short IJT. This difference score thus reflected the extent of postdecisional processing after incongruent decisions about mostly angry faces.

In the mediation model,[4] the association between physical aggression and the postdecisional-processing difference score (i.e., the *a* path) was significant, $b = 0.25$, $SE = 0.08$, $p = .002$, 95% CI = [0.10, 0.41]. The association between the postdecisional-processing difference score and angry rumination (i.e., the *b* path) was also significant, $b = 0.53$, $SE = 0.24$, $p = .028$, 95% CI = [0.06, 1.01]. In addition, the association between physical aggression and angry rumination (i.e., the *c* path, or total effect) was significant, $b = 0.80$, $SE = 0.16$, $p < .001$, 95% CI = [0.48, 1.11]. Furthermore, after controlling for the mediator (postdecisional-processing difference score), the association between physical aggression and angry rumination (i.e., the c' path, or direct effect) remained significant, $b = 0.66$, $SE = 0.17$, $p < .001$, 95% CI = [0.33, 0.99]. Finally, the analysis indicated a significant indirect effect of physical aggression on angry rumination through postdecisional processing of angry faces, $b = 0.13$, $SE = 0.07$, 95% CI = [0.03, 0.30]. Thus, consistent with Hypothesis 5, postdecisional processing mediated the association between physical aggression and angry rumination.

To examine the specificity of the indirect effect via postdecisional processing in the context of incongruent decisions about mostly angry faces (the element of postdecisional processing we found was associated with aggression in earlier analyses; see Fig. 4), we used PROCESS Model 6 to test indirect effects via multiple mediators. In addition to postdecisional processing in the context of incongruent decisions about mostly angry faces, we entered difference scores representing the three other interaction contrasts (see Fig. 4) as potential mediators as well. None of the indirect effects for the other difference scores were significant, which suggests that postdecisional processing in the context of incongruent decisions about angry faces is not only uniquely associated with physical aggression but also the only element of postdecisional processing through which physical aggression is linked to angry rumination.

Discussion

Previous research suggests that physical aggression is associated with aberrations in both the formation and maintenance of threat-based social decisions. The results of the present study indicate that these aberrations may stem, in part, from distinctive patterns of postdecisional processing. Because we used a novel

experimental task designed to assess postdecisional processing after facial emotion decisions, this study is the first empirical examination of how social decisions unfold in real time among physically aggressive individuals. It is worth highlighting that the validity of the task was established by a series of successful manipulation checks. Results indicated that at the emotion-decision-formation stage, physical aggression was associated with better differentiation of mostly angry (i.e., threatening) and mostly happy (i.e., nonthreatening) faces, but only at moderate levels of ambiguity. Moreover, we found that physical aggression was associated with steeper decreases in confidence over time when mostly angry (i.e., threatening) faces were identified as happy (i.e., nonthreatening). Finally, this pattern of postdecisional processing mediated the association between physical aggression and angry rumination.

The finding that physical aggression was associated with superior differentiation between threatening and nonthreatening faces under moderate ambiguity was consistent with our hypothesis about the formation of facial emotion decisions. However, two caveats should be noted.

First, we expected that more physically aggressive individuals would show a greater likelihood of identifying faces as angry. However, we did not necessarily expect that they would show a combination of tendencies toward both heightened anger identification and heightened happiness identification when these decisions were warranted (i.e., heightened differentiation between mostly angry and mostly happy faces). Yet this finding is consistent with previous research that suggested that aggressive individuals are more sensitive to subtle changes in the amount of anger displayed in faces and adjust their responses accordingly (Wilkowski & Robinson, 2012). Moreover, this finding adds to evidence that physical aggression is associated with more adept, and not biased, anger processing (Brennan & Baskin-Sommers, 2020).

Second, although we expected to detect an association between physical aggression and angry decisions under greater ambiguity, our effect was within the moderate-ambiguity condition but not the high-ambiguity condition. Performance in the high-ambiguity condition (55%/45% blends of each emotion) was quite poor (see Table 1). Performance near chance levels under high ambiguity likely created substantial noise that made it difficult to detect an effect of physical aggression (Siegelman, Bogaerts, & Frost, 2017). Furthermore, this pattern of results suggests that there may be boundary conditions to the association between physical aggression and the tendency to identify faces as angry as ambiguity increases. Although this association may emerge after ambiguity levels exceed a certain threshold, this association may be evident only up to a point, after which stimuli become too ambiguous and the evidence for decision-making becomes too degraded.

Turning to our next set of hypotheses regarding the extent to which confidence was affected by ambiguity and the emotion decision made, we did not find evidence that physical aggression was associated with less modulation of confidence as a function of ambiguity or heightened confidence in angry decisions. Both of these hypotheses were based on an earlier study by Brennan and Baskin-Sommers (2019), in which participants completed a social-decision-making task. In the task, participants gathered information about the negative and positive behaviors of a hypothetical person and then decided whether the person was "nasty" or "nice." This task differed from the present task in several important ways.

First, unlike the present study, the Brennan and Baskin-Sommers (2019) study did not directly manipulate ambiguity. Rather, it was inferred that more physically aggressive individuals made decisions under greater ambiguity because they gathered less information to support their hostile decisions, suggesting a weaker evidence base. Despite this, however, physically aggressive individuals reported greater confidence in their hostile decisions compared with less physically aggressive individuals. Thus, the true impact of ambiguity could not be quantified directly in the Brennan and Baskin-Sommers study.

Second, whereas the present study examined facial-emotion identification, the Brennan and Baskin-Sommers (2019) study examined trait judgments. Physically aggressive individuals may calibrate confidence differently for these distinct types of social decisions rather than display overconfidence across all decisions and situations. Finally, the stimuli in the Brennan and Baskin-Sommers (2019) task consisted of negative and positive behaviors, in contrast with emotional faces in the present task. The negative behaviors (e.g., "offended a man") could be conceptualized as indirect provocations; however, facial cues of anger do not, on their own, constitute provocations (da Cunha-Bang et al., 2017; Lemerise, Gregory, & Fredstrom, 2005; Lickley & Sebastian, 2018). Thus, the presence of provocation, even indirect, might contribute to heightened confidence in threat-based decisions among physically aggressive individuals (Bertsch, Böhnke, Kruk, & Naumann, 2009). Altogether, differences between studies in task design and stimuli may account for the inconsistencies observed.

In terms of postdecisional processing (i.e., change in confidence over time), physically aggressive individuals exhibited steeper decreases in confidence for

mostly angry faces. This finding indicated a pattern of more extensive postdecisional processing of threatening faces. Specifically, more physically aggressive individuals continued to accumulate evidence about threatening faces after they decided on the emotion displayed in these faces. The specificity of this interaction to mostly angry faces suggests that the predominantly threatening information conveyed in mostly angry faces was more readily processed and stored in memory than the predominantly nonthreatening information conveyed in mostly happy faces. This finding is consistent with previous studies suggesting that more physically aggressive individuals show preferential processing of threat-related information in general (e.g., Smith & Waterman, 2003) and stronger memory for angry faces in particular (d'Acremont & Van der Linden, 2007).

Furthermore, the finding that postdecisional processing of mostly angry faces depended on emotion decision provides insight into the effectiveness of postdecisional processing in physical aggression. Postdecisional processing can be considered effective to the extent that it steers decision-makers away from incongruent decisions and toward congruent decisions. In other words, more effective postdecisional processing brings decisions more in line with the preponderance of evidence available for decision-making, which, in the present study, was the dominant emotion displayed in the faces. Therefore, the fact that physically aggressive individuals only showed steeper decreases in confidence for incongruent, but not congruent, decisions about mostly angry faces is consistent with more effective postdecisional processing of threatening faces. This finding aligns with and extends previous research that linked physical aggression to more efficient evidence accumulation for anger during the formation of social decisions (Brennan & Baskin-Sommers, 2020). Across these studies, and consistent with predictions of decision-making theory (Pleskac & Busemeyer, 2010), physically aggressive individuals exhibit heightened evidence accumulation, which may support more effective processing of threatening social information at both the formation and maintenance stages of social decision-making.

Despite being more effective, the pattern of postdecisional processing observed in more physically aggressive individuals could nevertheless make threat-based decisions more likely to emerge over time when real threats exist. Threat-based decisions could become more likely because as the decision-maker loses faith in the initial non-threat-based decision, the alternative threat-based decision becomes more plausible. As a result, the decision might be reversed from non-threat-based to threat-based, and in turn, the decision-maker may become more likely to aggress to neutralize the newly recognized threat. This finding is consistent with observations that betrayal by someone considered to be a friend is a powerful trigger for aggression (Lawrence, 2006) and that violent retaliation is often delayed rather than immediate (Bushman & Anderson, 2001). An enhanced ability to recognize threats that were not initially detected may be acquired through chronic exposure to threatening environments (Guerra, Huesmann, & Spindler, 2003; Weiss, Dodge, Bates, & Pettit, 1992) and is likely adaptive in environments that contain real threats.

Recognizing real social threats that were not initially detected may also relate to angry rumination. The tendency to ruminate on social threats is robustly linked to physical aggression. The present results suggest that more effective postdecisional processing of social threat may play a role in the relationship between physical aggression and angry rumination. Specifically, more dramatic decreases in confidence over time for incongruent decisions about threatening stimuli might increase the plausibility of threat-based decisions, in turn leading angry ruminative content (e.g., thinking about how someone wronged you) to feature more prominently in awareness. The idea that physically aggressive individuals' threat-based decisions gain plausibility over time under these circumstances suggests that physically aggressive individuals might have to deploy even more cognitive control than less aggressive individuals to disengage from these decisions. This interpretation is consistent with work suggesting that angry rumination in physically aggressive individuals is related to a failure of cognitive control to interrupt perseverative thinking (Denson, 2013; Wilkowski & Robinson, 2010). These insights into the mechanisms of angry rumination in physical aggression lend themselves to clinical implications.

Distraction is a clinical tool that effectively reduces aggressive behavior (Gallagher & Parrott, 2011; Giancola & Corman, 2007; Subramani, Parrott, Latzman, & Washburn, 2019). One possibility is that distraction may interrupt the postdecisional accumulation of evidence that decreases confidence in decisions that others are nonthreatening. However, the present findings suggest that distraction might be needed even when others are initially seen as nonthreatening, presenting an obstacle to effectively identifying when to use distraction. Moreover, because physically aggressive individuals' postdecisional processing may be adaptive in threatening environments, mindfulness-based interventions targeted toward strengthening nonjudgmental awareness of (vs. eliminating) decisions that others are threatening could be beneficial (Wright, Day, & Howells, 2009). Identifying individuals who would benefit most

from intervention is crucial as well. Because the mechanisms identified in the present study are likely to be relatively entrenched by the time an individual reaches adulthood, intervening earlier in development (e.g., during adolescence; see Dickerson, Skeem, Montoya, & Quas, 2020) may be advantageous. Furthermore, negative emotionality appears important for contextualizing the association between physical aggression and postdecisional processing of social threat (see Table S1 in the Supplemental Material), which suggests that interventions targeting negative emotionality might be useful.

Before concluding, limitations of the present study should be noted. First, because our sample was limited to male offenders, it is unclear whether the results would generalize to other populations, such as female offenders or nonincarcerated individuals. However, because male offenders perpetrate physical violence at high rates, understanding aggression in this population is particularly important. Future research should seek to replicate findings in other samples.

Second, we used emotional face stimuli that displayed only anger and happiness. As a result, we do not know whether the steeper decrease in confidence for incongruent decisions about angry faces is specific to happy decisions or would apply more broadly to any incongruent decisions about angry faces (e.g., if mostly angry faces were identified as sad, afraid, surprised). Our decision to use only anger and happiness was based on several important considerations, including the desire to compare processing of social threat (i.e., anger) with nonthreat (i.e., happiness) and maintain consistency with previous research. However, future research should test the generalizability of findings by using face stimuli displaying a wider range of emotions.

Third, because we did not formally assess certain forms of psychopathology that have been linked to aberrant emotional processing (e.g., anxiety disorders, depression), we could not evaluate the impact of these factors on task performance.

Fourth, although confidence can be considered an indicator of metacognition (i.e., one's awareness of one's own cognitive processing), our analyses did not separate different components of metacognition (e.g., metacognitive sensitivity vs. bias). As a result, important questions remain regarding physically aggressive individuals' metacognition in the context of social decision-making, and applying a metacognitive framework in future research would likely be fruitful.

Finally, the presence of empty cells because of lack of response variability within some task conditions prevented us from testing our full statistical model. Future studies using this paradigm can avoid this limitation by increasing trial numbers within conditions or removing the low-ambiguity condition to ensure response variability.

The present study shows how social decision-making unfolds in real time in physical aggression and contributes to mounting evidence that physically aggressive individuals exhibit more effective anger processing capabilities. Because the maintenance stage of social decision-making represents an important but neglected topic as it relates to physical aggression, this work contributes to building a framework for understanding how and why physically aggressive individuals persist in seeing others as threatening. Finally, the development of a novel paradigm to examine postdecisional processing with social stimuli presents exciting possibilities for future research into whether other behaviors and disorders marked by rumination and aberrant processing of social threat (e.g., social anxiety disorder) are associated with distinct patterns of postdecisional processing as well.

Transparency

Action Editor: Kelly L. Klump
Editor: Kenneth J. Sher
Author Contributions

G. M. Brennan developed the study concept. G. M. Brennan and A. Baskin-Sommers designed the study and performed clinical and neuropsychological assessments. G. M. Brennan collected some of the task data. G. M. Brennan performed data analysis and interpretation under the supervision of A. Baskin-Sommers. G. M. Brennan drafted the manuscript, and A. Baskin-Sommers provided critical revisions. Both of the authors approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This work was supported by a Harry Frank Guggenheim Foundation grant and an American Psychology-Law Society grant.

ORCID iD

Grace M. Brennan https://orcid.org/0000-0003-0246-0820

Acknowledgments

We thank Connecticut Department of Correction staff for their support and the research assistants who helped collect these data. The data have not been made available on a permanent third-party archive because the combination of demographic and crime variables makes it possible to identify participants. However, for a deidentified subset of the data (or study materials), please contact the authors.

Notes

1. Previous studies from our research group used partially overlapping samples of incarcerated males—53% of participants in the present study participated in the Brennan and Baskin-Sommers (2020) study, and 29% of participants in the present

study participated in the Brennan and Baskin-Sommers (2019) study. However, participants completed these separate studies at least several months apart, and participants in the present study had not been exposed to the experimental stimuli previously.

2. Age was included as a covariate in this analysis (and all analyses to follow) because it was associated with task dependent variables.

3. To protect against violations of the assumption of sphericity, we report Huynh-Feldt corrected *p* values for all GLM analyses.

4. One participant was not included in the mediation analysis because the experimental session was cut short before he could complete the ARS measure.

References

AlMoghrabi, N., Huijding, J., & Franken, I. H. A. (2018). The effects of a novel hostile interpretation bias modification paradigm on hostile interpretations, mood, and aggressive behavior. *Journal of Behavior Therapy and Experimental Psychiatry*, *58*, 36–42. doi:10.1016/j.jbtep.2017.08.003

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.

Anestis, M. D., Anestis, J. C., Selby, E. A., & Joiner, T. E. (2009). Anger rumination across forms of aggression. *Personality and Individual Differences*, *46*, 192–196. doi:10.1016/j.paid.2008.09.026

Archer, J., & Haigh, A. (1997). Beliefs about aggression among male and female prisoners. *Aggressive Behavior*, *23*, 405–415. doi:10.1002/(SICI)1098-2337(1997)23:6 < 405::AID-AB1 > 3.0.CO;2-F

Bertsch, K., Böhnke, R., Kruk, M., & Naumann, E. (2009). Influence of aggression on information processing in the emotional Stroop task—An event-related potential study. *Frontiers in Behavioral Neuroscience*, *3*, Article 28. doi:10.3389/neuro.08.028.2009

Bierman, K. L., & Wargo, J. B. (1995). Predicting the longitudinal course associated with aggressive-rejected, aggressive (nonrejected), and rejected (nonaggressive) status. *Development and Psychopathology*, *7*, 669–682. doi:10.1017/S0954579400006775

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897X00357

Brennan, G. M., & Baskin-Sommers, A. R. (2019). Physical aggression is associated with heightened social reflection impulsivity. *Journal of Abnormal Psychology*, *128*, 404–414. doi:10.1037/abn0000448

Brennan, G. M., & Baskin-Sommers, A. R. (2020). Aggressive realism: More efficient processing of anger in physically aggressive individuals. *Psychological Science*, *31*, 568–581. doi:10.1177/0956797620904157

Bronson, J., & Carson, E. A. (2019). *Prisoners in 2017* (NCJ Publication No. 252156). Retrieved from https://www.bjs.gov/content/pub/pdf/p17.pdf

Burt, S. A., Mikolajewski, A. J., & Larson, C. L. (2009). Do aggression and rule-breaking have different interpersonal correlates? A study of antisocial behavior subtypes, negative affect, and hostile perceptions of others. *Aggressive Behavior*, *35*, 453–461. doi:10.1002/ab.20324

Bushman, B. J. (2002). Does venting anger feed or extinguish the flame? Catharsis, rumination, distraction, anger and aggressive responding. *Personality and Social Psychology Bulletin*, *28*, 724–731. doi:10.1177/0146167202289002

Bushman, B. J., & Anderson, C. A. (2001). Is it time to pull the plug on the hostile versus instrumental aggression dichotomy? *Psychological Review*, *108*, 273–279.

Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology*, *63*, 452–459.

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.

d'Acremont, M., & Van der Linden, M. (2007). Memory for angry faces, impulsivity, and problematic behavior in adolescence. *Journal of Abnormal Child Psychology*, *35*, 313–324. doi:10.1007/s10802-006-9092-1

da Cunha-Bang, S., Fisher, P. M., Hjordt, L. V., Perfalk, E., Persson Skibsted, A., Bock, C., . . . Knudsen, G. M. (2017). Violent offenders respond to provocations with high amygdala and striatal reactivity. *Social Cognitive & Affective Neuroscience*, *12*, 802–810. doi:10.1093/scan/nsx006

De Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development*, *73*, 916–934. doi:10.1111/1467-8624.00447

Denson, T. F. (2013). The multiple systems model of angry rumination. *Personality & Social Psychology Review*, *17*, 103–123. doi:10.1177/1088868312467086

Dickerson, K. L., Skeem, J. L., Montoya, L., & Quas, J. A. (2020). Using positive emotion training with maltreated youths to reduce anger bias and physical aggression. *Clinical Psychological Science*, *8*, 773–787. doi:10.1177/2167702620902118

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.

Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science*, *20*, 1183–1188. doi:10.1111/j.1467-9280.2009.02422.x

Gallagher, K. E., & Parrott, D. J. (2011). Does distraction reduce the alcohol-aggression relation? A cognitive and behavioral test of the attention-allocation model. *Journal of Consulting and Clinical Psychology*, *79*, 319–329. doi:10.1037/a0023065

Giancola, P. R., & Corman, M. D. (2007). Alcohol and aggression: A test of the attention-allocation model. *Psychological Science*, *18*, 649–655. doi:10.1111/j.1467-9280.2007.01953.x

Guerra, N. G., Huesmann, L. R., & Spindler, A. (2003). Community violence exposure, social cognition, and aggression among urban elementary school children. *Child Development*, *74*, 1561–1576. doi:10.1111/1467-8624.00623

Harris, J. A. (1997). A further evaluation of the Aggression Questionnaire: Issues of validity and reliability. *Behaviour Research and Therapy*, *35*, 1047–1053. doi:10.1016/S0005-7967(97)00064-8

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). New York, NY: Guilford Press.

Huesmann, L. R., Dubow, E. F., & Boxer, P. (2009). Continuity of aggression from childhood to early adulthood as a predictor of life outcomes: Implications for the adolescent-limited and life-course-persistent models. *Aggressive Behavior*, *35*, 136–149. doi:10.1002/ab.20300

Ireland, J. L., & Archer, J. (2004). Association between measures of aggression and bullying among juvenile and young offenders. *Aggressive Behavior*, *30*, 29–42. doi:10.1002/ab.20007

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1–16. doi:10.1068/v070821

Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking antisocial behavior, substance use, and personality: An integrative quantitative model of the adult externalizing spectrum. *Journal of Abnormal Psychology*, *116*, 645–666. doi:10.1037/0021-843x.116.4.645

Lawrence, C. (2006). Measuring individual responses to aggression-triggering events: Development of the Situational Triggers of Aggressive Responses (STAR) Scale. *Aggressive Behavior*, *32*, 241–252. doi:10.1002/ab.20122

Lemerise, E. A., Gregory, D. S., & Fredstrom, B. K. (2005). The influence of provocateurs' emotion displays on the social information processing of children varying in social adjustment and age. *Journal of Experimental Child Psychology*, *90*, 344–366. doi:10.1016/j.jecp.2004.12.003

Lickley, R. A., & Sebastian, C. L. (2018). The neural basis of reactive aggression and its development in adolescence. *Psychology, Crime & Law*, *24*, 313–333. doi:10.1080/1068316X.2017.1420187

Maoz, K., Eldar, S., Stoddard, J., Pine, D. S., Leibenluft, E., & Bar-Haim, Y. (2016). Angry-happy interpretations of ambiguous faces in social anxiety disorder. *Psychiatry Research*, *241*, 122–127. doi:10.1016/j.psychres.2016.04.100

McLaughlin, K. A., Aldao, A., Wisco, B. E., & Hilt, L. M. (2014). Rumination as a transdiagnostic factor underlying transitions between internalizing symptoms and aggressive behavior in early adolescents. *Journal of Abnormal Psychology*, *123*, 13–23. doi:10.1037/a0035358

Mellentin, A. I., Dervisevic, A., Stenager, E., Pilegaard, M., & Kirk, U. (2015). Seeing enemies? A systematic review of anger bias in the perception of facial expressions among anger-prone and aggressive populations. *Aggression and Violent Behavior*, *25*, 373–383. doi:10.1016/j.avb.2015.09.001

Murphy, P. R., Robertson, I. H., Harty, S., & O'Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. *eLife*, *4*, Article e11946. doi:10.7554/eLife.11946

Okuda, M., Picazo, J., Olfson, M., Hasin, D. S., Liu, S.-M., Bernardi, S., & Blanco, C. (2015). Prevalence and correlates of anger in the community: Results from a national survey. *CNS Spectrums*, *20*, 130–139. doi:10.1017/S1092852914000182

Peled, M., & Moretti, M. M. (2007). Rumination on anger and sadness in adolescence: Fueling of fury and deepening of despair. *Journal of Clinical Child & Adolescent Psychology*, *36*, 66–75. doi:10.1080/15374410709336569

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Penton-Voak, I. S., Thomas, J., Gage, S. H., McMurran, M., McDonald, S., & Munafo, M. R. (2013). Increasing recognition of happiness in ambiguous facial expressions reduces anger and aggressive behavior. *Psychological Science*, *24*, 688–697. doi:10.1177/0956797612459657

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*, 864–901. doi:10.1037/a0019737

Poulin, F., & Boivin, M. (1999). Proactive and reactive aggression and boys' friendship quality in mainstream classrooms. *Journal of Emotional and Behavioral Disorders*, *7*, 168–177. doi:10.1177/106342669900700305

Rand, D. G., Ohtsuki, H., & Nowak, M. A. (2009). Direct reciprocity with costly punishment: Generous tit-for-tat prevails. *Journal of Theoretical Biology*, *256*, 45–57.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922. doi:10.1162/neco.2008.12-06-420

Sauter, D. (2010). More than happy: The need for disentangling positive emotions. *Current Directions in Psychological Science*, *19*, 36–40. doi:10.1177/0963721409359290

Schönenberg, M., & Jusyte, A. (2014). Investigation of the hostile attribution bias toward ambiguous facial cues in antisocial violent offenders. *European Archives of Psychiatry and Clinical Neuroscience*, *264*, 61–69. doi:10.1007/s00406-013-0440-1

Schulte-Rüther, M., Markowitsch, H. J., Fink, G. R., & Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: A functional magnetic resonance imaging approach to empathy. *Journal of Cognitive Neuroscience*, *19*, 1354–1372. doi:10.1162/jocn.2007.19.8.1354

Siegel, J. Z., Mathys, C., Rutledge, R. B., & Crockett, M. J. (2018). Beliefs about bad people are volatile. *Nature Human Behaviour*, *2*, 750–756 doi:10.1038/s41562-018-0425-1

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, *49*, 418–432. doi:10.3758/s13428-016-0719-z

Smith, P., & Waterman, M. (2003). Processing bias for aggression words in forensic and nonforensic samples. *Cognition & Emotion*, *17*, 681–701. doi:10.1080/02699930302281

Subramani, O. S., Parrott, D. J., Latzman, R. D., & Washburn, D. A. (2019). Breaking the link: Distraction from emotional cues reduces the association between trait disinhibition and reactive physical aggression. *Aggressive Behavior*, *45*, 151–160. doi:10.1002/ab.21804

Sukhodolsky, D. G., Golub, A., & Cromwell, E. N. (2001). Development and validation of the Anger Rumination Scale. *Personality and Individual Differences*, *31*, 689–700. doi:10.1016/S0191-8869(00)00171-9

Teige-Mocigemba, S., Hölzenbein, F., & Klauer, K. C. (2016). Seeing more than others: Identification of subtle aggressive information as a function of trait aggressiveness. *Social Psychology*, *47*, 136–149. doi:10.1027/1864-9335/a000266

Thome, J., Liebke, L., Bungert, M., Schmahl, C., Domes, G., Bohus, M., & Lis, S. (2016). Confidence in facial emotion recognition in borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment*, *7*, 159–168. doi:10.1037/per0000142

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social Cognitive and Affective Neuroscience*, *3*, 119–127. doi:10.1093/scan/nsn009

Tremblay, P. F., & Ewart, L. A. (2005). The Buss and Perry Aggression Questionnaire and its relations to values, the Big Five, provoking hypothetical situations, alcohol consumption patterns, and alcohol expectancies. *Personality and Individual Differences*, *38*, 337–346. doi:10.1016/j.paid.2004.04.012

van den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). A common mechanism underlies changes of mind about decisions and confidence. *eLife*, *5*, Article e12192. doi:10.7554/eLife.12192

Van Zandt, T., & Maldonado-Molina, M. M. (2004). Response reversals in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1147–1166. doi:10.1037/0278-7393.30.6.1147

Weiss, B., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1992). Some consequences of early harsh discipline: Child aggression and a maladaptive social information processing style. *Child Development*, *63*, 1321–1335.

Wilkowski, B. M., & Robinson, M. D. (2008). The cognitive basis of trait anger and reactive aggression: An integrative analysis. *Personality and Social Psychology Review*, *12*, 3–21. doi:10.1177/1088868307309874

Wilkowski, B. M., & Robinson, M. D. (2010). The anatomy of anger: An integrative cognitive model of trait anger and reactive aggression. *Journal of Personality*, *78*, 9–38. doi:10.1111/j.1467-6494.2009.00607.x

Wilkowski, B. M., & Robinson, M. D. (2012). When aggressive individuals see the world more accurately: The case of perceptual sensitivity to subtle facial expressions of anger. *Personality and Social Psychology Bulletin*, *38*, 540–553. doi:10.1177/0146167211430233

Wright, S., Day, A., & Howells, K. (2009). Mindfulness and the treatment of anger problems. *Aggression and Violent Behavior*, *14*, 396–401. doi:10.1016/j.avb.2009.06.008

Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. *Journal of Experimental Psychology: General*, *144*, 489–510. doi:10.1037/xge0000062

Zimmer-Gembeck, M. J., & Nesdale, D. (2013). Anxious and angry rejection sensitivity, social withdrawal, and retribution in high and low ambiguous situations. *Journal of Personality*, *81*, 29–38. doi:10.1111/j.1467-6494.2012.00792.x