Fall 10-1-2021

Cognitive Mechanisms Supporting the Formation and Maintenance of Social Judgments in Physical Aggression

Grace Marie Brennan
Yale University Graduate School of Arts and Sciences, gmbrennan7@gmail.com

Abstract

Cognitive Mechanisms Supporting the Formation and
Maintenance of Social Judgments in Physical Aggression

Grace Marie Brennan

2021

Physical aggression is a harmful yet ubiquitous form of human behavior. A large body of research has established that physical aggression is rooted in aberrations at the formation and maintenance stages of social cognition. At the formation stage, more physically aggressive individuals are more likely to interpret ambiguous social stimuli as threatening; at the maintenance stage, more physically aggressive individuals are more likely to “hold on” to interpretations of others as threatening. However, very little research has examined the cognitive mechanisms that contribute to these aberrations during online social decision-making. The three experimental studies that comprise this dissertation apply theory and methods from the cognitive and decision sciences to specify the influences of putative cognitive mechanisms, namely initial bias (the starting point of an individual’s decision-making), efficiency of evidence accumulation (the quality of evidence extracted from a stimulus), and extent of evidence accumulation (the quantity of evidence gathered for a decision). The three studies provide a comprehensive perspective on social cognition by examining both lower-order facial emotion judgments and higher-order trait judgments in samples of incarcerated male offenders. Study 1 applies a form of computational modeling called diffusion modeling to parse the cognitive mechanisms contributing to the formation of lower-order facial emotion judgments. The findings of Study 1 suggest that more physically aggressive individuals display more efficient accumulation of anger-related evidence, which may help explain physically aggressive individuals’ heightened tendency to perceive ambiguous faces as threatening. Study 2 focuses on the extent of evidence accumulation and examines its role in the formation of higher-order trait judgments using a novel adaptation of an established experimental task. The findings of Study 2 suggest that more physically aggressive individuals display less extensive evidence accumulation while making trait judgments, particularly hostile trait judgments. Finally, Study 3 focuses on the maintenance of lower-order facial emotion judgments by examining post-decisional processing, again using a novel adaptation of an established experimental task. The findings of Study 3 suggest that more efficient accumulation of anger-related evidence in physically aggressive individuals is also evident *following* emotion decisions, which may help account for the persistence of threat-based social judgments over time in physical aggression. Taken together, this set of studies provides novel insights into the cognitive mechanisms driving aberrant social cognition in physical aggression. Implications for theory and clinical practice are discussed, as well as directions for future research.

Cognitive Mechanisms Supporting the Formation and Maintenance of Social Judgments in Physical Aggression

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Grace M. Brennan

Dissertation Director: Arielle R. Baskin-Sommers

December 2021

Acknowledgments

I would like to thank my advisor, Arielle Baskin-Sommers, for her tireless efforts to train and support me. She has been an endless source of constructive feedback and resources, and her mentorship has been instrumental in my development as a writer, teacher, clinician, project manager, critical thinker, and scientist. Throughout my time in graduate school, I have been able to turn to her for advice about a wide range of topics, and she has always helped me navigate and process difficult situations. She is incredibly generous with her time when it comes to helping students. I aspire to achieve anywhere near her level of commitment and generosity.

I would also like to thank the other members of my dissertation committee for generously offering up their time and feedback on my work: Dylan Gee, Ty Cannon, Maria Gendron, and BJ Casey. I would like to extend gratitude to Dylan Gee and Ty Cannon for providing detailed and thoughtful feedback on my work at multiple stages of my graduate school career. My scientific thinking has benefited tremendously from their input.

I would also like to thank Mary O'Brien for providing excellent clinical training and supporting students in countless ways. Additionally, I am grateful for the opportunities I have had to work with faculty members as a Teaching Fellow for various courses, particularly the opportunity to learn about teaching from Paul Bloom.

I am grateful to those who helped with collecting data for this dissertation, namely Scott Tillem, Emil Beckford, Andrew Del Vecchio, and Cole Rianda. I would also like to thank the members of the MoD Lab—past and present—who completed crucial research-related tasks, asked insightful questions, and provided helpful feedback on projects.

I would also like to thank the Connecticut Department of Correction staff for allowing us to conduct research within their space at Cheshire Correctional Institution. Additionally, I appreciate the inmates at Cheshire who participated in this research. Participants commonly agreed to participate based on the idea that the research might benefit others. I admire their altruism and willingness to share their experiences, and it is my hope that this research does indeed benefit others.

Finally, I am grateful to my family for their unconditional love and support. My father, Peter Brennan, has been my fiercest supporter throughout my life. He taught me the value of hard work and perseverance, and without his unrelenting belief in me I would not be where I am today. I would also like to thank my fiancée, Jared Bollinger, for being perpetually willing to review drafts of my writing, for supporting me in the pursuit of my goals, and for being a source of joy in my life.

Table of Contents

Chapter 1: General Introduction

Physical aggression, which includes violent crimes such as assault, robbery, and homicide, carries enormous costs for victims, perpetrators, and society at large. Each year in the United States, physical aggression results in over 1.6 million non-fatal injuries that require treatment in emergency departments (Sumner et al., 2015) and over 19,500 deaths (Centers for Disease Control and Prevention, 2019). Perpetrators of physical aggression experience severe impairments and poor life outcomes across a variety of domains, including relationships, physical health, mental health, and criminal justice system involvement (Bierman & Wargo, 1995; Huesmann, Dubow, & Boxer, 2009; Okuda et al., 2015; Poulin & Boivin, 1999). Of the 1.3 million inmates in state prisons in the United States, over half are currently serving sentences for a violent crime (Bronson & Carson, 2019). In sum, physical aggression is a destructive, costly, and common form of human behavior.

Remarkable progress has been made in identifying factors that contribute to physical aggression. Based on previous research, it is clear that physical aggression is influenced by both environmental (e.g., harsh and coercive parenting, peer rejection, affiliation with antisocial peers, exposure to violence; Guerra, Huesmann, & Spindler, 2003; Lansford, Malone, Dodge, Pettit, & Bates, 2010; Patterson, Reid, & Dishion, 1992; Powers, Bierman, & The Conduct Problems Prevention Research Group, 2013; Vitaro, Barker, Boivin, Brendgen, & Tremblay, 2006) and intra-individual (e.g., genetics, neurobiology, personality; Caspi et al., 2002; Chester, Lynam, Milich, & DeWall, 2017; Hare & McPherson, 1984) factors. At the nexus of these two types of influences is social cognition, or how individuals process and interpret social information in their environments (e.g., the faces and behaviors of others).

Decades of research link physical aggression to aberrations in social cognition (Dodge, 1980; Dodge & Crick, 1990; Lansford et al., 2006; Lochman & Dodge, 1994). From a decision-making perspective, social cognition can be conceptualized as proceeding through different stages of processing. First, at the formation stage, evidence is accumulated to inform an initial judgment about a stimulus (e.g., whether someone poses a potential threat or not; Crick & Dodge, 1994; Ratcliff & McKoon, 2008). Social information processing theory (Dodge & Crick, 1990; Crick & Dodge, 1994), an influential model of aggression, largely focuses on this stage of social cognition, highlighting the importance of encoding and interpreting social cues for promoting aggressive behavior (Dodge, 2006). Next, at the maintenance stage, which begins after an initial judgment has been made, evidence about the stimulus continues to be accumulated. Based on the incoming evidence, the initial judgment may gain or lose strength and may be revised (Pleskac & Busemeyer, 2010). Previous research establishes that physical aggression is associated with aberrations in social cognition that span these two stages.

At the formation stage, physically aggressive individuals are more likely to judge social stimuli as threatening. They display a heightened tendency to identify ambiguous faces as angry (Mellentin, Dervisevic, Stenager, Pilegaard, & Kirk, 2015; Schönenberg & Jusyte, 2014; Wilkowski & Robinson, 2012) and interpret others' ambiguous actions as being carried out with hostile intent (De Castro, Veerman, Koops, Bosch, & Monshouwer, 2002; Dodge, 1980). For example, studies examining hostile attribution biases assess individuals' responses to hypothetical vignettes and find that more physically aggressive individuals are more likely to make hostile interpretations of situations involving ambiguous provocation (Dodge, 1980; Lansford et al., 2010; Coccaro, Noblett, & McCloskey, 2009; Dodge, Price, Bachorowski, & Newman, 1990). At the maintenance stage, physically aggressive individuals' threat-based social judgments are more likely to persist over time. For example, studies examining angry rumination assess individuals' self-reported general tendencies and find that more physically aggressive individuals report higher levels of angry rumination, or perseverative 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). Angry rumination appears to strengthen threat-based judgments over time, making physically aggressive individuals less likely to disengage from them. Taken together, both heightened threat identification at the formation stage and heightened rumination at the maintenance stage of social cognition increase risk for physical aggression (Dodge, 2006; McLaughlin, Aldao, Wisco, & Hilt, 2014).

Despite advances in identifying aberrations in social cognition associated with physical aggression, limitations of previous research leave important questions unanswered. Previous studies examining the formation stage of social cognition tend to focus on outcomes of decision-making and only one behavioral measure in isolation. For example, most previous studies measure either response accuracy or frequency of a particular response, most notably the frequency of responses reflecting threat-based judgments. Very few studies have identified cognitive processes that lead to these outcomes during online decision-making or examined multiple behavioral indicators simultaneously (e.g., both responses and reaction time). Previous studies examining the maintenance stage of social cognition rely on self-report measures assessing the extent to which people endorse experiencing persistent anger-promoting thoughts about others in general. No research has directly examined how physically aggressive individuals' social judgments unfold in real time after they are formed. Thus, it remains unclear which cognitive mechanisms underlie the formation and maintenance of social judgments in physical aggression.

According to prominent theories of decision-making (Forstmann, Ratcliff, & Wagenmakers, 2016; Pleskac & Busemeyer, 2010; Ratcliff, 1978), multiple cognitive mechanisms exert distinct influences on the decision-making process and impact which decisions individuals make. These theories posit that decisions are made by accumulating evidence until one of two response thresholds is reached, at which point the corresponding decision is made. The key mechanisms in these theories are initial bias (the starting point of an individual's decision-making), efficiency of evidence accumulation (the quality of evidence extracted from a stimulus), and extent of evidence accumulation (the quantity of evidence gathered for a decision). Crucially, the process of evidence accumulation takes place not only *before* an initial decision has been made—but also it continues *after* an initial decision has been made (Pleskac & Busemeyer, 2010). Thus, decision-making is influenced by a complex interplay of multiple cognitive mechanisms, whose contributions to physically aggressive individuals' social cognition remain poorly understood.

This dissertation leverages insights and methodologies from the cognitive and decision sciences to identify mechanisms that contribute to aberrant social cognition in physical aggression. The three studies that comprise this dissertation form a comprehensive examination of social cognition across the stages of formation and maintenance, and also across levels of social judgments (i.e., lower-order facial emotion judgments and higher-order trait judgments). Study 1 applies a form of computational modeling called diffusion modeling to parse the cognitive mechanisms contributing to the formation of lower-order facial emotion judgments. Study 2 focuses on the extent of evidence accumulation and examines its role in the formation of higher-order trait judgments using a novel adaptation of an established experimental task. Finally, Study 3 focuses on the maintenance of lower-order facial emotion judgments by examining postdecisional processing, again using a novel adaptation of an established experimental task. Taken together, this set of studies aims to elucidate the cognitive mechanisms that contribute to the formation and maintenance of social judgments in physical aggression.

References: General Introduction

Anestis, M. D., Anestis, J. C., Selby, E. A., & Joiner, T. E. (2009). Anger rumination across forms of aggression. *Personality and Individual Differences, 46*(2), 192-196. doi:https://doi.org/10.1016/j.paid.2008.09.026

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*(4), 669-682. doi:10.1017/S0954579400006775

Bronson, J., & Carson, E. A. (2019). Prisoners in 2017 (NCJ Publication No. 252156). Retrieved from http://www.bjs.gov/index.cfm?ty=pbdetail&iid=6187

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*(6), 724-731. doi:10.1177/0146167202289002

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., . . . Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science, 297*(5582), 851-854. doi:10.1126/science.1072290

Centers for Disease Control and Prevention. (2019). National Violent Death Reporting System. Retrieved from https://www.cdc.gov/violenceprevention/datasources/nvdrs/index.html

Chester, D. S., Lynam, D. R., Milich, R., & DeWall, C. N. (2017). Physical aggressiveness and gray matter deficits in ventromedial prefrontal cortex. *Cortex, 97*, 17-22. doi:10.1016/j.cortex.2017.09.024

Coccaro, E. F., Noblett, K. L., & McCloskey, M. S. (2009). Attributional and emotional responses to socially ambiguous cues: Validation of a new assessment of social/emotional information processing in healthy adults and impulsive aggressive patients. *Journal of Psychiatric Research, 43*(10), 915-925. doi:10.1016/j.jpsychires.2009.01.012

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.

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*(2), 103-123. doi:10.1177/1088868312467086

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*(3), 791-814.

Dodge, K. A., & Crick, N. R. (1990). Social information-processing bases of aggressive behavior in children. *Personality and Social Psychology Bulletin, 16,* 8-22. doi:10.1177/0146167290161002

Dodge, K. A., Price, J. M., Bachorowski, J.-A., & Newman, J. P. (1990). Hostile attributional biases in severely aggressive adolescents. *Journal of Abnormal Psychology, 99*, 385-392. doi:10.1037/0021-843X.99.4.385

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology*, *67*, 641-666. doi:10.1146/annurev-psych-122414-033645

Guerra, N. G., Huesmann, L. R., & Spindler, A. (2003). Community violence exposure, social cognition, and aggression among urban elementary school children. *Child Development, 74*(5), 1561-1576. doi:10.1111/1467-8624.00623

Hare, R. D., & McPherson, L. M. (1984). Violent and aggressive behavior by criminal psychopaths. *International Journal of Law & Psychiatry*, *7*(1), 35-50. doi:10.1016/0160-2527(84)90005-0

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*(2), 136-149. doi:10.1002/ab.20300

Lansford, J. E., Malone, P. S., Dodge, K. A., Crozier, J. C., Pettit, G. S., & Bates, J. E. (2006). A 12-year prospective study of patterns of social information processing problems and externalizing behaviors. *Journal of Abnormal Child Psychology*, *34*(5), 715-724. doi:10.1007/s10802-006-9057-4

Lansford, J. E., Malone, P. S., Dodge, K. A., Pettit, G. S., & Bates, J. E. (2010). Developmental cascades of peer rejection, social information processing biases, and aggression during middle childhood. *Developmental Psychopathology*, *22*, 593-602. doi:10.1017/S0954579410000301

Lochman, J. E., & Dodge, K. A. (1994). Social-cognitive processes of severely violent, moderately aggressive, and nonaggressive boys. *Journal of Consulting and Clinical Psychology, 62,* 366-374.

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*(1), 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

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*(2), 130-139. doi:10.1017/S1092852914000182

Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial boys*. Eugene, OR: Castalia Publishing.

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*(1), 66-75. doi:10.1080/15374410709336569

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review, 117*(3), 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*(3), 168-177. doi:10.1177/106342669900700305

Powers, C. J., Bierman, K. L., & The Conduct Problems Prevention Research Group. (2013). The multifaceted impact of peer relations on aggressive-disruptive behavior in early elementary school. *Developmental Psychology, 49*(6), 1174-1186. doi:10.1037/a0028400

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59-108. doi:10.1037/0033-295X.85.2.59

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873-922. doi:10.1162/neco.2008.12-06-420

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

Sukhodolsky, D. G., Golub, A., & Cromwell, E. N. (2001). Development and validation of the Anger Rumination Scale. *Personality and Individual Differences, 31*(5), 689-700. doi:10.1016/S0191-8869(00)00171-9

Sumner, S. A., Mercy, J. A., Dahlberg, L. L., Hillis, S. D., Klevens, J., & Houry, D. (2015). Violence in the United States: Status, challenges, and opportunities. *Journal of the American Medical Association, 314*(5), 478-488. doi:10.1001/jama.2015.8371

Vitaro, F., Barker, E. D., Boivin, M., Brendgen, M., & Tremblay, R. E. (2006). Do early difficult temperament and harsh parenting differentially predict reactive and proactive aggression? *Journal of Abnormal Child Psychology, 34*(5), 685-695. doi:10.1007/s10802-006-9055-6

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*(1), 3-21. doi:10.1177/1088868307309874

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.

Chapter 2: Study 1

Aggressive realism: More efficient processing of anger in physically aggressive individuals

Published in:

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

Abstract

Physically aggressive individuals' heightened tendency to decide that ambiguous faces are angry is thought to contribute to their destructive interpersonal behavior. Although this tendency is commonly attributed to bias, other cognitive processes could account for the emotion identification patterns observed in physical aggression. Diffusion modeling is a valuable tool for parsing the contributions of several cognitive processes known to influence decision-making, including bias, drift rate (efficiency of information accumulation), and threshold separation (extent of information accumulation). In a sample of 90 incarcerated men, we applied diffusion modeling to an emotion-identification task. Physical aggression was positively associated with drift rate (i.e., more efficient information accumulation) for anger, and drift rate mediated the association between physical aggression and heightened anger identification. Physical aggression was not, however, associated with bias or threshold separation. These findings implicate processing efficiency for anger-related information as a potential mechanism driving aberrant emotion identification in physical aggression.

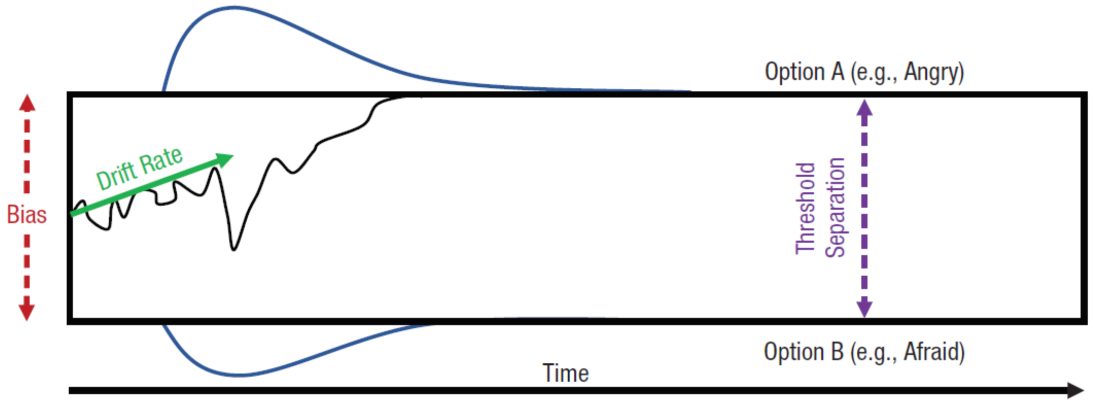
Introduction

Physical aggression is a harmful yet ubiquitous form of human behavior. Excessive physical aggression is associated with pervasive psychosocial impairments, including low-quality friendships, social rejection, marital discord, and involvement in the criminal justice system (Bierman & Wargo, 1995; Huesmann, Dubow, & Boxer, 2009; Poulin & Boivin, 1999). Decades of research suggest that physically aggressive behavior and its associated impairments arise, in part, from a pattern of interpreting social information in aberrant ways (Crick & Dodge, 1994).

One of the richest sources of social information is facial emotion (Marsh, Ambady, & Kleck, 2005). Substantial evidence indicates that physical aggression is associated with aberrant identification of facial emotions, most notably a heightened tendency to identify ambiguous faces as angry (Mellentin, Dervisevic, Stenager, Pilegaard, & Kirk, 2015; Schönenberg & Jusyte, 2014; Wilkowski & Robinson, 2012). This *anger perception bias*, a term that denotes impaired emotion identification and a preexisting inclination to make a particular interpretation irrespective of facial information, is theorized to drive physically aggressive behavior by fueling impressions of other individuals as hostile and threatening (Penton-Voak et al., 2013).

However, from a decision-making perspective, a response pattern observed at the behavioral level (i.e., a higher proportion of faces identified as angry) could arise from multiple cognitive processes, where "bias" is only one candidate (Ratcliff & McKoon, 2008). Because no research has looked beyond simple behavioral measures—for example, reaction time (RT) and accuracy—there has been no formal testing of the bias account, and the contributions of additional decision-making processes are unknown. Thus, although aberrant emotion identification in physical aggression is a reliable phenomenon, the underlying cognitive processes remain poorly understood.

Diffusion modeling is a form of computational modeling rooted in decision-making theory (Ratcliff, 1978) that can elucidate the cognitive processes involved in physically aggressive individuals' anger-identification patterns (Voss, Voss, & Lerche, 2015). Diffusion modeling is based on the premise that decisions are made by accumulating information until one of two response thresholds is reached, at which point the corresponding response is made (see Fig. 1). Within diffusion modeling, several processes could contribute to observed patterns of emotion identification in physical aggression. First, bias (the starting point of the decision-making process) could explain observed patterns if physically aggressive individuals require less information to identify faces as angry compared with other emotions (i.e., show a bias toward anger), predisposing them to identify faces as angry in a stimulus-nondependent manner. Second, drift rate (the rate at which information is accumulated) could explain observed patterns if physically aggressive individuals accumulate anger-related information more efficiently (i.e., show a higher drift rate for anger), leading them to identify faces as angry more swiftly without a decrement in accuracy. Finally, threshold separation (the amount of information accumulated for a decision) could contribute to observed patterns if physically aggressive individuals accumulate less information when identifying facial emotions (i.e., exhibit lower threshold separation), speeding up their responses and reducing their accuracy. Finally, any combination of these factors (e.g., lower threshold separation plus a bias toward anger) could result in an even greater likelihood of identifying faces as angry.



Parameter	Definition	Interpretation of Values
Bias	Decision Starting Point	Values Deviating From 0.5 Reflect a Bias Toward Option A (if > 0.5) or Toward Option B (if < 0.5)
Drift Rate	Average Slope of the Information-Accumulation Process	Larger Values Indicate More Efficient Processing
Threshold Separation	Amount of Information Accumulated for a Decision	Smaller Values Reflect a More Impulsive Decision-Making Style (i.e., Quicker, Less Accurate)

Fig. 1. Schematic representation of the diffusion model of decision-making. The decision process begins at a starting point that may represent an a priori response bias toward either option A or option B. Information is accumulated in favor of either option A or option B, and drift rate represents the average rate of information accumulation. The amount of information accumulated for a decision is represented by threshold separation, the distance between the two option thresholds. The “noisy” black line represents the information-accumulation process, whereas the blue curved lines above option A and below option B represent the reaction time distribution associated with each response. Diffusion modeling parameter estimation is based on these reaction time distributions. Further information about each of the parameters shown here is given below the graph. An additional parameter estimated by diffusion modeling (not directly examined in the present study and not illustrated here) is nondecision time, which represents the length of time taken for nondecision-related processes (e.g., encoding, motor execution).

In addition to unspecified contributions of various cognitive processes, the impact of contextual factors (e.g., apparent motion, background scene) on emotion identification in physical aggression is unknown. Previous research indicates that contextual factors influence aggressive individuals’ interpretations of social information more broadly. For example, aggressive individuals are more likely to make hostile interpretations of another’s actions under conditions of threat (Dodge & Somberg, 1987). However, the impact of contextual factors on physically aggressive individuals’ emotion identification has not been examined. It is possible that contextual threat amplifies physically aggressive individuals’ aberrant emotion identification via altered cognitive processes (e.g., more efficient anger processing under threat, more impulsive responding under threat). Thus, in addition to quantifying cognitive processes underlying emotion identification, it is important to examine potential contributions of context to aberrant emotion identification and related cognitive processes in physical aggression.

The primary aim of the present study was to apply diffusion modeling to estimate bias, drift rate, and threshold separation during a facial emotion-identification task to test whether any of these cognitive processes could account for the association between physical aggression and aberrant emotion identification (i.e., heightened anger identification). On the basis of evidence linking physical aggression to heightened anger identification (Mellentin et al., 2015; Schönenberg & Jusyte, 2014), we hypothesized that physical aggression would be associated with a higher likelihood of identifying faces as angry (Hypothesis 1).

Additionally, despite pervasive characterizations of emotion processing as “biased” in aggression, recent findings suggest that physically aggressive individuals possess superior anger identification abilities, exemplified by a heightened ability to discriminate between faces displaying different degrees of anger and an advanced capacity for extracting anger-related information from ambiguous faces (Wilkowski & Robinson, 2012). Rather than displaying signs of bias (i.e., showing a stimulus-nondependent tendency to identify faces as angry) or low threshold separation (i.e., responding less accurately), physically aggressive individuals display anger-identification patterns that may be most consistent with higher drift rate for anger because they appear to more efficiently and effectively accumulate information from subtle anger cues. Thus, we hypothesized that physical aggression would be associated with higher drift rate for anger (Hypothesis 2). Further, we hypothesized that drift rate would mediate the association between physical aggression and heightened anger identification (Hypothesis 3).

Finally, following research indicating that rapidly encroaching stimuli are perceived as more threatening (Coker-Appiah et al., 2013; Vieira, Tavares, Marsh, & Mitchell, 2017), we manipulated apparent movement of faces (by presenting looming or receding faces) to examine influences of contextual threat. We hypothesized that physical aggression would be associated with a higher likelihood of identifying looming (i.e., threatening) faces as angry (Hypothesis 4). Additionally, on the basis of evidence that aggression is associated with hyperreactivity to threat (Coccaro, McCloskey, Fitzgerald, & Phan, 2007; da Cunha-Bang et al., 2017) and impulsive decision-making under threat (Brennan & Baskin-Sommers, 2019; Verona & Bozzay, 2017), we hypothesized that physical aggression would be associated with lower threshold separation (i.e., greater impulsivity) under this condition (Hypothesis 5).

Method

Participants

Participants were men from a high-security correctional institution in Connecticut (for sample characteristics and correlations among key study variables, see Table S1 in the Supplemental Material available online); 96.94% of participants had been charged with a violent crime in their lifetime, and 56.12% had been charged with a violent institutional infraction while incarcerated (i.e., violations against persons, including fighting and assault on correctional staff). Because physical aggression is more pronounced in men compared with women, and because 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.

Prior to recruitment, study personnel received an institutional roster of inmates. Study personnel used this roster to review medical files and exclude individuals who had a history of psychosis or bipolar disorder, currently had mood or anxiety disorders, currently used psychotropic medication, had a family history of psychosis, had 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), had an IQ below 70, or had a reading level below fourth grade.

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 institutional or 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 Connecticut Department of Correction regulations, participants did not receive financial compensation. After providing consent, participants completed an initial session that involved a series of clinical and neuropsychological assessments. Participants who did not meet eligibility thresholds (detailed above) on any of these assessments were excluded from further participation. Eligible participants returned for a second session in which they completed the task followed by the aggression and emotional experience measures (see Measures section below). Both in-person sessions took place in a private testing space within the prison.

An a priori power analysis based on published studies on related topics (i.e., individual differences in facial emotion identification; Wilkowski & Robinson, 2012) indicated that a sample size of approximately 90 participants would be sufficient to detect moderate effects with

80% power. To ensure sufficient power to account for the normative loss of data due to invalid task performance, we collected data from 98 participants.

Measures

Buss-Perry Aggression Questionnaire (BPAQ). The BPAQ (Buss & Perry, 1992) is a 29-item self-report measure of aggression. Participants rate each item on a 5-point Likert-type 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 (9 items), Verbal Aggression (5 items), Anger (7 items), and Hostility (8 items). The BPAQ is a reliable, valid, and widely used measure of aggression (Harris, 1997; Tremblay & Ewart, 2005) and shows evidence of adequate reliability and validity in incarcerated samples (Archer & Haigh, 1997; Ireland & Archer, 2004). In the present study, we used the BPAQ Physical Aggression subscale as the measure of physical aggression, and our hypotheses centered on physical aggression on the basis of previous research (e.g., Wilkowski & Robinson, 2012). However, aggression is a multifaceted construct that can be conceptualized as having behavioral (i.e., Physical Aggression, Verbal Aggression), affective (i.e., Anger), and attitudinal (i.e., Hostility) components (Buss & Perry, 1992). Therefore, we also examined associations between these other aggression-related constructs and task performance (for analyses with the other BPAQ subscales and BPAQ Total score, see the Supplemental Material). Scores for the Physical Aggression subscale can range from 5 to 45, with higher scores indicating higher levels of physical aggressiveness. Internal consistencies for the Physical Aggression subscale (Cronbach's α = .79) and the BPAQ as a whole (Cronbach's α = .88) in the present sample were acceptable and comparable with reliability coefficients reported by Buss and Perry (1992).

Range and Differentiation of Emotional Experience Scale (RDEES). The RDEES (Kang & Shaver, 2004) is a 14-item self-report measure of the extent to which one's emotional experiences are broad in range and well differentiated. Participants rate each item on a 5-point Likert-type scale (1 = *does not describe me very well*, 5 = *describes me very well*). The measure consists of two subscales: Range (7 items; sample item: "I experience a wide range of emotions") and Differentiation (7 items; sample item: "I am aware that each emotion has a completely different meaning"). Scores on each subscale can range from 7 to 35, with higher scores indicating greater range and differentiation of emotional experiences, respectively. Internal consistencies for the both the Range and Differentiation subscales (Cronbach's αs = .67 and .80, respectively) in the present sample were acceptable. Following previous research indicating that emotion differentiation is associated with emotion-identification accuracy (Israelashvili, Oosterwijk, Sauter, & Fischer, 2019), we evaluated the validity of the task using both the Differentiation (convergent validity) and Range (divergent validity) subscales.

Ambiguous emotion-identification task. Participants completed a two-alternative forced-choice task in which they identified the emotion displayed in a series of ambiguous emotional faces.

Stimuli. Stimuli consisted of emotional face images from the Racially Diverse Affective Expression (RADIATE) face stimulus set (publicly available at http://fablab.yale.edu/page/assays-tools; Conley et al., 2018; Tottenham et al., 2009). Images of 39 unique male models of three racial/ethnic backgrounds (Black, White, and Hispanic) displaying anger, fear, and happiness were selected from the RADIATE set. The racial/ethnic composition of the face stimuli (i.e., 38.46% Black, 33.33% White, 28.21% Hispanic) roughly mirrored that found in our sample. Stimuli were generated by blending two images using face morphing software (Abrosoft, 2018, Fantamorph Deluxe for Mac, Version 5.5.0) to create 70%–30% blends. The 70%–30% level of blending was chosen to achieve a moderate level of ambiguity (Schönenberg & Jusyte, 2014) and elicit variable but sufficiently high accuracy levels to provide data suitable for diffusion modeling (Ratcliff & McKoon, 2008). Three types of emotion blends were created: anger–fear blends, anger–happiness blends, and fear–happiness blends. We chose anger, fear, and happiness to maximize consistency with previous studies that examined emotion identification in physical aggression—outside of anger (the primary emotion of interest in the present study), fear and happiness are the most frequently used negative and positive emotions, respectively (e.g., Schönenberg & Jusyte, 2014; Wilkowski & Robinson, 2012). Within each blend, one of the two emotions served as the dominant emotion. In total, six blends per model were created (3 emotion blend types × 2 dominant emotion types; see Fig. 2). The process of generating six different image types for each model resulted in 234 unique images.

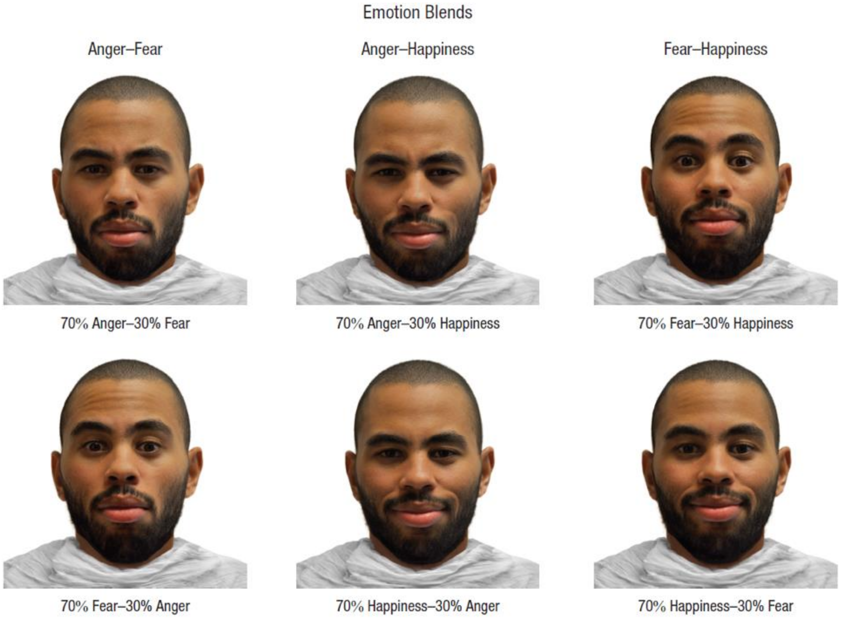


Fig. 2. Sample task stimuli. Stimuli displayed blends of anger–fear (left column), anger–happiness (middle column), and fear–happiness (right column). Within each blend type, one of the two emotions was the dominant (i.e., 70%) emotion.

The task consisted of three separate blocks: an anger–fear block, an anger–happiness block, and a fear–happiness block. Within each block, faces displayed blends of only two emotions. For example, in the anger–fear block, all faces displayed a blend of anger and fear. Because the study aims and hypotheses revolved around anger, the anger–fear and anger–happiness blocks were the primary blocks of interest; however, we included a fear–happiness block in order to examine identification patterns and decision-making parameters for fear and happiness outside of the context of anger. Within each block, half of the faces displayed mostly one emotion (e.g., 70% anger–30% fear), and half of the faces displayed mostly the other emotion (e.g., 70% fear–30% anger). Ordering of blocks was counterbalanced. Furthermore, within each block, half of the faces appeared to loom (i.e., move toward the participant), and half of the faces appeared to recede (i.e., move away from the participant). Each block consisted of 156 trials (39 unique faces × 2 dominant emotion types × 2 movement types) for a total of 468 trials in the task.

Task procedure. Participants were seated approximately 60 cm away from a 27-in. BenQ high-performance LED gaming monitor (Model XL2720Z; BenQ, Taipei, Taiwan). Participants were instructed to identify the emotion expressed in each face as quickly and accurately as possible using two keys on the keyboard. At the beginning of each block, one key was assigned to one of the two emotions represented in the faces, and another key was assigned to the other of the two emotions represented in the faces. Participants were told to press the left shift key to identify the face as one of the two emotions for that block and to press the right shift key to identify the face as the other emotion. Keyboard covers with corresponding labels were placed over the keyboard in each block to aid the participant in key–response mappings. Key–response mappings were counterbalanced across participants to counteract any effects of assigning a particular response option to either the dominant or non-dominant hand. Before each block began, participants completed 10 practice trials in which they pressed the corresponding key for the emotion word (e.g., “angry” or “afraid” prior to the anger–fear block) that appeared on the screen. To proceed to the next practice trial (and ultimately to the main task), participants were required to press the correct key on each practice trial (and were given multiple chances if needed).

Stimulus presentation and response collection were controlled using the Psychophysics Toolbox extension (Version 3; Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) as implemented in MATLAB 2017b (The MathWorks, Natick, MA). Stimuli were presented in random order for each participant. Each trial began with a fixation cross (500 ms), after which a face was displayed on the screen for a total of 1,520 ms. Following previous research (Vieira et al., 2017), we created movement effects by rapidly changing the visual angle of stimuli. Faces increased (on looming trials) or decreased (on receding trials) in size by a factor of 1.05, resulting in 19 frames (each lasting 80 ms) per trial (see Fig. 3). The intertrial interval varied randomly between 1,000 and 2,000 ms (average 1,500 ms).

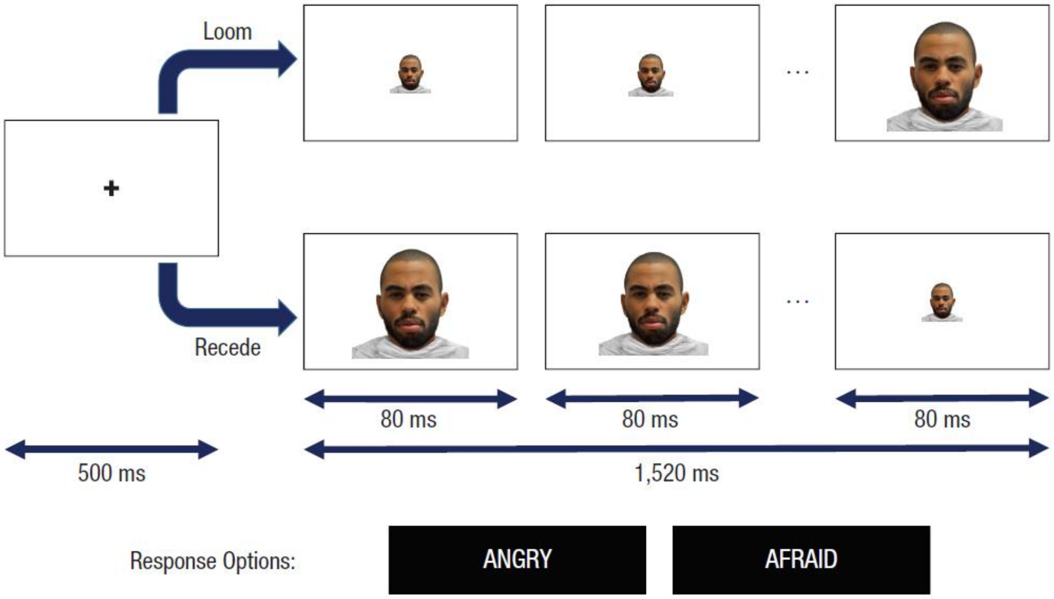


Fig. 3. Schematic representation of trial layout and timing in the ambiguous facial emotion-identification task. On each trial, participants viewed a serial presentation of images that either increased in visual angle (i.e., a looming trial, depicted it the upper row of images) or decreased in visual angle (i.e., a receding trial, depicted in the lower row of images). Shown here are the first frame, second frame, and last frame (of 19 total frames on each trial) for each trial type. Participants pressed one of two keys to identify the emotion displayed in the face.

Control emotion-identification task. After completing the main task, participants completed a control emotion-identification task, which assessed general emotion-identification accuracy as a way to examine the validity of the ambiguous emotion-identification task. Participants were instructed to choose the emotion displayed by each face that appeared on the screen. Stimuli in the control task consisted of unblended emotional face images (i.e., 100% display of one emotion) from the RADIATE stimulus set. The emotions displayed in the images were the same as those used in the main task: anger, fear, and happiness. Participants were given three response options (one for each emotion), which appeared as text in three separate panels below each face on the screen, and they used a mouse to click the panel corresponding to the emotion that each face displayed. There was no time limit imposed for responding. The control task consisted of 54 trials.

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 conditions was met: (a) no response given (or response given in < 300 ms) on more than 20% of trials, (b) accuracy at or below chance (i.e., ≤ 50%), or (c) statistical outliers (> 3 *SD*s from the mean) on task behavioral variables. Seven participants were excluded from analyses on the basis of these criteria, and the resulting sample consisted of 91 participants.

Task validation. Convergent validity of the ambiguous emotion-identification task was evaluated by examining associations between the RDEES Differentiation score and overall task accuracy, as well as between control task accuracy and overall task accuracy. Divergent validity of the task was evaluated by examining the association between RDEES Range score and overall task accuracy. On the one hand, we expected RDEES Differentiation (one's ability to identify subtle variations in emotional experiences) to be positively related to task accuracy (i.e., an enhanced ability to correctly identify the dominant emotion in ambiguous emotional faces; Israelashvili et al., 2019), and we also expected control task accuracy to be positively related to task accuracy, since the control task had similar demands but no stimulus movement effects and no time limit for responses. On the other hand, we did not expect RDEES Range (one's own experience of a range of different emotions) to be related to task accuracy.

A 3 (emotion blend: anger–fear, anger–happiness, fear–happiness) × 2 (movement: looming, receding) repeated measures general linear model (GLM), with RDEES Differentiation (*z* scored) as a continuous between-subjects independent variable and overall task accuracy as a dependent variable, revealed a main effect of differentiation on accuracy, $F(1,88) = 4.19$, $p = .044$, $\eta_p^2 = 0.05$, 90% confidence interval (CI) = [0.001, 0.13]; individuals with higher levels of differentiation exhibited higher emotion-identification accuracy overall ($b = 3.61$, $SE = 1.76$, 95% CI = [0.10, 7.11]), providing support for convergent validity of the task. Furthermore, we detected a moderately strong correlation between control task accuracy and overall task accuracy, $r(88) = .49$, providing additional support for convergent validity of the task (see Table S1 in the Supplemental Material). A 3 (emotion blend: anger–fear, anger–happiness, fear–happiness) × 2 (movement: looming, receding) repeated measures GLM, with RDEES Range (*z* scored) as a continuous between-subjects independent variable and overall task accuracy as a dependent variable, revealed no associations between RDEES Range and overall task accuracy, providing support for divergent validity of the task.

Diffusion modeling. Following established guidelines (Voss et al., 2015), we removed trials with no response (i.e., omissions) and trials with RTs less than 300 ms (i.e., premature responses) from individual participants' data before subjecting them to diffusion modeling. Rates of omissions and premature responses were low (i.e., omissions characterized, on average, 2.51% of trials per participant, and premature responses characterized 0.24% of trials).

We used *fast-dm-30* software (Voss & Voss, 2007; Voss et al., 2015) to estimate decision-making parameters on the basis of response and RT data from the task. The software was designed to estimate parameters from Ratcliff's (1978) diffusion model, in which decision-making is a noisy, continuous process of accumulating information until one of two decision thresholds (one for each response option) is met. The model uses RT distributions for the two response options to estimate decision-making parameters, including bias, threshold separation, and drift rate. By using the entire range of task performance across trials (rather than simple accuracy or mean RTs in isolation), diffusion modeling delivers several advantages over traditional methods, including increased reliability and the potential to yield novel mechanistic insights (Price, Brown, & Siegle, 2019). The Kolmogorov-Smirnov estimation procedure was used because it accounts for exact RT distributions (as opposed to binning RT data) and is robust to contaminants. Guided by theoretical and methodological considerations, we allowed the relative starting point to vary by emotion blend (i.e., block) only (because starting point is not impacted by stimulus features), and we allowed both threshold separation and drift rate to vary by all three conditions (i.e., emotion blend, dominant emotion, and movement). "Angry" responses were set as response option A, whereas non-"angry" responses were set as response option B (see Fig. 1), so that positive bias-parameter values would indicate a bias toward anger, and positive drift-rate values would indicate a drift rate toward anger (and conversely, negative values would indicate a bias toward the nonanger emotion—i.e., happiness or fear, depending on the block—and drift rate toward the nonanger emotion, respectively). Parameter values for threshold separation, by contrast, are not directional, and they typically range from 0.5 to 2.0. To maximize parsimony and accuracy of the model, we opted for a four-parameter model, in which our three parameters of interest plus nondecision time were allowed to vary, whereas the remaining parameters were fixed at 0 (Lerche & Voss, 2016; for correlations among the diffusion-modeling parameters, see Table S2 in the Supplemental Material).

Here, we provide a simplified illustration of how various patterns of RT distributions impact the diffusion-modeling parameter estimates. In a two-choice task, each response option (e.g., option A and option B in Fig. 1) has an RT distribution that is determined by the frequency of different RTs for that response across all trials within a given task condition. The frequency of RTs for one response determines the RT distribution for that response option (e.g., the blue curved line above option A in Fig. 1), whereas the frequency of RTs for the other response determines the RT distribution for that response option (e.g., the blue curved line below option B in Fig. 1). For example, if a participant has a large number of fast RTs and virtually no slow RTs when making one response, the RT distribution for that response option will be compressed toward the left (i.e., fast) end of the distribution. In terms of how various RT distribution patterns translate into diffusion-modeling parameter estimates, let us consider which RT distribution patterns would correspond to a bias toward one response option, higher drift rate toward one response option, and lower threshold separation. For bias, if the leading end of the RT distribution for option A is compressed toward the left without a corresponding compression in the RT distribution for option B, then diffusion modeling estimates a higher value for the bias parameter (i.e., greater than 0.5), indicating a bias toward option A. For drift rate, if there is an increased relative probability of faster RTs for option A (i.e., the RT distribution is taller for option A), then diffusion modeling estimates a higher value for the drift-rate parameter, indicating stronger drift rate toward option A. Finally, for threshold separation, if the RT distributions for both response options are compressed toward the left, then diffusion modeling estimates a lower value for the threshold separation parameter. Although both bias and lowered threshold separation are associated with reduced performance on a task, each parameter relates to performance in a distinct manner. Whereas a bias toward one response option would preferentially increase the likelihood of that response, lowered threshold separation, which denotes the extent of information accumulation for both response options, would decrease accuracy in general but not in favor of either response option.

Following parameter estimation, model fit was assessed using Kolmogorov-Smirnov test statistics (values > .05 generally indicate acceptable fit), along with visual inspection of quantile-

quantile (Q-Q) plots (which indicate acceptable fit if all data points lie near the main diagonal). These indices revealed that the model generally fitted the data well. On the basis of visual inspection, we deemed that one participant's data fitted poorly to the model, and this participant was excluded. Thus, the final sample consisted of 90 participants. Excluded participants did not differ from included participants in terms of physical aggression (95% CI for the mean difference = [-5.40, 5.78], *p* = .905).

The study protocol was approved by the Yale University Human Investigation Committee and was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. In this article, we report all of the dependent measures collected, all data exclusions, and all of the task conditions. This study was not preregistered. 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, requests for a deidentified subset of the data can e-mailed to the corresponding author.

Results

Emotion identification

Given previous research highlighting heightened anger identification in physical aggression, we started by conducting a 2 (emotion blend: anger–fear, anger–happiness) × 2 (dominant emotion: anger, non-anger) × 2 (movement: looming, receding) repeated measures GLM with BPAQ Physical Aggression (*z* scored) as a continuous between-subjects independent variable and the proportion of trials on which participants responded "angry" (i.e., anger identification) as a dependent variable (for additional analyses pertaining to robustness of results and specificity to physical aggression, see the Supplemental Material). The analysis revealed both task effects and physical-aggression-related effects.

In terms of task effects, there was a main effect of dominant emotion on anger identification, *F*(1,88) = 1845.92, $p < .001$, $\eta_p^2 = .95$, 90% CI = [.94, .96]; mostly angry faces were more likely to be identified as angry (*M* = 73.7%, 95% CI = [71.8%, 75.6%]) compared with mostly nonangry faces (*M* = 21.6%, 95% CI = [20.0%, 23.2%]). This main effect provides a key demonstration of task validity by indicating that participants were able to discriminate between the two types of faces and identify the dominant emotion.

There was also a main effect of movement on anger identification, *F*(1,88) = 5.16, $p = .025$, $\eta_p^2 = .10$, 90% CI = [.004, .15]; looming faces were more likely to be identified as angry (*M* = 48.2%, 95% CI = [46.8%, 49.6%]) compared with receding faces (*M* = 47.1%, 95% CI = [45.8%, 48.5%]). This effect provides support for the success of the movement manipulation and suggests that looming faces were perceived as more threatening. Additionally, there was an Emotion Blend × Movement interaction, *F*(1,88) = 9.78, $p = .002$, $\eta_p^2 = .06$, 90% CI = [.02, .20]; looming faces were more likely to be identified as angry, and this was particularly true for the anger–fear blended faces (*M* = 49.1% for looming faces, 95% CI = [47.2%, 51.0%]; *M* = 46.4% for receding faces, 95% CI = [44.3%, 48.6%]) compared with the anger–happiness blended faces (*M* = 47.3% for looming faces, 95% CI = [45.6%, 49.0%]; *M* = 47.8% for receding faces, 95% CI = [46.3%, 49.4%]). An Emotion Blend × Dominant Emotion interaction emerged as well, *F*(1,88) = 293.81, $p < .001$, $\eta_p^2 = .77$, 90% CI = [.70, .81], indicating that the difference in anger identification as a function of the dominant emotion in the face was greater for the anger–happiness blended faces (*M* = 79.5% for mostly angry faces, 95% CI = [77.7%, 81.3%]; *M* = 15.6% for mostly nonangry faces, 95% CI = [13.7%, 17.6%]) compared with the anger–fear blended faces (*M* = 67.9% for mostly angry faces, 95% CI = [65.0%, 70.8%]; *M* = 27.6% for mostly nonangry faces, 95% CI = [25.6%, 29.6%]), suggesting that participants had greater difficulty accurately identifying the dominant emotion for faces that displayed a blend of anger and fear. This finding indicates that the anger–fear blended faces were significantly more ambiguous than the anger–happiness blended faces.

In terms of physical-aggression-related effects, there was an Emotion Blend × Physical Aggression interaction, $F(1,88) = 4.07$, $p = .047$, $\eta_p^2 = .04$, 90% CI = [.0003, .13]; physical aggression was positively related to the proportion of “angry” responses for the anger–fear blended faces ($b = 0.02$, $SE = 0.01$, 90% CI = [0.01, 0.04], $p = .025$, $\eta_p^2 = .06$), but no association was detected for the anger–happiness blended faces ($b = -0.01$, $SE = 0.01$, 90% CI = [-0.01, 0.01] $p = .873$, $\eta_p^2 = .00$; see Fig. 4). Results remained unchanged after we controlled for participant race. Thus, the results were consistent with Hypothesis 1: Higher levels of physical aggression were associated with a heightened tendency to identify ambiguous faces as angry. However, we failed to find support for Hypothesis 4, as there was no interaction involving movement and physical aggression.

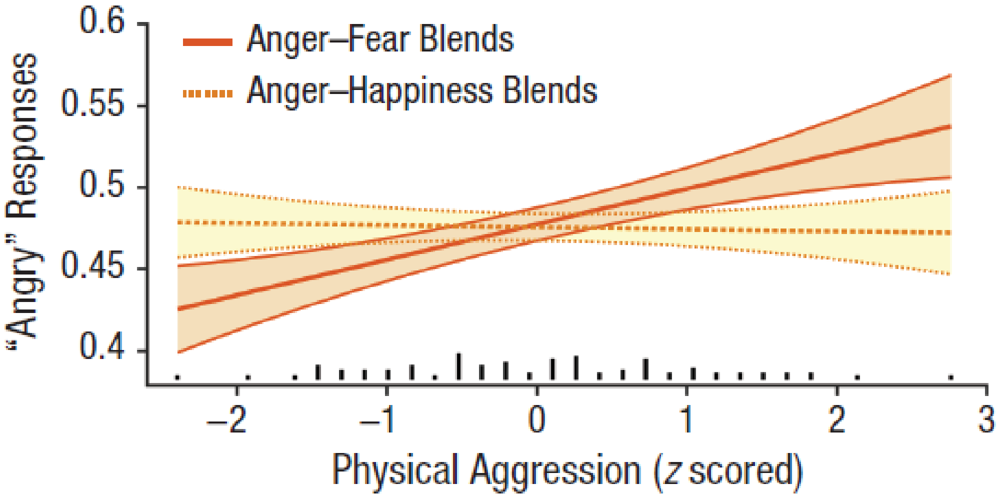


Fig. 4. The relationship between physical aggression and proportion of “angry” responses for anger–fear blended faces and anger–happiness blended faces. Error bands represent ± 1 *SE*, and the dot plot along the *x*-axis represents frequencies for Buss-Perry Aggression Questionnaire Physical Aggression scores.

Because physical aggression was associated with heightened anger identification for the anger–fear blended faces (but not the anger–happiness blended faces), it is possible that this association was confounded by fear identification. That is, if individuals with higher levels of physical aggression were generally less likely to identify faces as afraid, this could have accounted for their heightened tendency to identify faces as angry when they were presented with two response options: angry or afraid. To rule out fear identification as a potential confound of the association between physical aggression and anger identification, we analyzed fear identification outside of the context of anger (i.e., in the fear–happiness block) by conducting a 2 (dominant emotion: fear, happiness) × 2 (movement: looming, receding) repeated measures GLM with BPAQ Physical Aggression (*z* scored) as a continuous between-subjects independent variable and proportion of "afraid" responses as a dependent variable. The GLM revealed a significant main effect of dominant emotion on fear identification, $F(1,88) = 2940.83$, $p < .001$, $\eta_p^2 = .97$, 90% CI = [.96, .98]; mostly afraid faces were more likely to be identified as afraid ($M = 83.3\%$, 95% CI = [81.6%, 85.0%]) compared with mostly happy faces ($M = 19.1\%$, 95% CI = [16.9%, 21.2%]). There were no other task effects and, crucially, no physical-aggression-related effects associated with fear identification. Most notably, we failed to detect a main effect of physical aggression on fear identification, $F(1,88) = 0.07$, $p = .786$, $\eta_p^2 = .001$, 90% CI = [.00, .03]. The failure to detect physical-aggression-related effects associated with fear identification suggests that the association between physical aggression and heightened anger identification for anger–fear blended faces is not attributable to a diminished tendency to identify fear in faces.

Diffusion modeling parameters

To examine decision-making parameters estimated with diffusion modeling as a function of task conditions as well as physical aggression, we conducted a series of 2 (emotion blend:

anger–fear, anger–happiness) × 2 (dominant emotion: anger, nonanger) × 2 (movement: looming, receding) repeated measures GLMs with BPAQ Physical Aggression (*z* scored) as a continuous between-subjects independent variable and each diffusion modeling parameter as a dependent variable.

Bias. There were no task effects and no physical aggression-related effects associated with bias (all *p*s ≥ .352).

Drift rate. Examination of drift rate as a dependent variable revealed both task effects and physical aggression-related effects. In terms of task effects, there was a main effect of dominant emotion on drift rate, *F*(1,88) = 1161.87, *p* < .001, η_p^2 = .93, 90% CI = [.91, .94], indicating that drift rate toward anger was higher for mostly angry faces (*M* = 0.89, 95% CI = [0.82, 0.96]) compared with mostly nonangry faces (*M* = -0.94, 95% CI = [-1.00, -0.87]). There was also an Emotion Blend × Movement interaction, *F*(1,88) = 8.67, *p* = .004, η_p^2 = .09, 90% CI = [.02, .19], indicating that drift rate toward anger was higher for looming faces, but only for the anger–fear blended faces (*M* = 0.03 for looming faces, 95% CI = [-0.05, 0.10]; *M* = -0.07 for receding faces, 95% CI = [-0.15, 0.01]) and not for the anger–happiness blended faces (*M* = -0.05 for looming faces, 95% CI = [-0.13, 0.02]; *M* = 0.00 for receding faces, 95% CI = [-0.07, 0.07]). Additionally, there was an Emotion Blend × Dominant Emotion interaction, *F*(1,88) = 262.18, *p* < .001, η_p^2 = .75, 90% CI = [.67, .80], indicating that the difference between drift rates for mostly angry and mostly nonangry faces was greater for the anger–happiness blended faces (*M* = 1.14 for mostly angry faces, 95% CI = [1.07, 1.21]; *M* = -1.19 for mostly nonangry faces, 95% CI = [-1.28, -1.10]) compared with the anger–fear blended faces (*M* = 0.64 for mostly angry faces, 95% CI = [0.54, 0.74]; *M* = -0.68 for mostly nonangry faces, 95% CI = [-0.76, -0.60]). The fact that drift rate was more strongly differentiated according to dominant emotion for the anger–

happiness faces (i.e., information accumulation proceeded more efficiently) is again consistent with the idea noted above that the anger–fear blended faces were significantly more ambiguous than the anger–happiness blended faces.

The final task effect was a Movement × Dominant Emotion interaction, *F*(1,88) = 9.05, *p* = .003, $\eta_p{}^2$ = .09, 90% CI = [.02, .20], indicating that the difference between drift rates for mostly angry and mostly nonangry faces was greater for receding faces (*M* = 0.91 for mostly angry faces, 95% CI = [0.84, 0.99]; *M* = -0.98 for mostly nonangry faces, 95% CI = [-1.05, -0.90]) compared with looming faces (*M* = 0.87 for mostly angry faces, 95% CI = [0.80, 0.94]; *M* = -0.89 for mostly nonangry faces, 95% CI = [-0.96, -0.82]).

In terms of physical-aggression-related effects, there was an Emotion Blend × Physical Aggression interaction, *F*(1,88) = 5.32, *p* = .023, $\eta_p{}^2$ = .06, 90% CI = [.004, .15]; physical aggression was positively related to drift rate for the anger–fear blended faces (*b* = 0.11, *SE* = 0.04, 90% CI = [0.04, 0.18], *p* = .014, $\eta_p{}^2$ = .07), but no association was detected for the anger–happiness blended faces (*b* = -0.02, *SE* = 0.04, 90% CI = [-0.09, 0.04], *p* = .599, $\eta_p{}^2$ = .00; see Fig. 5). Results remained unchanged after we controlled for participant race. Thus, consistent with Hypothesis 2, higher levels of physical aggression were associated with higher drift rate for anger.

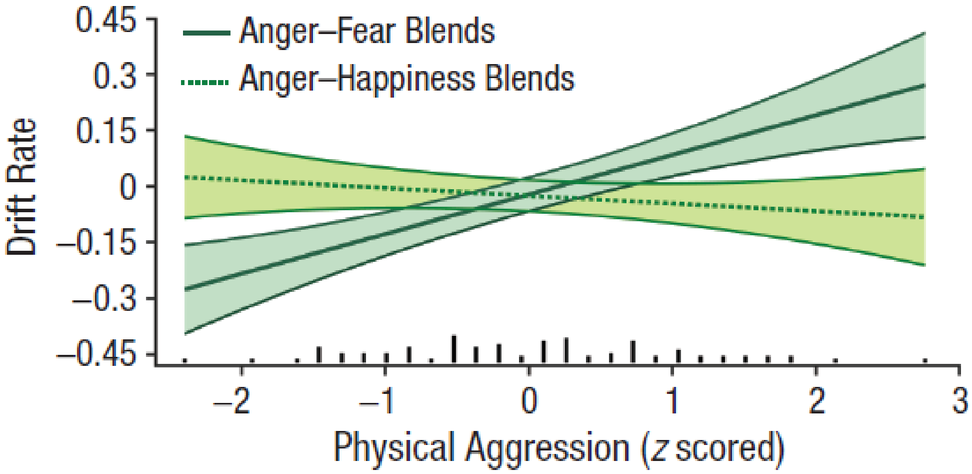


Fig. 5. The relationship between physical aggression and drift rate toward anger for anger–fear blended faces and anger–happiness blended faces. Error bands represent ± 1 *SE*, and the dot plot along the *x*-axis represents frequencies for Buss-Perry Aggression Questionnaire Physical Aggression scores.

Because physical aggression was associated with higher drift rate for anger when identifying the anger–fear blended faces (but not the anger–happiness blended faces), it is possible that this association was confounded by drift rate for fear. That is, if individuals with higher levels of physical aggression had a lower drift rate for fear in general, this could have accounted for their higher drift rate for anger when they were presented with two response options: angry or afraid. To rule out drift rate for fear as a potential confound of the association between physical aggression and drift rate for anger, we analyzed drift rate for fear outside of the context of anger (i.e., in the fear–happiness block) by conducting a 2 (dominant emotion: fear, happiness) × 2 (movement: looming, receding) repeated measures GLM with BPAQ Physical Aggression (*z* scored) as a continuous between-subjects independent variable and drift rate for fear as a dependent variable. The GLM revealed a significant main effect of dominant emotion on drift rate for fear, $F(1,88) = 1426.22$, $p < .001$, $\eta_p^2 = .94$, 90% CI = [.92, .95]; drift rate for fear was higher for mostly afraid faces ($M = 1.38$, 95% CI = [1.29, 1.47]) compared with mostly happy faces (*M* = -1.07, 95% CI = [-1.18, -0.96]). There were no other task effects and, crucially, no physical-aggression-related effects associated with drift rate for fear. Most notably, we failed to detect a main effect of physical aggression on drift rate for fear, $F(1,88) = 0.14$, $p = .714$, $\eta_p^2 = .00$, 90% CI = [.00, .04]. The failure to detect physical-aggression-related effects associated with drift rate for fear suggests that the association between physical aggression and higher drift rate for anger when identifying anger–fear blended faces is not attributable to less efficient processing of fear-related information.

Threshold separation. Examination of threshold separation as a dependent variable revealed task effects but no physical-aggression-related effects (indicating a failure to find support for Hypothesis 5). In terms of task effects, there was a main effect of emotion blend, $F(1,88) = 34.57$, $p < .001$, $\eta_p^2 = .28$, 90% CI = [.16, .39], indicating that threshold separation was greater for the anger–happiness blended faces (*M* = 1.25, 95% CI = [1.23, 1.27]) compared with the anger–fear blended faces (*M* = 1.19, 95% CI = [1.18, 1.21]). There was also a main effect of movement, $F(1,88) = 16.82$, $p < .001$, $\eta_p^2 = .16$, 90% CI = [.06, .27], indicating that threshold separation was lower for looming faces (*M* = 1.20, 95% CI = [1.19, 1.22]) compared with receding faces (*M* = 1.24, 95% CI = [1.22, 1.26]). Finally, there was an Emotion Blend × Dominant Emotion interaction, $F(1,88) = 6.11$, $p = .015$, $\eta_p^2 = .07$, 90% CI = [.01, .16], indicating that threshold separation was lower for mostly angry faces but only for the anger–happiness blended faces (*M* = 1.24 for mostly angry faces, 95% CI = [1.21, 1.26]; *M* = 1.26 for mostly nonangry faces, 95% CI = [1.24, 1.29]) and not for the anger–fear blended faces (*M* = 1.21 for mostly angry faces, 95% CI = [1.19, 1.23]; *M* = 1.18 for mostly nonangry faces, 95% CI = [1.16, 1.20]).

Mediation model

Given that the goal of the present study was to identify potential mechanisms supporting the link between physical aggression and heightened anger identification, we conducted a mediation analysis with BPAQ Physical Aggression as the independent variable, proportion of "angry" responses for the anger–fear blended faces as the dependent variable, and drift rate toward anger for the anger–fear blended faces as the mediator (see Fig. 6). The analysis was performed using the PROCESS macro for SPSS (Hayes, 2018), Model 4. We used a nonparametric resampling procedure (bootstrapping) with 5,000 samples to estimate the indirect effect. The analysis indicated a significant indirect effect of physical aggression on anger identification through drift rate (*b* = 0.004, *SE* = 0.002, 95% CI = [0.001, 0.007]). Thus, drift rate mediated the association between physical aggression and heightened anger identification, consistent with Hypothesis 3.

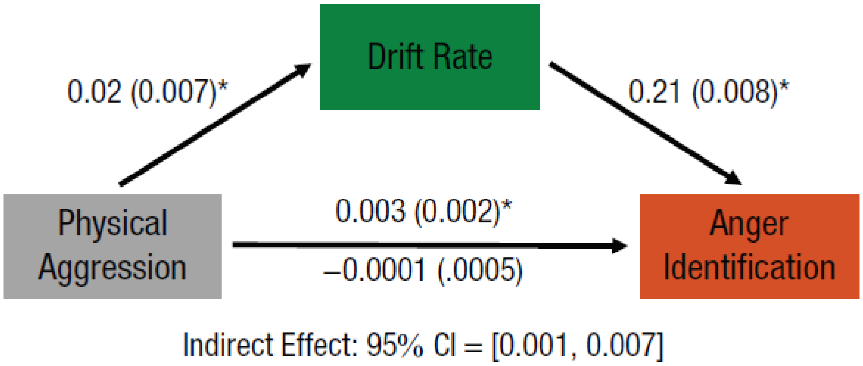


Fig. 6. Mediation model showing the influence of physical aggression on anger identification (i.e., proportion of "angry" responses) for the anger–fear blended faces, as mediated by drift rate toward anger. On the path from physical aggression to anger identification, the value above the arrow shows the total effect, and the value below the arrow shows the direct effect after controlling for the mediator. Unstandardized coefficients are shown, with standard errors in parentheses. The confidence interval (CI) for the indirect effect is also shown. Asterisks indicate significant paths ($p < .05$).

Discussion

Substantial evidence indicates that physically aggressive individuals exhibit a heightened tendency to identify anger in ambiguous faces. The present study was the first empirical endeavor to apply diffusion modeling to identify contributions of cognitive processes (i.e., bias, efficiency of information accumulation, extent of information accumulation) to this tendency. Results suggest that physically aggressive individuals' aberrant emotion identification (i.e., heightened anger identification for anger–fear faces) stems from more efficient processing of anger-related cues (i.e., higher drift rate) rather than from bias or impulsive responding (i.e., threshold separation). Moreover, higher drift rate mediated the association between physical aggression and heightened anger identification, highlighting the role of processing efficiency for anger-related information in physically aggressive individuals' propensity to arrive at aggression-promoting interpretations of social information.

The finding that physical aggression was associated with heightened anger identification for highly ambiguous (i.e., anger–fear) faces is consistent with previous research indicating aberrant social interpretations in aggression only under high ambiguity (Dodge, 1980; Mellentin et al., 2015; Schönenberg & Jusyte, 2014; Wilkowski & Robinson, 2012; Zimmer-Gembeck & Nesdale, 2013). Yet physical aggression was not associated with overall task accuracy (see Table S1 in the Supplemental Material), indicating that more physically aggressive individuals did not erroneously identify faces as angry. Indeed, physical aggression was positively correlated with accuracy for mostly angry anger–fear faces, suggesting that these individuals were *more* accurate under high ambiguity.

The present study's key contribution is demonstrating that more physically aggressive individuals appear to be more efficient at accumulating information related to anger under highly ambiguous conditions (i.e., for anger–fear faces), and this heightened efficiency may explain their tendency to see anger where less aggressive individuals do not. Although the concept of bias is inherent in the terms used to describe aggressive individuals' patterns of interpreting social information (e.g., anger-perception bias, hostile-attribution bias), our results do not support the contention that more physically aggressive individuals display impairments or biases in emotion identification. Instead, results suggest that these individuals are more adept at processing anger-related information, building on evidence that physical aggression relates to superior anger identification abilities (Wilkowski & Robinson, 2012). Because drift rate indexes information accumulation not only from perception but also from memory (Ratcliff, Smith, Brown, & McKoon, 2016), this finding can be interpreted in light of theory positing that aggressive individuals have stronger and more accessible hostile knowledge structures, which are essentially latent memories of hostility-related events (Anderson & Bushman, 2002). That is, as aggressive individuals accrue experiences of hostile interactions (brought about in part through their own aggressive behavior; Anderson, Buckley, & Carnagey, 2008), they activate and build on their existing hostile knowledge structures, making these structures more readily accessible to aid in interpreting incoming social information. Our findings suggest, however, that rather than drawing on knowledge structures to make biased interpretations of social information, physically aggressive individuals draw on knowledge structures to make more accurate interpretations, allowing them to adopt a realistic lens for viewing their often hostile worlds.

Although we did not find physical aggression-related effects of movement, we found that participants were generally more likely to identify looming faces as angry, and this tendency was stronger for the more ambiguous anger–fear faces. Whereas previous research has shown that looming objects and faces elicit greater threat-related neural activity (Coker-Appiah et al., 2013; Vieira et al., 2017), our findings provide the first demonstration that looming ambiguous faces are more likely to be identified as angry. Thus, threat-based reactivity to looming faces may impact actual emotion identification; individuals (regardless of level of physical aggressiveness) are more likely to “see” anger in rapidly encroaching faces. Moreover, following from the diffusion-modeling results, because threshold separation was lower for looming faces, it is possible that more impulsive responding in the context of threat leads individuals to identify anger in ambiguous faces. The failure to detect an association between physical aggression and heightened anger identification for looming faces is again inconsistent with the view that physically aggressive individuals exhibit impairments in interpreting social information, because they were no more or less likely to be misled by apparent movement, a contextual factor that was orthogonal to the emotion displayed in the faces.

Several 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. 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 female and nonincarcerated (e.g., community) samples. Second, we did not present faces of varying emotional intensities, a more direct manipulation of ambiguity. Results indicated that physical-aggression-related effects were specific to anger–fear faces rather than extending to anger–happiness as well, which may reflect the tendency among physically aggressive individuals to process anger differently only under high-ambiguity conditions. However, because we did not directly manipulate ambiguity, we cannot rule out the possibility that the physical-aggression-related effects for the more ambiguous stimuli (i.e., anger–fear faces) are attributable to the specific anger–fear blend rather than greater ambiguity, per se. However, failure to find physical-aggression-related differences in general fear identification and drift rate for fear provides evidence against the interpretation that the results are an artifact of the anger–fear blend. Future research should directly manipulate ambiguity and use other types of emotion blends (e.g., anger–sadness) to test the generalizability of the present findings to other ambiguous stimuli.

Overall, the present study contributes to mounting evidence that physical aggression is associated with aberrant processing of anger. Although researchers have used the term *bias* to describe physically aggressive individuals' anger processing aberrations, the present study suggests that their aberrant processing stems from efficiency and adeptness. Thus, physical aggression may be characterized by *aggressive realism*, or a tendency to more readily process anger when it is present in ambiguous social stimuli. Progress in understanding the mechanisms contributing to physical aggression may be made by investigating how seemingly adaptive capabilities can lead to maladaptive social behaviors.

References: Study 1

Abrosoft. (2018). FantaMorph Deluxe for Mac (Version 5.5.0) [Computer software]. Retrieved from https://www.fantamorph.com/download.html

Anderson, C. A., Buckley, K. E., & Carnagey, N. L. (2008). Creating your own hostile environment: A laboratory examination of trait aggressiveness and the violence escalation cycle. *Personality and Social Psychology Bulletin, 34*, 462-473.

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology, 53,* 27-51. doi:10.1146/annurev.psych.53.100901.135231

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

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.

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

Bronson, J., & Carson, E. A. (2019). *Prisoners in 2017* (NCJ Publication No. 252156). Retrieved from http://www.bjs.gov/index.cfm?ty=pbdetail&iid=6187

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

Coccaro, E. F., McCloskey, M. S., Fitzgerald, D. A., & Phan, K. L. (2007). Amygdala and orbitofrontal reactivity to social threat in individuals with impulsive aggression. *Biological Psychiatry*, *62*, 168-178. doi:10.1016/j.biopsych.2006.08.024

Coker-Appiah, D. S., White, S. F., Clanton, R., Yang, J., Martin, A., & Blair, R. J. (2013). Looming animate and inanimate threats: The response of the amygdala and periaqueductal gray. *Social Neuroscience, 8*, 621-630. doi:10.1080/17470919.2013.839480

Conley, M. I., Dellarco, D. V., Rubien-Thomas, E., Cohen, A. O., Cervera, A., Tottenham, N., & Casey, B. J. (2018). The racially diverse affective expression (RADIATE) face stimulus set. *Psychiatry Research, 270*, 1059-1067. doi:10.1016/j.psychres.2018.04.066

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.

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

Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development, 51*, 162-170.

Dodge, K. A., & Somberg, D. R. (1987). Hostile attributional biases among aggressive boys are exacerbated under conditions of threats to the self. *Child Development*, *58*, 213-224. doi:10.1111/1467-8624.ep7264206

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

Israelashvili, J., Oosterwijk, S., Sauter, D., & Fischer, A. (2019). Knowing me, knowing you: Emotion differentiation in oneself is associated with recognition of others' emotions. *Cognition and Emotion, 33,* 1461-1471. doi:10.1080/02699931.2019.1577221

Kang, S.-M., & Shaver, P. R. (2004). Individual differences in emotional complexity: Their psychological implications. *Journal of Personality*, *72*, 687-726.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What’s new in Psychtoolbox-3? *Perception, 36*(EVCP Abstract Suppl.), 14.

Lerche, V., & Voss, A. (2016). Model complexity in diffusion modeling: Benefits of making the model more parsimonious. *Frontiers in Psychology*, *7*, 1324-1324. doi:10.3389/fpsyg.2016.01324

Marsh, A., Ambady, N., & Kleck, R. (2005). The effects of fear and anger facial expressions on approach- and avoidance-related behaviors. *Emotion, 5*, 119-124.

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

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

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

Price, R. B., Brown, V., & Siegle, G. J. (2019). Computational modeling applied to the dot-probe task yields improved reliability and mechanistic insights. *Biological Psychiatry, 85*, 606-612. doi:https://doi.org/10.1016/j.biopsych.2018.09.022

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59-108. doi:10.1037/0033-295X.85.2.59

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

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model:

Current issues and history. *Trends in Cognitive Sciences, 20,* 260-281. doi:10.1016/j.tics.2016.01.007

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

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168,* 242-249. doi:10.1016/j.psychres.2008.05.006

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

Verona, E., & Bozzay, M. L. (2017). Biobehavioral approaches to aggression implicate perceived threat and insufficient sleep: Clinical relevance and policy implications. *Policy Insights from the Behavioral and Brain Sciences, 4,* 178-185. doi:10.1177/2372732217719910

Vieira, J. B., Tavares, T. P., Marsh, A. A., & Mitchell, D. G. V. (2017). Emotion and personal space: Neural correlates of approach-avoidance tendencies to different facial expressions as a function of coldhearted psychopathic traits. *Human Brain Mapping, 38,* 1492-1506. doi:10.1002/hbm.23467

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavioral Research Methods, 39,* 767-775. doi:10.3758/bf03192967

Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: A tutorial based on fast-dm-30. *Frontiers in Psychology, 6*, Article 336. doi:10.3389/fpsyg.2015.00336

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.

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

Chapter 3: Study 2

Physical aggression is associated with heightened social reflection impulsivity

Published in:

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

Abstract

Physical aggression harms individuals, disrupts social functioning across multiple forms of psychopathology, and leads to destruction within communities. Physical aggression is associated with aberrations in the interpretation of ambiguous information. However, the specific cognitive mechanisms supporting this link remain elusive. One potentially relevant cognitive mechanism is reflection impulsivity, the amount of information gathered during decision-making. Reflection impulsivity characterizes how individuals resolve ambiguity in the process of forming judgments when multiple interpretations of a stimulus are possible. In a sample of 98 incarcerated men, we examined reflection impulsivity using a novel social information sampling task. The primary aim of the study was to investigate the relationship between physical aggression and social reflection impulsivity. Additionally, we assessed the frequency of different social judgments (hostile vs. benign), the extent to which reflection impulsivity varied in the context of these different social judgments, and subjective certainty about social judgments. Finally, we investigated whether social reflection impulsivity moderated the relationship between physical aggressiveness and violent crime. Results indicated that more physically aggressive individuals displayed heightened social reflection impulsivity, which was amplified in the context of hostile judgments. Moreover, more physically aggressive individuals were more certain about their hostile judgments and more certain when judgments were made with unconstrained access to behavioral information. Finally, impulsive hostile judgments in physically aggressive individuals related to a more extensive history of assault charges. These findings suggest that physically aggressive individuals exhibit deficits in information gathering, leading to ill-informed and inflexible social judgments.

General Scientific Summary: Physical aggression is a costly form of human behavior that is evident across multiple forms of psychopathology. This study provides the first direct evidence that more physically aggressive individuals gather less evidence during social decision-making (i.e., exhibit heightened social reflection impulsivity), particularly while making hostile judgments, and yet they are nevertheless more certain about their hostile judgments. Moreover, physically aggressive individuals with more pronounced social reflection impulsivity have a more extensive history of assault charges, highlighting the real-world implications of this social–cognitive process.

Introduction

Aggressive behavior represents a pressing public health concern, not only because it leads to significant direct harm but also because it spreads within and devastates entire communities in the same manner as infectious disease (Patel, Simon, & Taylor, 2013). Aggression is commonly defined as behavior that is likely to result in physical, social, and/or emotional harm. Aggression can manifest in various forms (e.g., physical, verbal, relational), but no form of aggression generates greater public concern than physical aggression, which is behavior that inflicts bodily harm or conveys a threat of bodily harm. The manifestations of physical aggression include a range of acts from bullying, physical fighting, and throwing objects, to more severe forms of violence, such as assault and murder. Research indicates that each year nearly 17,000 people are victims of homicide in the United States, and over 1.6 million people are hospitalized for nonfatal injuries resulting from physical aggression (Sumner et al., 2015). The overall estimated costs associated with these deaths and injuries totals $96.8 billion (Centers for Disease Control and Prevention, 2017). Physical aggression thus exacts tremendous costs at all levels of society, from individuals to entire communities.

Central to many conceptualizations of the factors driving excessive physical aggression is the impact of ambiguity on information processing. Physically aggressive individuals are more likely to perceive anger in faces displaying ambiguous emotional expressions (Barth & Bastiani, 1997; Fine, Trentacosta, Izard, Mostow, & Campbell, 2004; Schönenberg & Jusyte, 2014; Schultz, Izard, & Bear, 2004; Wilkowski & Robinson, 2012), demonstrate a hostile attribution bias (i.e., a tendency to interpret others' ambiguous actions as signs of malicious intent; Chen, Coccaro, & Jacobson, 2012; De Castro, Veerman, Koops, Bosch, & Monshouwer, 2002; Dodge, 1980), and display reduced sensitivity to ambiguity during cost-benefit decision-making

(Buckholtz, Karmarkar, Ye, Brennan, & Baskin-Sommers, 2017). Together, these findings suggest that physically aggressive individuals tend to interpret ambiguous information negatively and fail to consider ambiguity while making decisions. Although the link between aberrations in processing ambiguity and physical aggression is relatively well-established and uncontroversial, less is known about specific underlying cognitive mechanisms that support this link.

One cognitive mechanism that plays a pivotal role in decision-making under ambiguity is reflection impulsivity. Reflection impulsivity is a construct that characterizes how individuals resolve ambiguity when multiple interpretations of a stimulus are possible (Kagan, 1965). More specifically, reflection impulsivity is commonly operationalized as the extent to which individuals gather information while making a decision (Clark, Robbins, Ersche, & Sahakian, 2006; Clark et al., 2003). Individuals with heightened reflection impulsivity gather less information while making a decision, which provides them with a weaker evidence base for their chosen response and thereby increases the likelihood that they will respond inaccurately (Evenden, 1999; Kagan, 1965). Consistent with the idea that reflection impulsivity hampers adaptive decision-making, multiple studies link heightened reflection impulsivity to substance abuse (Banca et al., 2016; Clark et al., 2006; Clark, Roiser, Robbins, & Sahakian, 2009; Solowij et al., 2012; Townshend, Kambouropoulos, Griffin, Hunt, & Milani, 2014), which in turn is associated with wide-ranging decision-making deficits (Clark & Robbins, 2002). Thus, heightened reflection impulsivity represents a key mechanism influencing impaired decision-making under ambiguity. Yet, despite evidence for pervasive abnormalities in physically aggressive individuals' decision-making under ambiguity, reflection impulsivity has not been studied as it relates to aggression.[1]

[1] Although reflection impulsivity might appear to overlap with constructs such as trait impulsivity and executive functioning (which have been studied extensively in relation to aggression), multiple studies establish reflection

Multiple cognitive theories of aggression suggest that aggressive individuals' decision-making abnormalities stem from a failure to adequately consider relevant information and jumping to conclusions (Crick & Dodge, 1994; Fontaine & Dodge, 2006; Tone & Davis, 2012; Wilkowski & Robinson, 2012). Translation of these theories into well-established models of decision-making (e.g., sequential sampling models; Ratcliff & Smith, 2004; Forstmann, Ratcliff, & Wagenmakers, 2016) emphasizes that decision-making unfolds through an iterative process of gathering information about a stimulus (e.g., a target person) until a sufficient quantity of evidence has been amassed. Each possible judgment about the stimulus (e.g., whether the target person is hostile or not) may require different quantities of evidence. Gathering information about a stimulus actively reduces ambiguity, steering the agent toward the judgment that has more evidence in its favor. Once the required quantity of evidence has been amassed for one judgment or another, the corresponding judgment is made and information gathering is terminated. Certainty about the judgment serves to strengthen the judgment (Pouget, Drugowitsch, & Kepecs, 2016). Furthermore, both information gathering and certainty can vary as a function of which judgment the agent makes. For example, an agent might require less evidence in the context of judging someone as hostile versus not hostile, or an agent might be more certain about their choice in the context of making a hostile (vs. non-hostile) judgment. Applying this framework to social decision-making in aggression offers the possibility to identify variations in these cognitive processes (i.e., information gathering, certainty) that may

impulsivity as a distinct construct (Clark et al., 2003; Clark et al., 2009; Crockett et al., 2012; Jepsen et al., 2018; Perales et al., 2009), a pattern of findings replicated in the present study (see Validity subsection of the Method section). Furthermore, although previous research indicates that aggressive individuals generate fewer response alternatives to socially provocative situations (Dodge, Lochman, Harnish, Bates, & Pettit, 1997; Fontaine & Dodge, 2006), reflection impulsivity occurs during an early stage of decision-making (i.e., when individuals are deciding how many cues to encode), whereas response generation occurs later in decision-making (i.e., when individuals are deciding how to respond to the cues they encoded).

help explain why aggressive individuals fail to consider relevant information during decision-making and interpret ambiguous social information in aberrant ways.

Present Study and Hypotheses

To examine cognitive processes implicated in the social decision-making of physically aggressive individuals, we administered a novel adaptation of the information sampling task, an experimental task developed by Clark and colleagues (2003), whose validity has been established (Clark et al., 2003, 2006, 2009; see Method section for validation of the novel adaptation in the present study). In a sample of incarcerated offenders with varying levels of physical aggressiveness, we measured reflection impulsivity in the context of social judgments (i.e., social reflection impulsivity), as well as the frequency of different social judgments (hostile vs. benign), and subjective certainty about those judgments.

The primary aim of the study was to examine the relationship between physical aggression and social reflection impulsivity. To this end, we hypothesized that (1) physical aggression would be associated with heightened social reflection impulsivity (i.e., negatively associated with information gathering), above and beyond the level of reflection impulsivity evident in decision-making more broadly. Secondary aims were to examine whether other aspects of the social decision-making process were associated with aggression. Based on previous research demonstrating a hostile attribution bias in aggression (Chen et al., 2012; De Castro et al., 2002; Dodge, 1980), we aimed to examine the relationship between physical aggression and frequency of hostile judgments in the task, hypothesizing that (2) physical aggression would be positively associated with frequency of hostile judgments. Additionally, based on theoretical conjectures that aggressive individuals tend to jump to conclusions prematurely in their social decision-making, particularly when those conclusions involve judging others as hostile (Crick & Dodge, 1994; Fontaine & Dodge, 2006; Tone & Davis, 2012; Wilkowski & Robinson, 2012), we hypothesized that (3) physical aggression would be associated with heightened reflection impulsivity (i.e., negatively associated with information gathering) particularly in the context of hostile social judgments. Additionally, based on theory suggesting that aggressive individuals hold more rigid beliefs about others' hostility (Dodge, 2006), we hypothesized that (4) physical aggression would be positively associated with subjective certainty about hostile social judgments. Finally, a tertiary aim of the study was to examine whether social reflection impulsivity interacted with physical aggressiveness to predict real-world physically aggressive behavior (i.e., assault charges). To this end, we hypothesized that (5) higher physical aggression combined with higher social reflection impulsivity would be associated with the greatest number of assault charges.

Method

Participants

Participants were 98 male offenders from a high-security correctional institution in Connecticut who ranged in age from 21 to 59 (*M* = 35.33, *SD* = 10.54). 54.1% of participants identified as African American, 44.9% identified as White, and 1% identified as American Indian. 21.4% of participants identified as Hispanic (see Supplemental Table 1 in the Supplemental Material for sample characteristics and correlations among key study variables). Additionally, 95.7% of participants in the final sample had been charged with a violent crime in their lifetime (see Supplemental Table 2 in the Supplemental Material), and 46.2% had been charged with a violent institutional infraction while incarcerated (i.e., violations against persons, including fighting and assault on correctional staff; see Supplemental Table 3 in the Supplemental Material). We used a prescreen of institutional files and assessment materials to exclude individuals who had: a history of psychosis or bipolar disorder, current mood/anxiety disorders, current psychotropic medication, a family history of psychosis, medical problems that could impede comprehension of or performance on the experimental task (e.g., uncorrectable auditory or visual deficits, three or more serious head injuries), IQ below 70, or reading level below 4th grade (see Supplemental Measures in the Supplemental Material).

An a priori power analysis based on published studies on related topics (i.e., individual differences in reflection impulsivity; Clark et al., 2006, 2009; Townshend et al., 2014) indicated that a sample size of approximately 90 participants would be sufficient to detect moderate 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 98 participants.

Aggression Measures

Buss-Perry Aggression Questionnaire (AQ; Buss & Perry, 1992). The AQ is a 29-item self-report measure of aggression. Participants rate each item on a 5-point Likert scale (1 = *extremely uncharacteristic of me* to 5 = *extremely characteristic of me*). The four questionnaire subscales, established through factor analysis, are Physical Aggression (9 items), Verbal Aggression (5 items), Anger (7 items), and Hostility (8 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). Analyses in the present study focused on the AQ Physical Aggression subscale (see Supplemental Results in the Supplemental Material for additional analyses with AQ Total score). Internal consistency for the Physical Aggression subscale (Cronbach's α = .77) and the AQ as a whole (Cronbach's α = .84) in the present sample was acceptable and comparable to reliability coefficients reported by Buss and Perry (1992).

Criminal charges. Self-reported number of assault charges, a severe and legally sanctioned form of physical aggression, were cross-validated using official State of Connecticut Department of Correction files and mittimus reports.

Experimental Task

Whereas the original information sampling task provides a measure of reflection impulsivity based on how much information participants gather while making a decision about which of two colors is dominant in a visual array, the social information sampling task developed for the present study provides a measure of reflection impulsivity in a social decision-making context. More specifically, participants made decisions about which of two attributes was predominantly displayed by a person who engaged in a range of behaviors. In the social information sampling task, participants were presented with information about a person's behaviors and instructed to decide whether the person was nasty (hostile judgment) or nice (benign judgment; Dodge, 2006) based on the behaviors. Stimulus presentation and response collection were controlled using the Psychtoolbox extension (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) as implemented in MATLAB 2017b (Mathworks).

Stimuli. Stimuli consisted of behavioral descriptions that contained three words using the following construction: Verb + Article + Object. The verb in each description was either positively valenced (consistent with a "nice" behavior; e.g., "helped a man") or negatively valenced (consistent with a "nasty" behavior; e.g., "offended a man"). Twenty positively valenced verbs and 20 negatively valenced verbs were selected from the Affective Norms for English Words (ANEW) database (Bradley & Lang, 1999) on the basis of readability (i.e., comprehensible to individuals with reading ability as low as the 4th-grade level) and being mild to moderate rather than extreme in terms of valence/arousal (e.g., we included "hit" but not “killed”). Overall, the 20 positively valenced words did not differ from the 20 negatively valenced words in terms of extremeness of valence (i.e., distance from a “neutral” rating).

Conditions. The task consisted of 20 trials equally divided into *two* conditions: *partial information* and *full information*. In both conditions (per trial), participants were presented with a display containing 25 boxes arranged in a 5 × 5 grid. Participants were told that each grid represented one person, and each of the 25 boxes in that grid contained a description of a behavior performed by that person. The visibility of the behavioral descriptions at the onset of each trial varied according to the task condition.

In the *partial information condition,* all 25 boxes were gray (i.e., showed no behavioral descriptions) at the start of the trial. When participants clicked a box, the behavioral description inside the box was revealed. This description remained visible for the duration of the trial to minimize demands on working memory. In this condition, participants were instructed to open as many boxes as they wanted while deciding whether the person was mainly nasty or nice. In the *full information condition*, the information inside each of the 25 boxes was visible from the onset of the trial. Thus, participants could view the full extent of available information about the person without having to open any boxes. On each trial, participants indicated their decision about the person by clicking one of two panels (one labeled “nasty,” one labeled “nice”) at the bottom of the screen. Finally, in both conditions, participants rated how certain they were about their decision using a sliding rating bar (see Figure 1).

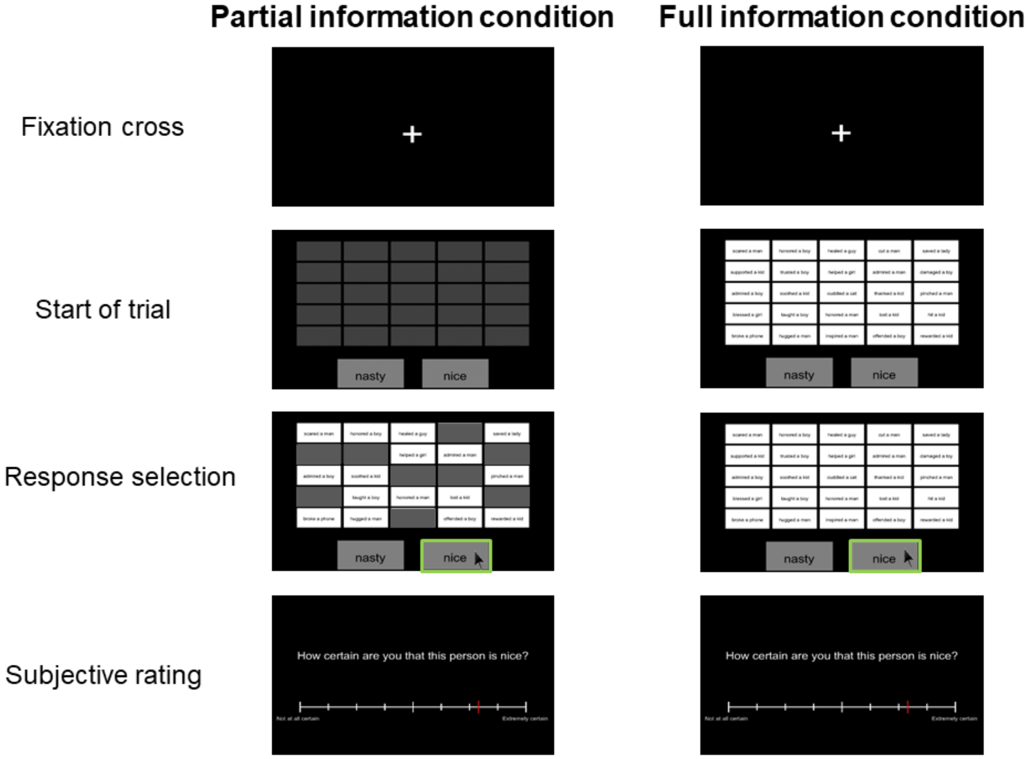


Figure 1. Schematic representation of the experimental task. The vertical sequence (top to bottom) in the left column depicts stages of trial progression in the partial information condition, while the sequence in the right column depicts stages of trial progression in the full information condition. In the partial information condition (left), participants were initially presented with a grid of dark gray boxes, with behavioral descriptions “inside” the dark gray boxes. Participants clicked individual boxes to “open” them and reveal the behavioral descriptions inside. Participants were instructed to open as many boxes as they wanted while deciding whether the person was mainly nasty or nice. In the full information condition (right), all of the information inside the boxes was visible from the beginning of each trial. Participants selected their choice by clicking on the corresponding light gray panel at the bottom of the screen (third row). On the next screen, participants were asked to rate their level of certainty regarding their decision (bottom row).

The full information condition always followed the partial information condition to avoid potential contamination of the social reflection impulsivity measure (i.e., the primary dependent variable in the present study, which was derived from the partial information condition). If participants were exposed to the full information condition first, they may have learned that their decision-making was facilitated when they had access to all of the available information, and this could have influenced them to open more boxes than they otherwise would have in the partial information condition. There is no reason to believe that placing the partial information condition first would impact performance in the full information condition. Thus, based on concerns about asymmetric transfer effects (i.e., participants learning to use one strategy in an early portion of an experiment and then carrying that strategy into subsequent portions of the experiment; see Poulton [1982] for a discussion of the disadvantageousness of counterbalancing in the context of asymmetric transfer effects) and our desire to obtain as pure a measure of social reflection impulsivity as possible, we held condition order constant across participants.

Trials. On each trial (10 per condition), the 25 boxes contained a ratio of approximately 3:2 in terms of positively valenced versus negatively valenced behaviors (or vice versa, depending on the trial). Positively valenced behaviors made up the majority of behaviors on half of the 10 trials in each condition, and negatively valenced behaviors made up the majority on the other half of the trials. Within each trial, positively valenced words did not differ from negatively valenced words in terms of extremeness of valence.

Each trial began with a 1-s fixation cross displayed in the center of the screen to indicate the start of the next trial, and each trial lasted 40 s. If participants did not click a response panel within the 40 s allotted for the trial, the trial ended and the next trial began. If participants clicked a response panel within 40 s, a black screen was displayed for the remainder of the trial time. In this way, each trial lasted 40 s regardless of decision speed, so as to discourage rushed responding. Across participants, stimuli were presented in the same order regardless of the order in which specific boxes were opened to maximize consistency of exposure to information. Before completing the 20 experimental trials, participants completed 8 practice trials (4 partial information, 4 full information), during which they received feedback regarding whether they made their decisions within the allotted amount of time.

Nonsocial control task. Following completion of the social information sampling task, participants completed a separate control task, which assessed reflection impulsivity in a nonsocial context. The layout and number of trials were the same as in the social task; however, instead of making decisions about people, participants made decisions about baskets of fruits and vegetables. Each of the 25 boxes on a trial contained a description of a type of fruit or vegetable that was either red (e.g., "is a strawberry") or green (e.g., "is a cucumber"). Fruits and vegetables were chosen for inclusion in the stimuli on the basis of being clearly either red or green, and on the basis of readability. The average letter count of the descriptions on each trial matched those for the social task so that descriptions inside the boxes would not take longer or shorter to read in either task. Participants were instructed to open as many boxes as they wanted while deciding whether the basket of fruits and vegetables was mainly red or green.

Participants always played the control task after the social task for reasons similar to those noted above (see Conditions subsection). Specifically, the social task provided our primary measures, and we wanted to avoid the potential contamination of responses in the social task due to asymmetric transfer effects (Poulton, 1982). Our goal was to encourage participants to respond in the social task as they would in a real-life social scenario, and accordingly participants were told that there were no "right" answers. Thus, ordering the tasks so that the control task followed the social task reduced the likelihood that participants would use contrived strategies in the primary social task (e.g., counting boxes) and in this way fostered more natural responding.

Key variables. The primary dependent variable derived from the social information sampling task was social reflection impulsivity. Additionally, we examined the frequency of different social judgments (hostile vs. benign) and subjective certainty about those judgments. Social reflection impulsivity was operationalized as the average number of boxes opened across trials in the partial information condition in the social task, with fewer boxes opened denoting higher social reflection impulsivity. Frequency of different social judgments was operationalized as the percentage of choices made that were nasty (i.e., percentage of judgments that were hostile) across trials within each condition of the social task. Subjective certainty was operationalized as the average certainty rating given by participants across trials within each condition of the social task.

The key variable derived from the nonsocial control task was a general measure of reflection impulsivity (i.e., average number of boxes opened in the partial information condition), which was assessed so that the role of general reflection impulsivity in the relationship between physical aggression and social reflection impulsivity could be examined. "Accuracy" was derived as a secondary measure from both the social task and the non-social task, as a means of assessing task validity, and was operationalized as the percentage of choices that matched the dominant behavior (social task) or color (nonsocial task) on each trial.

Validity. The reliability and validity of the social information sampling task was established through a series of analyses modeled after the validity analyses conducted by Clark et al. (2003). First, we calculated internal consistency and found that the social reflection impulsivity measure exhibited excellent reliability across trials (Cronbach's $\alpha = .97$). Second, we confirmed that less information gathering was related to lower "accuracy" in both the social task, $r = .54$, $p < .001$, 95% confidence interval (CI) [.43, .63], and the nonsocial task, $r = .73$, $p <$

.001, 95% CI [.62, .83]. Third, we established divergent validity by demonstrating that, consistent with previous research (Clark et al., 2003, 2009; Crockett, Clark, Smillie, & Robbins, 2012; Jepsen et al., 2018; Perales et al., 2009), extent of information gathering was associated with neither trait impulsivity (i.e., MPQ-B Constraint, see Supplemental Measures in the Supplemental Material; social task: *r* = -.02, *p* = .822; nonsocial task: *r* = -.02, *p* = .836) nor executive functioning (i.e., Color-Word Interference Test Inhibition/Naming contrast scaled score, see Supplemental Measures in the Supplemental Material; social task: *r* = -.07, *p* = .506; nonsocial task: *r* = -.03, *p* = .762).

Procedure

Before recruitment, study personnel received an institutional roster of inmates. Study personnel used this roster to review institutional files and exclude individuals who clearly did not meet eligibility criteria (see Participants section above). 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 not go into their institutional files and would not affect any pending legal status or sentencing they might be facing. In keeping with Connecticut Department of Correction regulations, participants did not receive financial compensation. They were informed that they could withdraw from the study at any time. All participants provided written informed consent according to the procedures set forth by the Yale University Human Investigation Committee. After providing consent, participants completed an initial session that involved a series of clinical and neuropsychological assessments (e.g., Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition (DSM–5)*, Wide Range Achievement Test; see Supplemental Measures in the Supplemental Material). Participants who did not meet eligibility thresholds on any of these assessments were excluded from further participation. Then, after completing questionnaires assessing personality (e.g., Multidimensional Personality Questionnaire–Brief; see Supplemental Measures in the Supplemental Material), participants returned for a second session in which they completed the experimental task followed by aggression questionnaires (e.g., AQ; see Aggression Measures). Both in-person sessions took place in a private testing space within the prison. Finally, study personnel reviewed records to obtain a measure of criminal charges for each participant (see Aggression Measures).

Results

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 conditions was met: (a) no response given (i.e., the participant did not respond in time) on more than 25% of trials, (b) statistical outliers (>3 *SD*s from the mean) on any key task variables, or (c) extreme difficulties comprehending the task as noted by the experimenter. Five participants were excluded from analyses based on these criteria, and accordingly the final sample consisted of 93 participants. Excluded participants did not differ from included participants in terms of age or physical aggression ($p > .7$).

Social Reflection Impulsivity

A linear regression, with AQ Physical Aggression (*z*-scored) as an independent variable, age and race/ethnicity as covariates[2], and social reflection impulsivity as a dependent variable indicated that AQ Physical Aggression was negatively associated with extent of social information gathering, $B = -1.85$, $SE = 0.74$, $p = .014$, 90% CI [-3.07, -0.63]. Consistent with

[2] Age and race/ethnicity were included as covariates in these analyses (and all analyses to follow) because these demographic variables were associated with task dependent variables.

Hypothesis 1, more physically aggressive participants demonstrated greater social reflection impulsivity (see Figure 2).

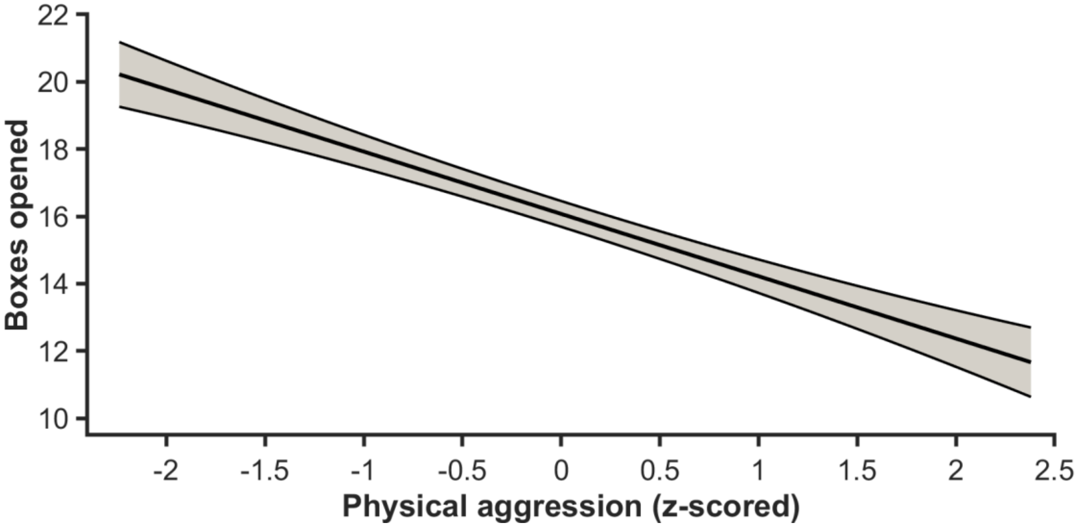


Figure 2. The relationship between physical aggression and social information gathering. Participants with higher levels of physical aggression opened fewer boxes (i.e., gathered less information) in the social information sampling task. Error band represents 90% confidence interval.

To determine whether this effect persisted even after controlling for reflection impulsivity in the nonsocial control task, another linear regression was run with the addition of nonsocial information gathering as a covariate. The analysis showed that AQ Physical Aggression was negatively associated with extent of information gathering in the social task, *B* = -1.02, *SE* = 0.40, *p* = .012, 90% CI [-1.68, -0.36], above and beyond the effects of more general reflection impulsivity as measured in the nonsocial control task. In other words, more physically aggressive participants demonstrated greater social reflection impulsivity even after controlling for general reflection impulsivity.

Frequency of Hostile Social Judgments

A two-way (condition: partial information, full information) repeated measures GLM, with AQ Physical Aggression (*z*-scored) as a continuous between-subjects independent variable and hostile social judgment frequency as a dependent variable, failed to detect a main effect of condition, $F(1,91) = .01$, $p = .914$, $\eta_p^2 < .01$, 90% CI [0.00, 0.01], or physical aggression, $F(1,91) = 3.90$, $p = .051$, $\eta_p^2 = .04$, 90% CI [0.00, 0.12]. Furthermore, the analysis failed to detect a Condition × Physical Aggression interaction, $F(1,91) = .01$, $p = .907$, $\eta_p^2 < .01.$, 90% CI [0.00, 0.01]. Thus, Hypothesis 2 (i.e., that physically aggressive individuals would display a higher frequency of hostile judgments) was not supported.

Reflection Impulsivity in the Context of Hostile Versus Benign Social Judgments

A two-way (judgment: nasty, nice) repeated measures GLM with AQ Physical Aggression (*z*-scored) as a continuous between-subjects independent variable and reflection impulsivity as a dependent variable failed to detect a main effect of judgment on reflection impulsivity, $F(1,88) = 1.43$, $p = .235$, $\eta_p^2 = .02$, 90% CI [0.00, 0.08]. However, there was a main effect of physical aggression, $F(1,88) = 5.05$, $p = .027$, $\eta_p^2 = .05$, 90% CI [0.003, 0.15], indicating that physically aggressive individuals gathered less information across both nasty and nice judgments. Furthermore, there was a Judgment × Physical Aggression interaction, $F(1,88) = 4.81$, $p = .031$, $\eta_p^2 = .05.$, 90% CI [0.003, 0.14]. In terms of this interaction, there was a simple main effect of physical aggression in the context of nasty judgments, $B = -2.18$, $SE = 0.75$, $p = .005$, $\eta_p^2 = .09$, 90% CI [-3.43, -0.93], but we failed to detect a simple main effect of aggression in the context of nice judgments, $B = -1.33$, $SE = 0.86$, $p = .125$, $\eta_p^2 = .03$, 90% CI [-2.75, 0.10].[3] Together, in line with Hypotheses 1 and 3, these results indicate that more physically aggressive individuals demonstrated greater reflection impulsivity overall in the social task, and their reflection impulsivity was particularly heightened in the context of hostile judgments (see Figure 3).

[3] These results remained unchanged after adding nonsocial (i.e., general) reflection impulsivity as a covariate in the analysis.

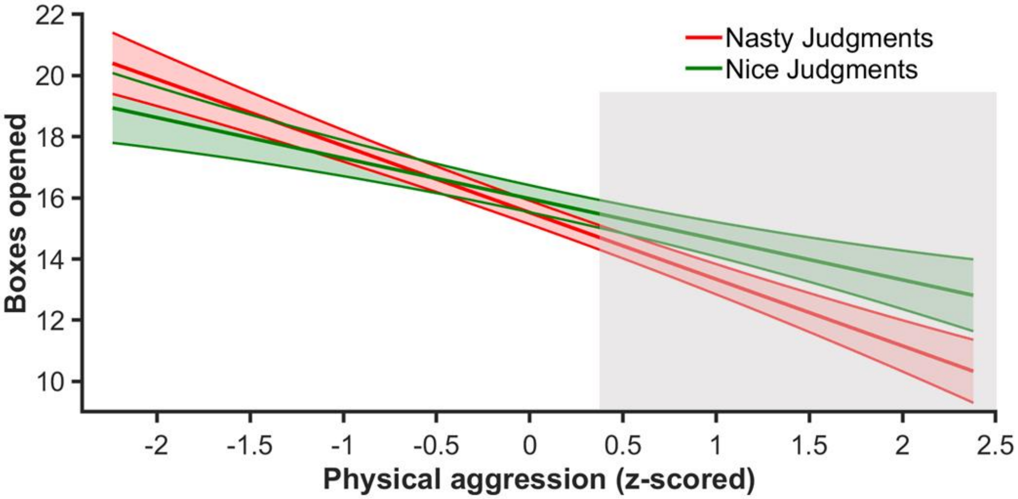


Figure 3. The relationship between physical aggression and information gathering in the context of hostile (“nasty”) versus benign (“nice”) judgments. Participants with higher levels of physical aggression gathered less information in the context of nasty judgments, but we failed to detect an effect of aggression on information gathering in the context of nice judgments. Error bands represent 90% CI. Region of significance is shown in gray shading: specifically, at *z*-scored values of physical aggression around 0.37 (i.e., AQ Physical Aggression scores around 27) and higher (representing 31 participants), there is a difference between reflection impulsivity for hostile versus benign judgments.

Subjective Certainty

A 2 (condition: partial information, full information) × 2 (judgment: nasty, nice) repeated measures GLM with AQ Physical Aggression (*z*-scored) as a continuous between-subjects independent variable and subjective certainty as a dependent variable failed to detect a main effect of condition, $F(1,84) = .002$, $p = .961$, $\eta_p^2 = .01$, 90% CI [0.00, 0.001], or physical aggression, $F(1,84) = 3.73$, $p = .057$, $\eta_p^2 = .04$, 90% CI [0.00, 0.13]. However, consistent with previous research (Rand, Ohtsuki, & Nowak, 2009; Siegel, Mathys, Rutledge, & Crockett, 2018), there was a main effect of judgment on certainty, $F(1,84) = 4.26$, $p = .042$, $\eta_p^2 = .05$, 90% CI [0.001, 0.14], such that participants were generally more certain when judging someone as nice ($M = 41.49$) than when judging someone as nasty ($M = 35.09$). Furthermore, there was a Judgment × Physical Aggression interaction, $F(1,84) = 3.97$, $p = .049$, $\eta_p^2 = .05$, 90% CI

[0.00005, 0.13], and a Condition × Physical Aggression interaction, *F*(1,84) = 4.60, *p* = .035, η_p^2 = .05, 90% CI [0.002, 0.14]. In terms of the Judgment × Physical Aggression interaction, there was a simple main effect of aggression in the context of nasty judgments, *B* = 7.80, *SE* = 3.47, *p* = .027, η_p^2 = .05, 90% CI [2.03, 13.57], but we failed to detect a simple main effect of aggression in the context of nice judgments, *B* = 1.00, *SE* = 3.36, *p* = .767, η_p^2 = .001, 90% CI [-4.58, 6.58]. In terms of the Condition × Physical Aggression interaction, there was a simple main effect of aggression in the full information condition, *B* = 7.46, *SE* = 3.22, *p* = .023, η_p^2 = .06, 90% CI [2.11, 12.82], but we failed to detect a simple main effect of aggression in the partial information condition, *B* = 2.09, *SE* = 2.81, *p* = .458, η_p^2 = .01, 90% CI [-2.57, 6.75]. Together, these results indicate that more physically aggressive individuals endorsed greater certainty particularly in the context of nasty judgments (vs. nice; see Figure 4A), consistent with Hypothesis 4, as well as greater certainty particularly in the context of full information (vs. partial; see Figure 4B; see Supplemental Results in the Supplemental Material for robustness, specificity, and additional exploratory analyses).

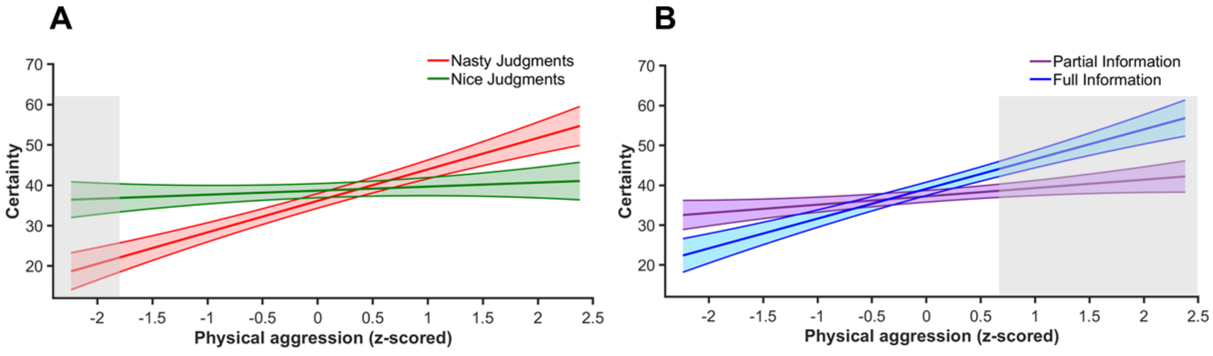


Figure 4. The relationship between physical aggression and certainty in the context of hostile ("nasty") versus benign ("nice") judgments (A) and in the context of partial versus full information conditions (B). Participants with higher levels of physical aggression endorsed greater certainty when judging someone as nasty, but we failed to detect an effect of aggression on certainty for nice judgments (A). Furthermore, participants with higher levels of physical aggression endorsed greater certainty in the full information condition, but we failed to detect an

effect of aggression on certainty in the partial information condition (B). Error bands represent 90% CI. Regions of significance are shown in gray shading: specifically, at *z*-scored values of physical aggression around -1.84 (i.e., AQ Physical Aggression scores around 12) and lower (representing 2 participants), there is a difference between reflection impulsivity in the context of hostile versus benign judgments (A). Additionally, at *z*-scored values of physical aggression around 0.67 (i.e., AQ Physical Aggression scores around 29) and higher (representing 25 participants), there is a difference between reflection impulsivity in the partial versus full information conditions (B).

Social Reflection Impulsivity and "Real-World" Behavior

The relevance of social reflection impulsivity for moderating the association between physical aggression and assault charges was assessed using a negative binomial regression with AQ Physical Aggression and hostile reflection impulsivity (the extent of information gathering in the context of nasty judgments) as continuous independent variables and number of assault charges as a count-based dependent variable. In the model examining effects of aggression (*z*-scored) and hostile reflection impulsivity (*z*-scored) as well as their interaction, χ^2/df = 1.17, $p <$ .001, only the Aggression × Hostile Reflection Impulsivity interaction predicted number of assault charges, odds ratio (OR) = 0.58, $p = .002$, 95% CI [0.42, 0.82]. Specifically, consistent with Hypothesis 5, the greatest number of assault charges resulted from a combination of high physical aggression and low information gathering in the context of hostile judgments (i.e., high hostile reflection impulsivity; see Figure 5).

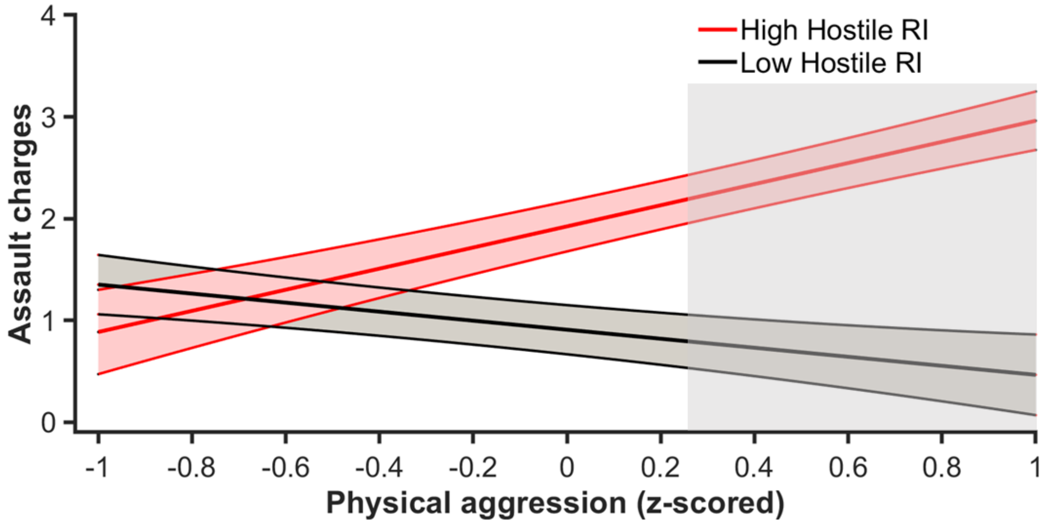


Figure 5. Simple slopes plotted 1 SD above the mean and 1 SD below the mean for hostile reflection impulsivity. Higher aggression was related to more assault charges at high levels of hostile reflection impulsivity (*B* = 0.76, *p* = .006), but we failed to detect an effect of aggression at low levels of hostile reflection impulsivity (*B* = -0.31, *p* = .182). Error bands represent 90% CI. Region of significance is shown in gray shading: specifically, at *z*-scored values of physical aggression around 0.26 (i.e., AQ Physical Aggression scores around 26) and higher (representing 40 participants), there is an effect of hostile reflection impulsivity on assault charges. RI = reflection impulsivity.

Discussion

Physically aggressive individuals interpret ambiguous information in aberrant ways, which appears to bias their social cognition and exacerbate their aggressive behavior. The results of the present study suggest that these aberrations may stem, in part, from tendencies toward reflection impulsivity, a cognitive mechanism underlying decision-making under ambiguity. Using a novel experimental task designed to assess information gathering during social decision-making, this study is the first empirical demonstration that physical aggression is associated with heightened reflection impulsivity. Specifically, we found that more physically aggressive individuals gathered less information during social decision-making. Furthermore, physically aggressive individuals' tendency toward greater social reflection impulsivity was amplified in the context of hostile judgments. However, despite their tendency to base hostile judgments on fewer pieces of information, more physically aggressive individuals reported greater certainty about their hostile judgments. More physically aggressive individuals also demonstrated greater certainty when they were presented with the full range of available social information compared with partial information. Finally, translating the present findings to a real-world measure of violent behavior, physically aggressive individuals who displayed more pronounced hostile reflection impulsivity (i.e., reflection impulsivity in the context of hostile judgments) had the most assault charges, indicating that this specific form of reflection impulsivity may play a role in violent offending.

Consistent with previous research and models of decision-making, information gathering is a key process that supports social decision-making and diminishes the ambiguity surrounding decisions (Clark et al., 2006; Forstmann et al., 2016). In research on aggression using vignette-based methodology, studies have shown that physically aggressive youth tend to make decisions more rapidly and with less consideration of available cues (Dodge & Newman, 1981; Slaby & Guerra, 1988), thereby leaving more room for ambiguity in their decision-making. Extending this pattern, in the present study physically aggressive individuals engaged in limited information gathering (i.e., higher reflection impulsivity) while deciding whether someone was hostile or benign.

Notably, we found that physically aggressive individuals engaged in limited information gathering particularly in the context of deciding that someone was hostile, which may reflect a self-protective tendency. Physically aggressive individuals typically must navigate more hostile environments from an early age (Anderson, Buckley, & Carnagey, 2008; Guerra, Rowell Huesmann, & Spindler, 2003; Weiss, Dodge, Bates, & Pettit, 1992), making it particularly likely that they will be exposed to social threats. Because the threat of mistreatment looms large when faced with a potentially hostile person, extensive information gathering or indecision under these circumstances could result in vulnerability to exploitation. The tendency toward heightened reflection impulsivity when making hostile judgments may allow aggressive individuals to constrict the timeframe during which they are vulnerable (i.e., by spending less time opening boxes) and thereby protect themselves from threat (i.e., mistreatment by hostile individuals; Hoglund & Leadbeater, 2007). The tendency to rapidly judge others as hostile may serve an adaptive function in the short term by reducing vulnerability to threats but likely serves maladaptive functions as well, such as blocking opportunities to develop positive social relationships.

Despite the fact that physically aggressive individuals' hostile judgments were based on less information, we found that they were characterized by greater certainty. In general, judgments marked by greater certainty exert stronger influences on behavior (Fazio & Zanna, 1978) and are more persistent and less amenable to new information (Tormala & Rucker, 2007). Related to aggression, greater certainty may heighten aggressive individuals' propensity to initiate and continue to engage in aggressive behavior over time. In terms of initiating acts of aggression, heightened certainty about hostile judgments may lead aggressive individuals to be more likely to act on these judgments by confronting or aggressing against the supposedly hostile individual. For example, an aggressive individual, driven by an inflated sense of certainty, may exhibit stronger determination to carry out violent retaliation against a perceived enemy, despite the fact that their reason for desiring revenge may be based on limited information. Additionally, in terms of continuing to engage in aggression over time, greater certainty that others are hostile may promote self-serving cognitive distortions (e.g., derogating and shifting blame to victims) that allow individuals to justify their harmful behavior (Slaby & Guerra, 1988; van Leeuwen, Rodgers, Gibbs, & Chabrol, 2014). Being more certain about a victim's hostility (one possible form of victim derogation) and clinging to that judgment even after inflicting harm on the victim may facilitate aggressive individuals' justification of their aggression on the basis that it neutralized the ostensible threat posed by the victim, thereby reducing sympathy for the victim and undermining motivation to change. Thus, less flexible judgments about others' hostility may promote aggression against perceived enemies, as well as contribute to the maintenance of a chronic pattern of aggressive behavior.

In addition to being more certain about their hostile judgments, we found that physically aggressive individuals reported greater certainty when they had full and unconstrained access to all available social information (i.e., in the full information condition) compared with when they gathered the information themselves (i.e., in the partial information condition). On the one hand, when all individuals were exposed to equal amounts of information and thus should have experienced comparable levels of certainty, aggressive individuals' certainty was bolstered. On the other hand, when individuals chose how much information to gather and certainty should have tracked the amount of information gathered (i.e., gathering less information should have resulted in less certainty), aggressive individuals gathered less information but their sense of certainty paradoxically was not diminished. Taken together, it appears that aggressive individuals do not appropriately adjust their level of certainty to the level of ambiguity present in the decision-making context. This interpretation is consistent with previous research indicating that aggressive individuals do not appropriately adjust their cost-benefit decisions according to varying levels of ambiguous information (Buckholtz et al., 2017). Overall, expressing more certainty when judging others as hostile (based on less information) and when exposed to equal amounts of information reflects an inflexible and overconfident style of social decision-making (see Supplemental Results in the Supplemental Material for a follow-up analysis of a potential contributing factor to aggressive individuals' certainty).

Before concluding, limitations of the present study should be noted. First, the fact that we did not find support for our hypothesis that physically aggressive individuals would be more likely to judge others as hostile may reflect limitations of our experimental design. The social information sampling task was specifically designed to measure reflection impulsivity, and consequently it may not have been an adequately sensitive measure of hostile attribution bias. Effect sizes for the association between aggression and hostile attribution biases are quite small, particularly in adult samples (De Castro et al., 2002), and multiple studies failed to find an association (Coccaro, Fanning, Fisher, Couture, & Lee, 2017; Helfritz-Sinville & Stanford, 2014). The present study, though adequately powered to detect moderate effect sizes associated with reflection impulsivity, was likely underpowered to detect smaller effect sizes associated with hostile attribution biases. More specific experimental design elements may have contributed to the null finding as well. As noted in the Method section, stimulus words that were extreme in terms of valence or arousal were excluded; the relatively low-intensity behaviors that served as stimuli in the present study may have had a minimal impact on aggressive participants' tendency to judge people in the task as hostile (Skowronski & Carlston, 1987). Additionally, previous research indicates that hostile attribution biases are more likely to arise when the decision-making context is self-relevant (Dodge & Frame, 1982), threatening (Dodge & Somberg, 1987), and spontaneous rather than deliberate (Zelli, Rowell Huesmann, & Cervone, 1995). However, none of these factors were introduced or manipulated in the present study. Future research should examine whether these factors influence the likelihood of hostile attributions in the social information sampling task.

Second, the order of the experimental conditions and tasks was not counterbalanced, raising the question of whether the ordering of experimental components impacted the present results. For example, it is possible that strategies used in the partial information condition could have carried over into the full information condition, and that strategies used in the social task could have carried over into the nonsocial task. However, our decision to present experimental components in a fixed order was based on concerns about asymmetric transfer effects (see Method section), which are rooted in evidence that social and nonsocial decision-making are subserved by separable processes (Van Overwalle, 2011). While the lack of counterbalancing and the specific ordering of experimental components were deliberate decisions made to reduce unknown or unwanted influences on the primary dependent variable (social reflection impulsivity), future research could examine the impact of different condition/task orders on social information sampling task performance.

In summary, more physically aggressive individuals displayed a more impulsive and less flexible social decision-making style, particularly in the context of hostile judgments. Furthermore, aggressive individuals who made more ill-informed hostile judgments had the most extensive history of assault charges, highlighting the relevance of social decision-making aberrations for understanding real-world violence. The present study contributes to the mounting evidence that physically aggressive individuals exhibit a host of general cognitive deficits (Giancola, Martin, Tarter, Pelham, & Moss, 1996; Hancock, Tapscott, & Hoaken, 2010; Kuin, Masthoff, Kramer, & Scherder, 2015) and a pervasive pattern of aberrant ambiguity processing (Buckholtz et al., 2017; Dodge, 2006). Moreover, the findings pinpoint a previously unidentified mechanism, social reflection impulsivity, which may contribute to the distinctive ways in which aggressive individuals construe and navigate their social worlds.

References: Study 2

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Anderson, C. A., Buckley, K. E., & Carnagey, N. L. (2008). Creating your own hostile environment: A laboratory examination of trait aggressiveness and the violence escalation cycle. *Personality and Social Psychology Bulletin, 34,* 462-473. http://dx.doi.org/10.1177/0146167207311282

Archer, J., & Haigh, A. (1997). Beliefs about aggression among male and female prisoners. *Aggressive Behavior, 23,* 405-415. http://dx.doi.org/10.1002/(SICI)1098-2337(1997)23:6_405::AID-AB1_3.0.CO;2-F

Banca, P., Lange, I., Worbe, Y., Howell, N. A., Irvine, M., Harrison, N. A., . . . Voon, V. (2016). Reflection impulsivity in binge drinking: Behavioural and volumetric correlates. *Addiction Biology, 21*, 504-515. http://dx.doi.org/10.1111/adb.12227

Barth, J. M., & Bastiani, A. (1997). A longitudinal study of emotion recognition and preschool children’s social behavior. *Merrill-Palmer Quarterly, 43*, 107-128.

Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction manual and affective ratings (Technical Report C-1)*. Gainesville: The Center for Research in Psychophysiology, University of Florida.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433-436. http://dx.doi.org/10.1163/156856897X00357

Buckholtz, J. W., Karmarkar, U., Ye, S., Brennan, G. M., & Baskin-Sommers, A. R. (2017). Blunted ambiguity aversion during cost-benefit decisions in antisocial individuals. *Scientific Reports, 7,* 2030. http://dx.doi.org/10.1038/s41598-017-02149-6

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

Centers for Disease Control and Prevention. (2017). National Violent Death Reporting System. Retrieved from https://www.cdc.gov/ViolencePrevention/NVDRS/

Chen, P., Coccaro, E. F., & Jacobson, K. C. (2012). Hostile attributional bias, negative emotional responding, and aggression in adults: Moderating effects of gender and impulsivity. *Aggressive Behavior, 38,* 47-63. http://dx.doi.org/10.1002/ab.21407

Clark, L., & Robbins, T. W. (2002). Decision-making deficits in drug addiction. *Trends in Cognitive Sciences, 6,* 361-363. http://dx.doi.org/10.1016/S1364-6613(02)01960-5

Clark, L., Robbins, T. W., Ersche, K. D., & Sahakian, B. J. (2006). Reflection impulsivity in current and former substance users. *Biological Psychiatry, 60,* 515-522. http://dx.doi.org/10.1016/j.biopsych.2005.11.007

Clark, L., Roiser, J., Imeson, L., Islam, S., Sonuga-Barke, E., & Sahakian, B. J. (2003). Validation of a novel measure of reflection impulsivity for use in adult patient populations. *Journal of Psychopharmacology, 17(Suppl.),* A36.

Clark, L., Roiser, J. P., Robbins, T. W., & Sahakian, B. J. (2009). Disrupted reflection impulsivity in cannabis users but not current or former ecstasy users. *Journal of Psychopharmacology, 23,* 14-22. http://dx.doi.org/10.1177/0269881108089587

Coccaro, E. F., Fanning, J. R., Fisher, E., Couture, L., & Lee, R. J. (2017). Social emotional information processing in adults: Development and psychometrics of a computerized video assessment in healthy controls and aggressive individuals. *Psychiatry Research, 248,* 40-47. http://dx.doi.org/10.1016/j.psychres.2016.11.004

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. http://dx.doi.org/10.1037/0033-2909.115.1.74

Crockett, M. J., Clark, L., Smillie, L. D., & Robbins, T. W. (2012). The effects of acute tryptophan depletion on costly information sampling: Impulsivity or aversive processing? *Psychopharmacology, 219*, 587-597. http://dx.doi.org/10.1007/s00213-011-2577-9

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. http://dx.doi.org/10.1111/1467-8624.00447

Delis, D. C., Kaplin, E., & Kramer, J. (2001). *Delis Kaplin executive function system*. San Antonio, TX: The Psychological Corporation.

Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development, 51*, 162-170. http://dx.doi.org/10.2307/1129603

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

Dodge, K. A., & Frame, C. L. (1982). Social cognitive biases and deficits in aggressive boys. *Child Development, 53*, 620-635. http://dx.doi.org/ 10.2307/1129373

Dodge, K. A., Lochman, J. E., Harnish, J. D., Bates, J. E., & Pettit, G. S. (1997). Reactive and proactive aggression in school children and psychiatrically impaired chronically assaultive youth. *Journal of Abnormal Psychology, 106*, 37-51. http://dx.doi.org/10.1037/0021-843X.106.1.37

Dodge, K. A., & Newman, J. P. (1981). Biased decision-making processes in aggressive boys. *Journal of Abnormal Psychology, 90*, 375-379. http://dx.doi.org/10.1037/0021-843X.90.4.375

Dodge, K. A., & Somberg, D. R. (1987). Hostile attributional biases among aggressive boys are exacerbated under conditions of threats to the self. *Child Development, 58*, 213-224. http://dx.doi.org/10.2307/1130303

Evenden, J. (1999). The pharmacology of impulsive behaviour in rats V: The effects of drugs on responding under a discrimination task using unreliable visual stimuli. *Psychopharmacology, 143*, 111-122. http://dx.doi.org/10.1007/s002130050926

Fazio, R. H., & Zanna, M. P. (1978). On the predictive validity of attitudes: The roles of direct experience and confidence. *Journal of Personality, 46*, 228-243. http://dx.doi.org/10.1111/j.1467-6494.1978.tb00177.x

Fine, S. E., Trentacosta, C. J., Izard, C. E., Mostow, A. J., & Campbell, J. L. (2004). Anger perception, caregivers' use of physical discipline, and aggression in children at risk. *Social Development, 13*, 213-228. http://dx.doi.org/10.1111/j.1467-9507.2004.000264.x

First, M. B., Williams, J., Karg, R. S., & Spitzer, R. L. (2015). *Structured clinical interview for DSM-5–Research Version*. Arlington, VA: American Psychiatric Association.

Fontaine, R. G., & Dodge, K. A. (2006). Real-time decision making and aggressive behavior in youth: A heuristic model of response evaluation and decision (RED). *Aggressive Behavior, 32*, 604-624. http://dx.doi.org/10.1002/ab.20150

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology, 67*, 641-666. http://dx.doi.org/10.1146/annurev-psych-122414-033645

Garofalo, C., & Wright, A. G. (2017). Alcohol abuse, personality disorders, and aggression: The quest for a common underlying mechanism. *Aggression and Violent Behavior, 34*, 1-8. http://dx.doi.org/10.1016/j.avb.2017.03.002

Giancola, P. R., Martin, C. S., Tarter, R. E., Pelham, W. E., & Moss, H. B. (1996). Executive cognitive functioning and aggressive behavior in preadolescent boys at high risk for substance abuse/dependence. *Journal of Studies on Alcohol, 57*, 352-359. http://dx.doi.org/10.15288/jsa.1996.57.352

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

Hancock, M., Tapscott, J. L., & Hoaken, P. N. (2010). Role of executive dysfunction in predicting frequency and severity of violence. *Aggressive Behavior, 36*, 338-349. http://dx.doi.org/10.1002/ab.20353

Hare, R. D. (2003). *Manual for the revised psychopathy checklist* (2nd ed.). Toronto, Ontario, Canada: Multi-Health Systems.

Hare, R. D., & McPherson, L. M. (1984). Violent and aggressive behavior by criminal psychopaths. *International Journal of Law and Psychiatry, 7*, 35-50. http://dx.doi.org/10.1016/0160-2527(84)90005-0

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

Helfritz-Sinville, L. E., & Stanford, M. S. (2014). Hostile attribution bias in impulsive and premeditated aggression. *Personality and Individual Differences, 56*, 45-50. http://dx.doi.org/10.1016/j.paid.2013.08.017

Hoglund, W. L., & Leadbeater, B. J. (2007). Managing threat: Do social–cognitive processes mediate the link between peer victimization and adjustment problems in early adolescence? *Journal of Research on Adolescence*, *17*, 525-540. http://dx.doi.org/10.1111/j.1532-7795.2007.00533.x

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

Jepsen, J. R. M., Rydkjaer, J., Fagerlund, B., Pagsberg, A. K., Jespersen, R. A. F., Glenthøj, B. Y., & Oranje, B. (2018). Overlapping and disease specific trait, response, and reflection impulsivity in adolescents with first-episode schizophrenia spectrum disorders or attention-deficit/hyperactivity disorder. *Psychological Medicine, 48*, 604-616. http://dx.doi.org/10.1017/S0033291717001921

Kagan, J. (1965). Individual differences in the resolution of response uncertainty. *Journal of Personality and Social Psychology, 2*, 154-160. http://dx.doi.org/10.1037/h0022199

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*, 1.

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*, *136*, 768-821. http://dx.doi.org/ 10.1037/a0020327

Krueger, R. F., Hicks, B. M., Patrick, C. J., Carlson, S. R., Iacono, W. G., & McGue, M. (2002). Etiologic connections among substance dependence, antisocial behavior and personality: Modeling the externalizing spectrum. *Journal of Abnormal Psychology, 111,* 411-424. http://dx.doi.org/10.1037/0021-843X.111.3.411

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. http://dx.doi.org/ 10.1037/0021-843X.116.4.645

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: “Seizing” and “freezing.” *Psychological Review, 103*, 263-283. http://dx.doi.org/10.1037/0033-295X.103.2.263

Kruglanski, A. W., Webster, D. M., & Klem, A. (1993). Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of Personality and Social Psychology, 65,* 861-876. http://dx.doi.org/10.1037/0022-3514.65.5.861

Kuin, N., Masthoff, E., Kramer, M., & Scherder, E. (2015). The role of risky decision-making in aggression: A systematic review. *Aggression and Violent Behavior, 25*, 159-172. http://dx.doi.org/10.1016/j.avb.2015.07.018

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, 1*, 173-181. https://doi.org/10.1023/A:1026595011371

Patel, D. M., Simon, M. A., & Taylor, R. M. (2013). *Contagion of violence: Workshop summary*. Washington, DC: The National Academies Press.

Patrick, C. J., Curtin, J. J., & Tellegen, A. (2002). Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment, 14*, 150-163. http://dx.doi.org/10.1037/1040-3590.14.2.150

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437-442. http://dx.doi.org/10.1163/156856897X00366

Perales, J. C., Verdejo-García, A., Moya, M., Lozano, O., & Pérez-García, M. (2009). Bright and dark sides of impulsivity: Performance of women with high and low trait impulsivity on neuropsychological tasks. *Journal of Clinical and Experimental Neuropsychology, 31*(8), 927-944. http://dx.doi.org/10.1080/13803390902758793

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: Distinct probabilistic quantities for different goals. *Nature Neuroscience, 19*, 366-374. http://dx.doi.org/10.1038/nn.4240

Poulton, E. C. (1982). Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychological Bulletin, 91*, 673-690. http://dx.doi.org/10.1037/0033-2909.91.3.673

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., . . . Liu, J. (2006). The reactive–proactive aggression questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior, 32*, 159-171. http://dx.doi.org/10.1002/ab.20115

Raine, A., Lencz, T., Bihrle, S., LaCasse, L., & Colletti, P. (2000). Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. *Archives of General Psychiatry, 57*, 119-127. http://dx.doi.org/10.1001/archpsyc.57.2.119

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. http://dx.doi.org/10.1016/j.jtbi.2008.09.015

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*, 333-367. http://dx.doi.org/10.1037/0033-295X.111.2.333

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. http://dx.doi.org/10.1007/s00406-013-0440-1

Schultz, D., Izard, C. E., & Bear, G. (2004). Children's emotion processing: Relations to emotionality and aggression. *Development and Psychopathology, 16*, 371-387. http://dx.doi.org/10.1017/S0954579404044566

Séguin, J. R., Nagin, D., Assaad, J.-M., & Tremblay, R. E. (2004). Cognitive-neuropsychological function in chronic physical aggression and hyperactivity. *Journal of Abnormal Psychology, 113*, 603-613. http://dx.doi.org/10.1037/0021-843X.113.4.603

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

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology, 52,* 689-699. http://dx.doi.org/10.1037/0022-3514.52.4.689

Slaby, R. G., & Guerra, N. G. (1988). Cognitive mediators of aggression in adolescent offenders: I. Assessment. *Developmental Psychology, 24*, 580-588. http://dx.doi.org/10.1037/0012-1649.24.4.580

Solowij, N., Jones, K. A., Rozman, M. E., Davis, S. M., Ciarrochi, J., Heaven, P. C., . . . Yücel, M. (2012). Reflection impulsivity in adolescent cannabis users: A comparison with alcohol-using and non-substance-using adolescents. *Psychopharmacology, 219*, 575-586. http://dx.doi.org/10.1007/s00213-011-2486-y

Sumner, S. A., Mercy, J. A., Dahlberg, L. L., Hillis, S. D., Klevens, J., & Houry, D. (2015). Violence in the United States: Status, challenges, and opportunities. *Journal of the American Medical Association, 314*, 478-488. http://dx.doi.org/10.1001/jama.2015.8371

Tone, E. B., & Davis, J. S. (2012). Paranoid thinking, suspicion, and risk for aggression: A neurodevelopmental perspective. *Development and Psychopathology, 24*, 1031-1046. http://dx.doi.org/10.1017/S0954579412000521

Tormala, Z. L., & Rucker, D. D. (2007). Attitude certainty: A review of past findings and emerging perspectives. *Social and Personality Psychology Compass, 1*, 469-492. http://dx.doi.org/10.1111/j.1751-9004.2007.00025.x

Townshend, J. M., Kambouropoulos, N., Griffin, A., Hunt, F. J., & Milani, R. M. (2014). Binge drinking, reflection impulsivity, and unplanned sexual behavior: Impaired decision-making in young social drinkers. *Alcoholism: Clinical and Experimental Research, 38*(4), 1143-1150. http:// dx.doi.org/10.1111/acer.12333

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. http://dx.doi.org/10.1016/j.paid.2004.04.012

van Leeuwen, N., Rodgers, R. F., Gibbs, J. C., & Chabrol, H. (2014). Callous-unemotional traits and antisocial behavior among adolescents: The role of self-serving cognitions. *Journal of Abnormal Child Psychology, 42*, 229-237. http://dx.doi.org/10.1007/s10802-013-9779-z

Van Overwalle, F. (2011). A dissociation between social mentalizing and general reasoning. *Neuroimage, 54*, 1589-1599. http://dx.doi.org/10.1016/j.neuroimage.2010.09.043

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. http://dx.doi.org/10.2307/1131558

Weiss, J. L., & Schell, R. E. (1991). Estimating WAIS-R IQ from the Shipley Institute of Living Scale: A replication. *Journal of Clinical Psychology, 47*, 558-562. http://dx.doi.org/10.1002/1097-4679 (199107)47:4<558::AID-JCLP2270470414>3.0.CO;2-W

Wilkinson, G. S. (1993). *WRAT3: The Wide Range Achievement Test administration manual* (3rd ed.). Wilmington, DE: Wide Range, Inc.

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. http://dx.doi.org/10.1177/0146167211430233

Zachary, R. A. (1986). *Shipley Institute of Living Scale: Revised manual*. Los Angeles, CA: Western Psychological Services.

Zelli, A., Rowell Huesmann, L., & Cervone, D. (1995). Social inference and individual differences in aggression: Evidence for spontaneous judgments of hostility. *Aggressive Behavior, 21*, 405-417. https://doi.org/10.1002/1098-2337(1995)21:6<405::AID-AB2480210602>3.0.CO;2-N

Chapter 4: Study 3

Physical aggression is associated with more effective postdecisional processing of social threat

Published in:

Brennan, G.M. & Baskin-Sommers, A. (2021). Physical aggression is associated with more effective postdecisional processing of social threat. Advance online publication. *Clinical Psychological Science.* https://doi.org/10.1177/2167702620966294

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.

Introduction

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.

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.

[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.

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 α = .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 α = .92) was excellent.

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 &

Table 1

Sample Characteristics for Final Sample and Correlations among Task Variables

Variable	1	2	3	4	5	6	7	8	9	10	Mean	SD	Range
1) Age	--										33.19	8.35	20-57
2) Race[a]	.11	--											
White													
Black													
Asian													
Multiracial													
3) Ethnicity[b]	-.17	-.27*	--										
Not Hispanic													
Hispanic													
4) Educational attainment[c]	-.02	-.11	-.11	--									
Middle school or below													
Some high school													
Completed high school													
Some college													
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 including 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

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 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

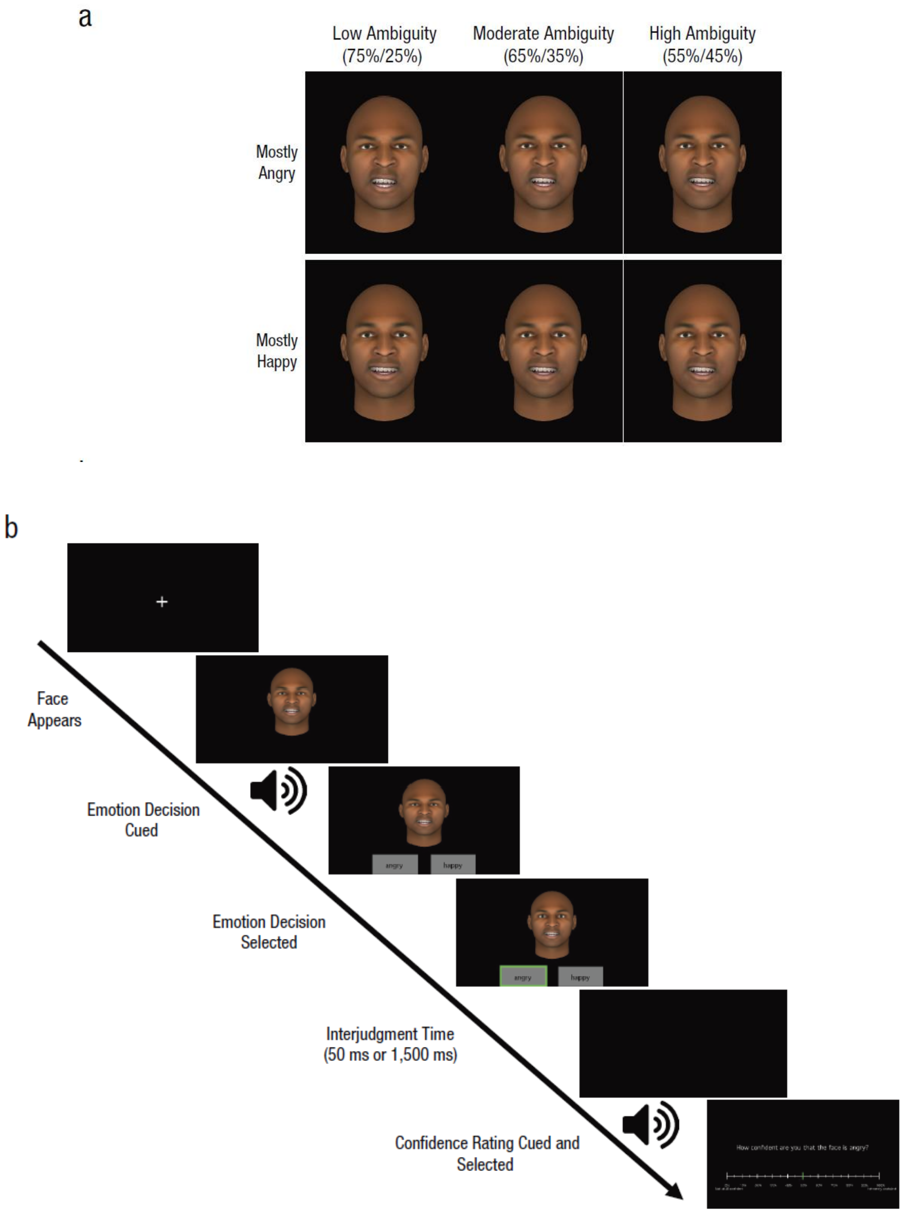

a
Low Ambiguity (75%/25%)
Moderate Ambiguity (65%/35%)
High Ambiguity (55%/45%)
Mostly Angry
Mostly Happy
b
+
Face Appears
Emotion Decision Cued
angry
happy
Emotion Decision Selected
Interjudgment Time (50 ms or 1,500 ms)
Confidence Rating Cued and Selected
How confident are you that the face is angry?

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.

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,[1] 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

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

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],[2] 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* =

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

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].

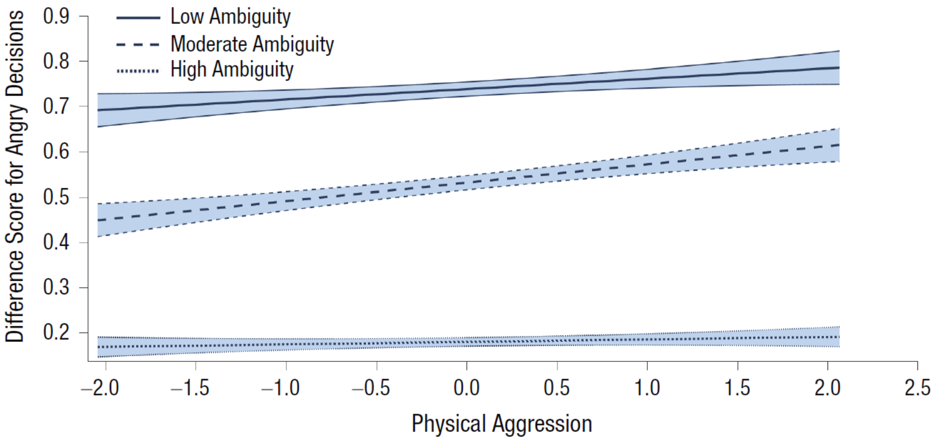


Fig. 2. The relationship between physical aggression (*z*-scored) and difference score for the proportion of angry decisions at low, moderate, and high ambiguity. Difference scores were calculated by subtracting the proportion of angry decisions for mostly happy faces from the proportion of angry decisions 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, $F(1, 73) = 124.00$, $p < .001$, $\eta_p^2 = .63$, 90% CI = [.51, .70], and the contrast between moderate ambiguity and high ambiguity, $F(1, 73) = 43.65$, $p < .001$, $\eta_p^2 = .37$, 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, $F(1, 73) = 54.71$, $p < .001$, $\eta_p^2 = .43$, 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, $F(2, 146) = 20.79$, $p < .001$, $\eta_p^2 = .22$, 90% CI = [.12, .31]. Second, we detected an Ambiguity × IJT interaction, $F(2, 146) = 5.86$, $p = .004$, $\eta_p^2 = .07$, 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.

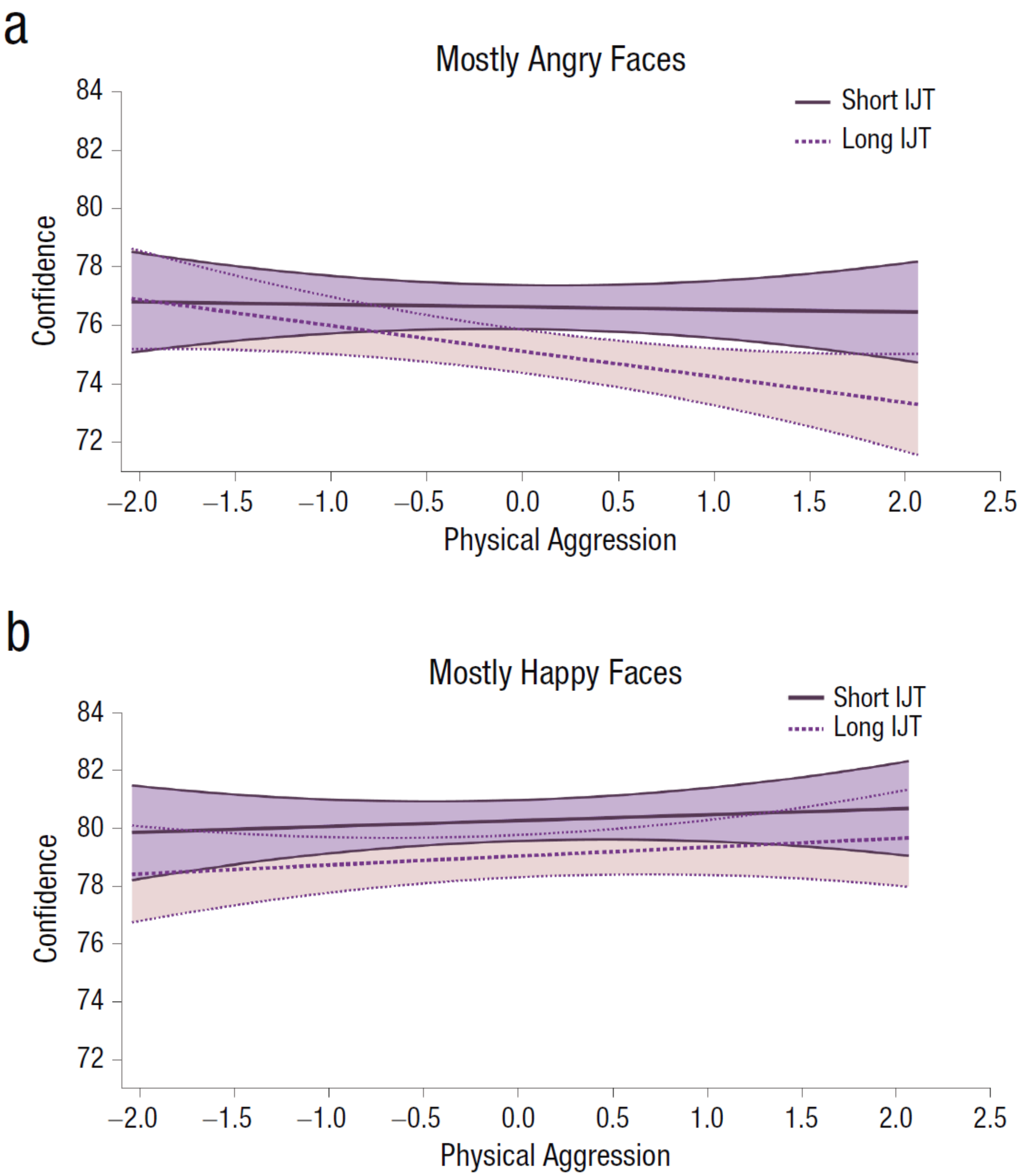


Fig. 3. The relationship between physical aggression (*z*-scored) and confidence as a function of interjudgment time (IJT) for mostly angry faces (a) and mostly happy faces (b). Participants with higher levels of physical aggression showed steeper decreases in confidence from the short IJT to the long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

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 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).

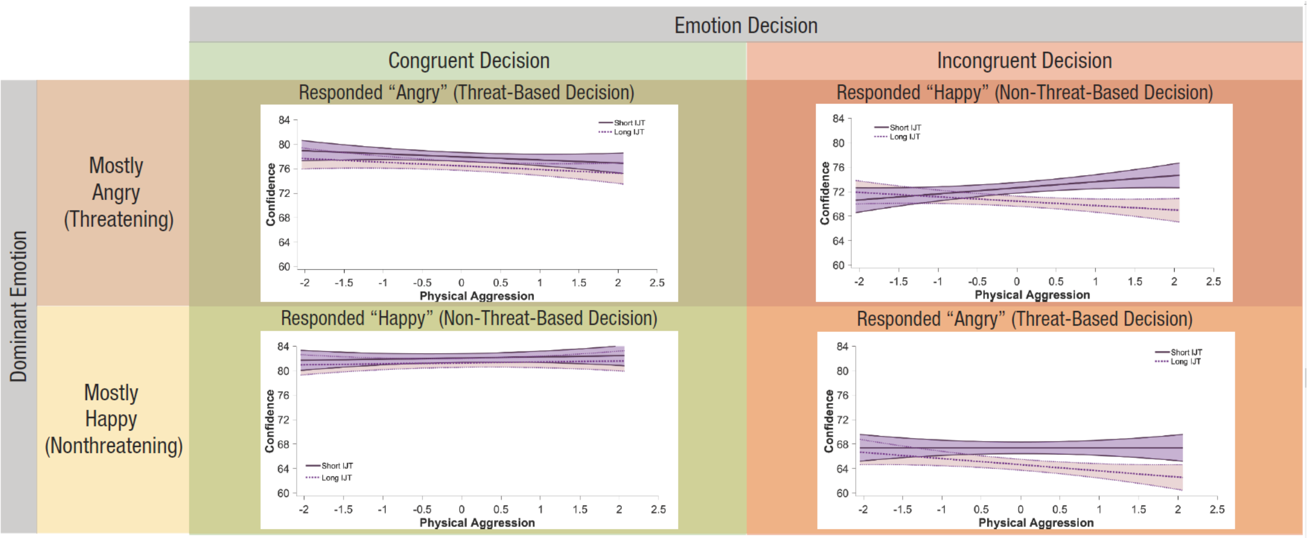


Fig. 4. The relationship between physical aggression (*z*-scored) and confidence as a function of interjudgment time (IJT) for faces that were mostly angry compared with mostly happy and for emotion decisions that were congruent (i.e., matched the dominant emotion displayed in the face) compared with 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,[3] 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 =

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

[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.

References: Study 3

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 (DSM-5)*. Washington, D. C.: American Psychiatric Publishing.

Anestis, M. D., Anestis, J. C., Selby, E. A., & Joiner, T. E. (2009). Anger rumination across forms of aggression. *Personality and Individual Differences, 46*(2), 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*(6), 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*(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*(4), 669-682. doi:10.1017/S0954579400006775

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433-436.

Brennan, G. M., & Baskin-Sommers, A. R. (2019). Physical aggression is associated with heightened social reflection impulsivity. *Journal of Abnormal Psychology, 128*(5), 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*. Advance online publication. doi:10.1177/0956797620904157

Bronson, J., & Carson, E. A. (2019). Prisoners in 2017 (NCJ Publication No. 252156). Retrieved from http://www.bjs.gov/index.cfm?ty=pbdetail&iid=6187

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*(6), 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*(1), 273-279.

Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology, 63*(3), 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.

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*(5), 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*(2), 103-123. doi:10.1177/1088868312467086

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*(3), 791-814.

Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science*, *20*(10), 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*(3), 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*(7), 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*(5), 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*(11), 1047-1053. doi:10.1016/S0005-7967(97)00064-8

Hayes, A. F. (2017). *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*(2), 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*(1), 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.

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*(4), 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*(3), 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*(4), 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*(3), 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*(1), 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*. 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*(2), 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*(1), 66-75. doi:10.1080/15374410709336569

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 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*(3), 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*(3), 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*(1), 45-57.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873-922. doi:10.1162/neco.2008.12-06-420

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*(8), 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*. 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*(2), 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*(5), 681-701. doi:10.1080/02699930302281

Smith, P., & Waterman, M. (2004). Role of experience in processing bias for aggressive words in forensic and non-forensic populations. *Aggressive Behavior, 30*(2), 105-122. doi:10.1002/ab.20001

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*(2), 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*(5), 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*(3), 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*(2), 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*(2), 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*(2), 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*, 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*(6), 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*(6), 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*(1), 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*(1), 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.

Wright, S., Day, A., & Howells, K. (2009). Mindfulness and the treatment of anger problems. *Aggression and Violent Behavior, 14*(5), 396-401. doi:https://doi.org/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*(2), 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*(1), 29-38. doi:10.1111/j.1467-6494.2012.00792.x

Chapter 5: General Discussion

The findings from this dissertation provide new evidence regarding the cognitive mechanisms driving aberrant social cognition in physical aggression. The three studies that comprise this dissertation identified physical aggression-related differences in putative cognitive mechanisms across two stages (i.e., formation and maintenance of social judgments) and two levels (i.e., lower-order facial emotion judgments and higher-order trait judgments) of social cognition. Findings from Studies 1 and 3 indicated that physical aggression was associated with greater adeptness at the formation and maintenance stages of lower-order social judgments—specifically, more adept processing of anger (i.e., threat) in ambiguous faces. Physical aggression related to more efficient accumulation of evidence related to anger (i.e., higher drift rate for anger) during the formation of judgments about highly ambiguous faces. This heightened efficiency of evidence accumulation for anger appeared to carry over into the maintenance stage as well, leading more physically aggressive individuals to experience steeper decreases in confidence following non-threatening judgments of threatening faces. Findings from Study 2 indicated that physical aggression was associated with a reduced extent of evidence accumulation during the formation of higher-order trait judgments, particularly hostile judgments, suggesting greater impulsivity. Despite exhibiting greater impulsivity while forming hostile judgments, more physically aggressive individuals were nevertheless more certain about these judgments. Thus, the formation of higher-order judgments in Study 2 appeared to be driven by mechanisms distinct from those driving the formation of lower-order judgments in Studies 1 and 3.

These findings have important theoretical implications. Social information processing theory (Dodge & Crick, 1990; Crick & Dodge, 1994) is arguably the most influential theory of social cognition in aggression. Guided by this theory, investigations of social information processing in physical aggression have revealed that more physically aggressive individuals show a heightened tendency to interpret ambiguous social stimuli as threatening or hostile. Although such general tendencies can reflect the workings of multiple candidate cognitive processes, these tendencies have traditionally been attributed to bias and interpreted as evidence of social cognitive impairments. The findings of this dissertation are inconsistent with this perspective. In Study 1, rather than displaying a bias toward threatening interpretations of ambiguous emotional faces (i.e., an anger perception bias), more physically aggressive individuals displayed more efficient processing of anger-related information. In Study 2, rather than displaying a stronger tendency to judge others as hostile (i.e., a hostile attribution bias), more physically aggressive individuals accumulated less evidence to support their hostile judgments (while showing no differences in frequency of hostile judgments). Finally, in Study 3, rather than displaying a heightened tendency to judge ambiguous emotional faces as angry, more physically aggressive individuals displayed a heightened ability to differentiate between threatening and non-threatening emotional faces under moderate ambiguity. Moreover, they displayed a pattern of postdecisional processing that suggested a heightened ability to recognize social threat *only* when a face was indeed threatening, and not in the absence of social threat. Taken together, the findings of this dissertation contribute a new level of specificity regarding the cognitive mechanisms driving aberrant social cognition in physical aggression. Moreover, they provide evidence for a refined mechanistic account of social cognition in physical aggression that stands in contrast with the traditional bias account.

This refined mechanistic account has implications for the treatment of individuals with elevated physical aggression as well. A number of existing interventions for aggressive behavior focus on modifying so-called interpretation biases (AlMoghrabi, Huijding, & Franken, 2018;

Hawkins & Cougle, 2013; Hiemstra, De Castro, & Thomaes, 2019; Penton-Voak et al., 2013; Van Bockstaele, van der Molen, van Nieuwenhuijzen, & Salemink, 2020; Vassilopoulos, Brouzos, & Andreou, 2014). Based on the findings of this dissertation, social cognitive biases may not be the most appropriate intervention target. The failure to detect evidence of bias in physically aggressive individuals' social cognition represents a failure to implicate bias as a cognitive mechanism that influences physically aggressive individuals' social cognition. Thus, it appears worthwhile to identify other cognitive mechanisms that might serve as intervention targets.

The findings of this dissertation suggest that it may be more appropriate to target efficiency of evidence accumulation for anger, extent of evidence accumulation for hostile trait judgments, or postdecisional processing of social threat. However, targeting these mechanisms should be weighed against the extent to which real threats are present in an individual's environment. The processes identified in this dissertation are likely to be adaptive in some ways and to confer an advantage in terms of identifying the presence of real threat. Thus, for individuals who regularly encounter the threat of physical harm in their environment, it may not be desirable to alter these processes. Instead, interventions could focus on breaking the link between making threat-based social judgments (e.g., identifying someone's face as angry) and responding with physical aggression. Mindfulness-based interventions could train physically aggressive individuals to notice their threat-based social judgments in a less judgmental manner (e.g., noting that someone's angry facial expression might be due to a variety of reasons, which may not involve a desire to inflict harm). This more mindful stance could promote greater consideration of how to respond less reactively and more effectively to signs of threat. Moreover, given that cognitive processes during both the formation and maintenance stages appear relevant for promoting threat-based social judgments, mindfulness training could be applied not only to the initial judgments themselves, but also to the subsequent processing of information (e.g., bringing mindful awareness to feeling less sure that someone was not looking at you “the wrong way”).

Future research is needed to address important questions that remain. The studies in this dissertation examined the formation of lower-order social judgments (Study 1), the formation of higher-order social judgments (Study 2), and the maintenance of lower-order social judgments (Study 3). The question of how higher-order social judgments are maintained in physical aggression remains open. Given that the formation of higher-order social judgments, particularly threat-based judgments, in Study 2 was characterized by heightened impulsivity, would a similar pattern of less extensive evidence accumulation characterize the maintenance of these judgments as well? A paradigm designed to assess postdecisional processing for higher-order social judgments would shed light on this question as well as provide additional evidence regarding the extent to which the mechanisms associated with lower-order versus higher-order social judgments are distinct in physical aggression.

Although the findings of this dissertation suggest that distinct mechanisms operate in the context of lower-order versus higher-order social judgments in physical aggression, more research is needed to address this question. Based on the potential influence of methodological differences (e.g., the presence of indirect provocation in Study 2 but not Studies 1 and 3), it is not entirely clear whether differences in findings (i.e., heightened impulsivity in Study 2 but not Study 1, heightened confidence in threat-based judgments in Study 2 but not Study 3) are rooted in the lower-order versus higher-order distinction. Additional research using different types of paradigms will be instrumental for evaluating the distinctiveness of mechanisms for lower-order versus higher-order social judgments. In particular, future studies should manipulate provocation in order to examine the impact of provocation on the cognitive mechanisms underlying the formation and maintenance of social judgments in physical aggression. Finally, future research should examine whether physically aggressive individuals' heightened efficiency of evidence accumulation for threat-related social information (i.e., facial cues of anger in Study 1) extends to other forms of threat-related social information as well (e.g., dominance smiles; Niedenthal, Mermillod, Maringer, & Hess, 2010).

In conclusion, the three studies that comprise this dissertation illuminate novel cognitive mechanisms supporting the formation and maintenance of social judgments in physical aggression. Previous research has identified patterns in the judgments physically aggressive individuals make about social information, and new approaches that apply insights and tools from the cognitive and decision sciences (e.g., computational modeling) may uncover the contributions of previously unrecognized cognitive mechanisms (Smeijers, Bulten, & Brazil, 2019). A better understanding of these cognitive mechanisms may be translated into more precise interventions that have the potential to significantly reduce the impact of physical aggression and violent crime on society.

References: General Discussion

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

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.

Dodge, K. A., & Crick, N. R. (1990). Social information-processing bases of aggressive behavior in children. *Personality and Social Psychology Bulletin, 16*, 8-22. doi:10.1177/0146167290161002

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

Hiemstra, W., De Castro, B. O., & Thomaes, S. (2019). Reducing aggressive children’s hostile attributions: A cognitive bias modification procedure. *Cognitive Therapy and Research, 43*(2), 387-398. doi:10.1007/s10608-018-9958-x

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences, 33*(6), 417-433.

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

Smeijers, D., Bulten, E. B. H., & Brazil, I. A. (2019). The computations of hostile biases (CHB) model: Grounding hostility biases in a unified cognitive framework. *Clinical Psychology Review, 73*, 101775. doi:10.1016/j.cpr.2019.101775

Van Bockstaele, B., van der Molen, M. J., van Nieuwenhuijzen, M., & Salemink, E. (2020). Modification of hostile attribution bias reduces self-reported reactive aggressive behavior in adolescents. *Journal of Experimental Child Psychology, 194*, 104811. doi:10.1016/j.jecp.2020.104811

Vassilopoulos, S. P., Brouzos, A., & Andreou, E. (2014). A multi-session attribution modification program for children with aggressive behaviour: Changes in attributions, emotional reaction estimates, and self-reported aggression. *Behavioural and Cognitive Psychotherapy, 43*(5), 538-548. doi:10.1017/S1352465814000149

Appendix A: Study 1 Supplemental Material

Supplemental Results

Specificity of findings to physical aggression

To assess whether the present findings generalize to other aggression-related constructs or aggression in general, the primary analyses were run with BPAQ Verbal Aggression, Anger, Hostility, and Total scores (all *z* scored) as separate independent variables instead of Physical Aggression. We failed to detect associations between any other aggression variable and study dependent variables (i.e., anger identification, diffusion model parameters).

Controlling for anxiety

To test whether anxiety contributed to physically aggressive individuals' task performance, analyses were re-run with the addition of Welsh Anxiety Inventory (WAI; Welsh, 1956) total score (*z* scored) as a covariate. When controlling for anxiety, results remained unchanged except for one difference in the analysis of anger identification. Specifically, the significant emotion blend×physical aggression interaction became non-significant, $F(1,85)=3.55$, $p=.063$, $\eta_p{}^2=0.04$, 90% CI [0.00, 0.13]. Results also revealed a main effect of anxiety on anger identification, $F(1,85)=7.01$, $p=.010$, $\eta_p{}^2=0.08$, 90% CI [0.01, 0.18], as well as drift rate for anger, $F(1,85)=4.18$, $p=.044$, $\eta_p{}^2=0.05$, 90% CI [0.001, 0.14].

Controlling for childhood physical abuse

To test whether childhood physical abuse contributed to physically aggressive individuals' task performance, analyses were re-run with the addition of Childhood Trauma Questionnaire-Short Form (CTQ-SF; Bernstein et al., 2003) Physical Abuse (*z* scored) as a covariate. When controlling for physical abuse, results remained unchanged except for one difference in the analysis of anger identification. Specifically, the significant emotion blend×physical aggression interaction became non-significant, *F*(1,87)=3.41, *p*=.068, $\eta_p{}^2$=0.04, 90% CI [0.00, 0.12].

Controlling for exposure to violence

To test whether exposure to violence played a role in physically aggressive individuals' task performance, analyses were re-run with the addition of Exposure to Violence Scale (ETV; Selner-O'Hagan, Kindlon, Buka, Raudenbush, & Earls, 1998) total score (*z* scored) as a covariate. When controlling for exposure to violence, results remained unchanged except for two differences. First, in the anger identification analysis, the significant emotion blend×physical aggression interaction became non-significant, *F*(1,85)=2.69, *p*=.105, $\eta_p{}^2$=0.03, 90% CI [0.00, 0.11]. Second, in the drift rate analysis, the significant emotion blend×physical aggression interaction became non-significant, *F*(1,85)=3.56, *p*=.063, $\eta_p{}^2$=0.04, 90% CI [0.00, 0.13].

Moderation by target race

To test whether target race affected task performance, the anger identification GLM analysis was re-run with the addition of target race as a within-subjects factor. Follow-up simple interaction contrasts, with White as the reference category, were used to yield the following comparisons: White versus Black targets and White versus Hispanic targets.

Crucially, we failed to detect any interactions between physical aggression and target race, justifying our choice to collapse across target race in our main analyses. Interestingly, we detected a significant main effect of target race, *F*(2,178)=35.43, *p*<.001, $\eta_p{}^2$=0.29, 90% CI [0.19, 0.36]. Examination of the simple interaction contrasts indicated that the White versus Black contrast was significant, *F*(1,89)=45.96, *p*<.001, $\eta_p{}^2$=0.34, 90% CI [0.21, 0.45], whereas the White versus Hispanic contrast was not, *F*(1,89)=0.06, *p*=.807, $\eta_p{}^2$=0.00, 90% CI [0.00, 0.03]. Mean anger identification scores indicated that participants were more likely to identify

White faces as angry (*M*=49.6%, 95% CI [48.2%, 50.9%]) compared to Black faces (*M*=44.5%, 95% CI [42.9%, 46.1%]); participants did not differ in anger identification for White and Hispanic (*M*=49.7%, 95% CI [48.1%, 51.3%]) faces. We also detected a significant dominant emotion×target race interaction, $F(2,178)=20.04$, $p<.001$, $\eta_p^2=0.18$, 90% CI [0.10, 0.26], indicating that Black faces were less likely to be identified as angry, particularly for mostly angry faces (*M* = 69.9%, 95% CI [67.2%, 72.5%], compared to 78.2%, 95% CI [75.9%, 80.5%], for White faces and 78.8%, 95% CI [67.2%, 72.5%] for Hispanic faces) and not mostly non-angry faces (*M* = 20.8%, 95% CI [18.7%, 22.9%], compared to 21.3%, 95% CI [18.8%, 23.8%], for White faces and 22.9% 95% CI [20.6%, 25.1%] for Hispanic faces).

Supplemental Discussion

Based on the supplemental results and consistent with previous research (Wilkowski & Robinson, 2012), the effects reported in the main results appear to be specific to physical aggression and not generalizable to other aggression-related constructs (i.e., verbal aggression, anger, hostility) or general aggression. Since physical aggression is overt and more likely to evoke anger and aggression from others, physically aggressive individuals may accumulate more experience that enhances their anger processing abilities. Moreover, since anger and hostility are not necessarily expressed outwardly, and since verbal aggression does not have to be direct (e.g., yelling at someone down the hall), physical aggression may be more relevant for face-to-face interactions (Wilkowski & Robinson, 2012).

Analyses further suggested that anxiety, childhood physical abuse, and exposure to violence may contribute to physically aggressive individuals' tendency to identify highly ambiguous faces as angry. However, the effect of physical aggression on drift rate for anger appeared robust to the influences of anxiety and physical abuse, whereas exposure to violence may help account for the effect of physical aggression on drift rate for anger. However, the relatively strong correlation between exposure to violence and physical aggression (i.e., *r*(86)=.45, see Table S3) highlights that it is unclear what remains of the physical aggression variable after partialling out the variance associated with exposure to violence (Lynam, Hoyle, & Newman, 2006). Future research should use longitudinal designs to parse the respective contributions of physical aggression and exposure to violence on anger processing. Overall, physical aggression appears to be more robustly related to drift rate for anger than to anger identification per se. This pattern suggests that our computational modeling approach allowed us to identify a mechanism that is more robustly linked to physical aggression than the anger identification variable yielded by traditional behavioral analysis methods (i.e., proportion of "angry" responses).

Finally, the analyses including target race indicated that target race did not interact with physical aggression to impact anger identification. However, target race impacted participants' anger identification overall. It is possible that the lower rates of anger identification seen for Black faces, particularly mostly angry faces, reflects a characteristic of the RADIATE stimulus set (Conley et al., 2018), since participants in the original validation study appeared to identify angry Black faces less accurately compared to White and Hispanic angry faces.

Supplemental References

Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., . . . Zule, W. (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse and Neglect, 27*, 169-190.

Conley, M. I., Dellarco, D. V., Rubien-Thomas, E., Cohen, A. O., Cervera, A., Tottenham, N., & Casey, B. J. (2018). The Racially Diverse Affective Expression (RADIATE) face stimulus set. *Psychiatry Research, 270*, 1059-1067. doi:https://doi.org/10.1016/j.psychres.2018.04.066

Lynam, D. R., Hoyle, R. H., & Newman, J. P. (2006). The perils of partialling: Cautionary tales from aggression and psychopathy. *Assessment, 13*, 328–341.

Pollak, S. D., & Kistler, P. (2002). Early experience is associated with the development of categorical representations for facial expressions of emotion. *Proceedings of the National Academy of Science, 99*, 9072-9076.

Pollak, S. D., Messner, M., Kistler, D. J., & Cohn, J. F. (2008). Development of perceptual expertise in emotion recognition. *Cognition, 110*, 242-247.

Pollak, S. D., & Sinha, P. (2002). Effects of early experience on children's recognition of facial displays of emotion. *Developmental Psychology, 38*, 784-791.

Selner-O'Hagan, M., Kindlon, D., Buka, S., Raudenbush, S., and Earls, F. (1998). Assessing exposure to violence in urban youth. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*, 215-224.

Welsh, G. S. (1956). Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine* (pp. 264-281). Minneapolis: University of Minnesota Press.

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.

Table S1

Sample Characteristics for Final Sample and Correlations among Key Variables

Variable	N	Mean	SD	Range	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1) Age	90	34.06	8.74	21-59	--													
2) Race					.16	--												
White	34																	
Black	54																	
Asian	1																	
American Indian	1																	
3) Ethnicity					-.08	-.35*	--											
Not Hispanic	73																	
Hispanic	17																	
4) BPAQ Physical Aggression	90	24.34	6.40	9-42	-.14	.02	.12	--										
5) RDEES Range	90	25.08	5.01	14-35	.09	-.05	.23*	.06	--									
6) RDEES Differentiation	90	25.42	5.26	11-35	.01	-.07	.14	.14	.45*	--								
7) Overall task accuracy	90	76.80%	5.91%	58.12-87.18%	-.08	-.19	.15	.16	.07	.21*	--							
8) Accuracy: AF predominantly angry	90	67.89%	14.36%	15.38-91.03%	-.20	-.18	.20	.25*	.15	.26*	.66*	--						
9) Accuracy: AF predominantly afraid	90	69.49%	10.19%	47.44-93.59%	.13	-.10	-.08	-.05	-.07	.05	.51*	-.05	--					
10) Accuracy: AH predominantly angry	90	79.50%	8.35%	48.72-97.44%	-.09	-.32*	.20	.07	-.01	.05	.44*	.30*	.15	--				
11) Accuracy: AH predominantly happy	90	81.78%	10.81%	46.15-97.44%	.01	.10	-.02	.11	.10	.11	.63*	.20	.32*	-.18	--			
12) Accuracy: FH predominantly afraid	90	83.30%	8.08%	57.69-98.72%	-.09	-.21*	.15	.05	-.10	-.09	.38*	.25*	.11	.45*	-.08	--		
13) Accuracy: FH predominantly happy	90	78.82%	11.23%	44.87-96.15%	.03	-.003	-.10	.02	.09	.22*	.65*	.24*	.28*	-.03	.66*	-.19	--	
14) Control task accuracy	90	90.14%	8.29%	31.48-98.15%	-.20	-.10	.14	.18	.03	.10	.49*	.41*	.31*	.25*	.21	.18	.23*	--

Note. Correlations including race and ethnicity used Spearman's ρ (all other correlations used Pearson's r); BPAQ = Buss Perry Aggression Questionnaire, RDEES = Range and Differentiation of Emotional Experiences Scale, AF = anger–fear blended faces, AH = anger–happiness blended faces, FH = fear–happiness blended faces. * $p < .05$

Table S2

Correlations among Diffusion Modeling Parameters

Parameter	1	2	3
1) Bias toward anger	--		
2) Drift rate toward anger	-.25*	--	
3) Threshold separation	-.03	.02	--

Note. * *p* < .05

Table S3

Correlations among Physical Aggression and Covariates

Variable	1	2	3	4
1) BPAQ Physical Aggression	--			
2) WAI	.25*	--		
3) ETV Total	.45*	.11	--	
4) CTQ Physical Abuse	.15	.04	.34*	--

Note. BPAQ = Buss Perry Aggression Questionnaire, WAI = Welsh Anxiety Inventory, ETV = Exposure to Violence, CTQ = Childhood Trauma Questionnaire. * *p* < .05

Appendix B: Study 2 Supplemental Material

Supplemental Method

Supplemental Measures

Shipley Institute of Living Scale (Zachary, 1986). The Shipley Institute of Living Scale is a measure of intelligence that consists of two subtests: vocabulary, a 40-item subtest in which participants choose a word (out of four options) that is synonymous with the word provided; and pattern matching, a 20-item subtest in which participants complete verbal and numerical patterns by writing in correct answers. Examiners convert raw scores on each subtest and then the total raw score to age-corrected T-scores. The total age-corrected T-score can then be used to estimate a participant's WAIS-R Full-scale IQ score, which has been shown to be an accurate means of predicting IQ (Weiss & Schell, 1991)**.**

Wide Range Achievement Test-III Reading Subtest (WRAT3 Reading Subtest; Wilkinson, 1993). The WRAT3 Reading Subtest is a measure of reading grade-level consisting of 42 items. Participants are instructed to pronounce a series of words aloud until they make 10 consecutive errors. If they make 10 errors, they are then asked to read a string of 15 letters aloud. Participants are awarded one point for each correctly pronounced word and letter, for a maximum score of 42. The final reading grade-level is determined by normed age adjustments.

Delis-Kaplan Executive Function System (D-KEFS): Color-Word Interference Test (CWIT; **Delis, Kaplin, & Kramer, 2001).** The D-KEFS comprises nine stand-alone tests designed to evaluate executive functions such as problem solving, inhibition, and flexibility of thinking. One of these nine tests is the CWIT, which consists of four conditions: Condition 1 (Color Naming), Condition 2 (Word Reading), Condition 3 (Inhibition; color names are written in ink colors that are different from the color itself), and Condition 4 (Inhibition/Switching;

participants must name the ink color when the word is not boxed and read the word when it is boxed). Participants are scored based on number of corrected and uncorrected errors that they make, as well as total time taken to complete each condition. Errors and total completion time for the condition are summed to create a raw score, which is then scaled (by age level) to determine Cumulative Percentile rank. We used the Inhibition/Naming contrast scaled score, which provides a measure of the difference between performance on Condition 3 (Inhibition) and performance on Condition 1 (Color Naming), as our measure of inhibition to identify potential contributions of executive functioning to the social reflection impulsivity-physical aggression relationship.

Multidimensional Personality Questionnaire–Brief (MPQ-B; Patrick, Curtin, & Tellegen, 2002). The MPQ-B is a shortened, 155-item measure adapted from the MPQ that assesses personality at the trait and structural levels. Participants respond to each of the 155 items by selecting one of two responses, typically "true" or "false." The MPQ-B consists of 12 primary trait scales, and three higher-order factor scores can be derived from the measure as well. One of these factor scores is Constraint, which is a measure of disinhibition and impulsivity (lower scores on Constraint indicate higher levels of disinhibition and impulsivity). We used the Constraint score as a measure of trait impulsivity to identify potential contributions of trait impulsivity to the social reflection impulsivity-physical aggression relationship. The MPQ-B is a reliable and valid measure of personality, and the Constraint score is a commonly used measure of impulsivity in research on aggression and externalizing behaviors (Kotov, Gamez, Schmidt, & Watson, 2010; Krueger et al., 2002).

Structured Clinical Interview for DSM-5 Disorders Substance Use Disorders Module (SCID-5 SUD; First, Williams, Karg, & Spitzer, 2015). The SCID-5 SUD was used to determine diagnoses of current (past year) or past (prior to the past year) alcohol use disorder (AUD) and substance use disorder (SUD). A diagnosis of AUD or SUD was given if at least two symptoms were present.

Psychopathy Checklist–Revised (PCL-R; Hare, 2003). The PCL-R is an interview-based measure that assesses 20 items related to psychopathic traits and behavior (e.g., glibness/superficial charm, shallow affect, impulsivity, poor behavior controls). Interviewers score each item from 0 to 2, with 0 indicating that the item does not apply to the individual, 1 indicating that the item applies to a certain extent, and 2 indicating that the item applies to the individual. Scores can range from 0 to 40, with higher scores indicating higher resemblance to a prototypical psychopath. Inter-rater reliability based on 21.42% of the sample was .98 for PCL-R total score. Information gathered as part of the PCL-R was used to assess Antisocial Personality Disorder (APD) symptoms based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for APD (American Psychiatric Association, 2013).

Reactive-Proactive Aggression Questionnaire (RPQ; Raine et al., 2006). The RPQ is a 23-item self-report measure designed to assess aggression according to motivations for engaging in aggressive acts. Participants rate each item on a scale of 0-2 (0="never," 1="sometimes," 2="often") based on how often they perform specific aggressive behaviors. The RPQ consists of two subscales: a reactive aggression subscale (11 items) and a proactive aggression subscale (12 items). Higher scores for each subscale indicate higher levels of reactive and proactive aggression, respectively. For this sample, good internal consistency was demonstrated for each subscale (Proactive Aggression Cronbach's α=.87, Reactive Aggression Cronbach's α=.79).

Need for Closure Scale (NCS; Kruglanski, Webster, & Klem, 1993). The NCS is a 42-item self-report questionnaire that measures an individual's motivated proclivity to obtain firm answers and to avoid ambiguity (Kruglanski et al., 1993; Webster & Kruglanski, 1996). Participants respond to each of 42 items using a 6-point Likert scale (1 ="strongly disagree" to 6="strongly agree"). The NCS consists of five subscales: desire for predictability (8 items), need for order (10 items), intolerance of ambiguity (9 items), decisiveness (7 items), and closed-mindedness (8 items). Internal consistency for the NCS was acceptable (Cronbach's α=.77).

Supplemental Results

Controlling for social information sampling task "accuracy"

In order to ensure that highly aggressive individuals' performance on the social information sampling task was not attributable to differences in "accuracy" (e.g., identifying someone as "nasty" when most of the behaviors they engaged in were negative), all primary analyses were run with "accuracy" (z-scored) included as a covariate. When controlling for "accuracy," all of the results remained the same. Crucially, even after controlling for "accuracy," the association between physical aggression and social reflection impulsivity remained, *B*=-1.46, *SE*=0.63, *p*=.023, 90% CI [-2.46, -0.62].

Controlling for executive functioning

In order to rule out inhibition (a major component of executive functioning) as a potential confound, all primary analyses were run with the D-KEFS CWIT Inhibition/Naming Contrast Scaled Score (z-scored) included as a covariate. When controlling for inhibition, all of the results remained the same except for one slight difference in the analysis of subjective certainty. Specifically, when controlling for inhibition, we failed to detect a condition×physical aggression interaction, $F(1,82)=3.27$, $p=.074$, $\eta_p^2=.04$, 90% CI [0.00, 0.13]. Crucially, however, even after controlling for inhibition, the association between physical aggression and social reflection impulsivity remained, *B*=-1.86, SE=0.76, *p*=.016, 90% CI [-3.05, -0.78].

Controlling for IQ and reading ability

In order to ensure that highly aggressive individuals' performance on the social information sampling task was not attributable to differences in IQ, which is inversely related to aggressive behavior (Séguin, Nagin, Assaad, & Tremblay, 2004), or reading ability, which may have impacted performance of the task since stimuli were read, all primary analyses were run with IQ (z-scored) and reading level (z-scored) included as covariates. When controlling for IQ, all of the results remained the same except for some slight differences in the analyses of frequency of social judgments and subjective certainty. First, in terms of frequency of social judgments, controlling for IQ introduced a main effect of physical aggression on social judgments, *F*(1,91)=4.40, *p*=.039, $\eta_p{}^2$=.05, 90% CI [0.001, 0.14], such that more physically aggressive individuals made *fewer* hostile social judgments, *B*=-0.04, *SE*=0.02, *p*=.038, 90% CI [-0.07, -0.01]. Second, in terms of subjective certainty, when controlling for IQ, we failed to detect a condition×physical aggression interaction, *F*(1,84)=3.09, *p*=.083, $\eta_p{}^2$=.04, 90% CI [0.00, 0.12]. The same exact pattern was found when controlling for reading ability (condition×physical aggression interaction: *F*(1,84)=3.71, *p*=.057, $\eta_p{}^2$=.04, 90% CI [0.00, 0.13]).

Controlling for trait impulsivity

In order to rule out trait impulsivity as a potential confound, all primary analyses were run with trait impulsivity (MPQ-B Constraint z-scored) included as a covariate. When controlling for trait impulsivity, all of the results remained the same. Crucially, even after controlling for trait impulsivity, the association between physical aggression and social reflection impulsivity remained, *B*=-1.80, SE=0.77, *p*=.022, 90% CI [-2.95, -0.71].

Controlling for substance use disorders

Based on research indicating that substance misuse is associated with heightened reflection impulsivity (Clark, Robbins, Ersche, & Sahakian, 2006; Clark, Roiser, Robbins, & Sahakian, 2009) and frequently co-occurs with aggressive behavior (Garofalo & Wright, 2017; Krueger, Markon, Patrick, Benning, & Kramer, 2007), we wanted to ensure that the relationship between physical aggression and social reflection impulsivity was not attributable to substance misuse. To do this, linear regression analyses were run with substance use disorder diagnosis (a dummy variable indicating whether the participant has ever in his lifetime met criteria for a substance use disorder) included as a covariate. In these analyses, the association between physical aggression and social reflection impulsivity remained, *B*=-1.92, *SE*=0.74, *p*=.011, 90% CI [-3.15, -0.70]. The association also remained after controlling for reflection impulsivity in the non-social task as well (i.e., both SUD diagnosis and non-social reflection impulsivity were entered as covariates), *B*=-0.99, *SE*=0.40, *p*=.016, 90% CI [-1.66, -0.32].

Controlling for Psychopathy and Antisocial Personality Disorder symptoms

In order to ensure that highly aggressive individuals’ performance on the social information sampling task was not attributable to Psychopathy or APD symptoms (both of which are robustly associated with aggressive behavior; Hare & McPherson, 1984; Raine, Lencz, Bihrle, LaCasse, & Colletti, 2000), all primary analyses were run with PCL-R total score (z-scored) and APD symptom count (z-scored) included as covariates, respectively. When controlling for Psychopathy, all of the results remained the same except for some slight differences in the analyses of frequency of social judgments and subjective certainty. First, in terms of frequency of social judgments, controlling for Psychopathy introduced a main effect of physical aggression on social judgments, *F*(1,91)=4.83, *p*=.030, η_p^2=.05, 90% CI [0.003, 0.15], such that more physically aggressive individuals made *fewer* hostile social judgments, *B*=-0.40, *SE*=0.02, *p*=.029, 90% CI [-0.07, -0.01]. The same pattern was found when controlling for APD symptoms, *F*(1,91)=5.94, *p*=.017, η_p^2=.06, 90% CI [0.01, 0.16], such that more physically aggressive individuals made *fewer* hostile social judgments when controlling for APD symptoms, *B*=-0.43, *SE*=0.02, *p*=.016, 90% CI [-0.07, -0.01]. Second, in terms of subjective certainty, when controlling for Psychopathy, we failed to detect a judgment×physical aggression interaction, *F*(1,84)=3.68, *p*=.058, η_p^2=.04, 90% CI [0.00, 0.13]. The same pattern was found when controlling for APD symptoms, *F*(1,84)=3.42, *p*=.068, η_p^2=.04, 90% CI [0.00, 0.13]).

Generalizability of findings to different types of aggression

In order to assess whether the findings reported here apply to general aggression (operationalized as the AQ Total score), the primary analyses were run with AQ Total score (z-scored) as the independent variable instead of the AQ Physical Aggression score. We failed to detect an association between general aggression and social reflection impulsivity, *B*=-0.16, *SE*=0.11, *p*=.132, 90% CI [-0.34, 0.02]. However, there was a judgment×general aggression interaction, *F*(1,88)=6.65, *p*=.012, η_p^2<.01, 90% CI [0.01, 0.17]. Specifically, participants with higher levels of general aggression demonstrated higher reflection impulsivity when they judged a person as nasty compared to when they judged a person a nice. However, we failed to detect a simple main effect of general aggression on reflection impulsivity in the context of hostile judgments, *B*=-1.44, *SE*=0.76, *p*=.061, 90% CI [-2.70, -0.18], as well as in the context of benign judgments, *B*=-0.46, *SE*=0.85, *p*=.592, 90% CI [-1.88, 0.96]. Finally, the 2 (condition: partial information, full information) × 2 (judgment: nasty, nice) repeated measures GLM revealed a main effect of general aggression on certainty, such that more aggressive individuals endorsed higher certainty about their social judgments, $F(1,84)=7.03$, $p=.010$, $\eta_p^2=.08$, 90% CI [.01, .18]. We failed to detect any other effects in the remaining analyses.

In order to assess whether the findings reported here apply to different forms of aggression (reactive and proactive aggression), the primary analyses were run with RPQ Reactive Aggression (z-scored) and RPQ Proactive Aggression (z-scored) scores as the independent variables. Regression analyses failed to detect associations between proactive aggression and social reflection impulsivity, $B=-0.17$, $SE=0.11$, $p=.121$, 90% CI [-0.34, 0.01], and between reactive aggression and social reflection impulsivity, $B=-0.11$, $SE=0.11$, $p=.301$, 90% CI [-0.29, 0.07].[8] We also failed to detect a judgment×aggression interaction (with social reflection impulsivity as the DV) for both proactive and reactive aggression. However, in terms of certainty, there were main effects of both proactive aggression, $F(1,84)=6.96$, $p=.010$, $\eta_p^2=.08$, 90% CI [0.01, 0.18], and reactive aggression, $F(1,84)=4.16$, $p=.044$, $\eta_p^2=.05$, 90% CI [0.001, 0.14], on certainty, such that more proactively aggressive individuals and more reactively aggressive individuals endorsed higher certainty about their social judgments overall. Furthermore, both proactive aggression, $F(1,84)=6.59$, $p=.012$, $\eta_p^2=.07$, 90% CI [0.01, 0.17], and reactive aggression, $F(1,84)=4.40$, $p=.039$, $\eta_p^2=.05$, 90% CI [0.001, 0.14], interacted with condition to predict certainty. For proactive aggression, more proactively aggressive individuals were more certain in the full condition compared to the partial condition (simple main effect of proactive aggression in the full information condition: $B=10.39$, $SE=4.24$, $p=.001$, 90% CI [5.18, 15.59], simple main effect of proactive aggression in the partial information condition: $B=3.67$,

[8] We also examined residualized scores on RPQ Reactive Aggression (variance associated with RPQ Proactive Aggression partialled out) and RPQ Proactive Aggression (variance associated with RPQ Reactive Aggression partialled out) as independent variables. Results were similar for residualized reactive and proactive aggression scores.

SE=2.79, *p*=.192, 90% CI [-0.96, 8.30]). The same pattern of results was found for reactive aggression: more reactively aggressive individuals were more certain in the full information condition compared to the partial information condition (simple main effect of reactive aggression in the full information condition: *B*=9.18, *SE*=3.17, *p*=.005, 90% CI [3.90, 14.45], simple main effect of reactive aggression in the partial information condition: *B*=4.31, *SE*=2.78, *p*=.124, 90% CI [-0.30, 8.93]). In addition, there was a condition×judgment×reactive aggression interaction, *F*(1,84)=8.92, *p*=.004, η_p^2=.10, 90% CI [0.02, 0.20], such that more reactively aggressive participants were more certain when making hostile judgments in the full information condition, *B*=14.67, *SE*=4.10, *p*=.001, 90% CI [7.85, 21.48].

Role of need for closure in aggressive individuals' judgment certainty

Since the finding that physical aggression was related to greater certainty in the full (versus partial) information condition was unexpected, we conducted a follow-up analysis to determine whether a potential third variable might shed light on the relationship (MacKinnon, Krull, & Lockwood, 2000). We reasoned that need for closure, an individual's motivation to have a firm understanding and avoid ambiguity (Kruglanski & Webster, 1996), might contribute to an individual's sense of certainty about a decision, particularly when they are under the impression that they have all of the information relevant for making the decision. We noted that this construct was positively associated with physical aggression in the present sample (see Supplemental Table 1 in the Supplemental Material), making it a potential candidate for elucidating physically aggressive individuals' performance on the social information sampling task. After adding the NCS Total score as a covariate in the GLM, we failed to detect a condition×physical aggression interaction, *F*(1, 84)=2.58, *p*=.112, η_p^2=.03, 90% CI [0.00, 0.11], suggesting that a need for closure may be implicated in heightening physically aggressive individuals' certainty in the full information condition. The judgment×physical aggression interaction, on the other hand, remained, $F(1, 84)=3.99$, $p=.049$, $\eta_p^2=.05$, 90% CI [0.0001, 0.13].

Supplemental Discussion

Across supplemental analyses, the effects reported in the main analyses are largely specific to physical aggression, with few effects seen for general aggression and subtypes of aggression (i.e., reactive and proactive). Furthermore, the effects reported in the main analyses are robust, with little impact of task "accuracy," executive functioning, IQ, reading ability, trait impulsivity, substance use disorders, Psychopathy, and APD symptoms. Therefore, the social information sampling task appears to tap into deficits that are highly specific to physical aggression, rather than being associated with other forms of aggression, psychiatric diagnoses, or decrements in IQ or reading ability.

In addition to examining the robustness and specificity of the main analyses, we also explored whether need for closure affects the relationship between aggression and certainty in the context of varying levels of available information. Broadly speaking, heightened need for closure represents a stronger need to reduce ambiguity in decision-making. More specifically, in the present study, heightened need for closure appears to intensify the conviction with which aggressive individuals seize upon social judgments when they are under the impression that they have all of the relevant information about a person. Heightened need for closure can lead to confident but misguided social judgments because in the real world it is never possible to have all of the information needed for judging another person (e.g., the thoughts and intentions of others are never completely knowable and thus always retain some degree of ambiguity). Taken together, the results of this supplemental analysis indicate that heightened need for closure may strengthen the sense of certainty aggressive individuals feel when judging others, and future research should more directly examine the role of need for closure in aggressive individuals' social decision-making and social behavior.

Supplemental References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington: American Psychiatric Publishing.

Clark, L., Robbins, T. W., Ersche, K. D., & Sahakian, B. J. (2006). Reflection impulsivity in current and former substance users. *Biological Psychiatry, 60*(5), 515-522.

Clark, L., Roiser, J., Robbins, T., & Sahakian, B. (2009). Disrupted reflection impulsivity in cannabis users but not current or former ecstasy users. *Journal of Psychopharmacology, 23*(1), 14-22.

Delis, D. C., Kaplin, E., & Kramer, J. (2001). *Delis Kaplin Executive Function System*. San Antonio Texas, TX: The Psychological Corporation.

Evenden, J. (1999). The pharmacology of impulsive behaviour in rats V: The effects of drugs on responding under a discrimination task using unreliable visual stimuli. *Psychopharmacology, 143*(2), 111-122.

First, M. B., Williams, J., Karg, R. S., & Spitzer, R. L. (2015). *Structured Clinical Interview for DSM-5-Research Version*. Arlington, VA: American Psychiatric Association.

Garofalo, C., & Wright, A. G. (2017). Alcohol abuse, personality disorders, and aggression: The quest for a common underlying mechanism. *Aggression and Violent Behavior, 34,* 1-8.

Hare, R. D. (2003). *Manual for the Revised Psychopathy Checklist* (2nd ed.). Toronto, Ontario, Canada: Multi-Health Systems.

Hare, R. D., & McPherson, L. M. (1984). Violent and aggressive behavior by criminal psychopaths. *International Journal of Law & Psychiatry, 7*(1), 35-50. doi:http://dx.doi.org/10.1016/0160-2527(84)90005-0

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking “big” personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin, 136*(5), 768.

Krueger, R. F., Hicks, B. M., Patrick, C. J., Carlson, S. R., Iacono, W. G., & McGue, M. (2002). Etiologic connections among substance dependence, antisocial behavior and personality: Modeling the externalizing spectrum. *Journal of Abnormal Psychology, 111*(3), 411-424. doi:10.1037/0021-843X.111.3.411

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*(4), 645-666. doi:10.1037/0021-843x.116.4.645

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: "Seizing" and "freezing.". *Psychological Review, 103*(2), 263-283. doi:10.1037/0033-295X.103.2.263

Kruglanski, A. W., Webster, D. M., & Klem, A. (1993). Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of Personality and Social Psychology, 65*(5), 861.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, 1*(4), 173-181. doi:10.1023/a:1026595011371

Patrick, C. J., Curtin, J. J., & Tellegen, A. (2002). Development and Validation of a Brief Form of the Multidimensional Personality Questionnaire. *Psychological Assessment, 14*(2), 150-163. doi:10.1037//1040-3590.14.2.150

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., . . . Liu, J. (2006). The reactive–proactive aggression questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior, 32*, 159-171.

Raine, A., Lencz, T., Bihrle, S., LaCasse, L., & Colletti, P. (2000). Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. *Archives of General Psychiatry, 57*(2), 119-127.

Séguin, J. R., Nagin, D., Assaad, J.-M., & Tremblay, R. E. (2004). Cognitive-neuropsychological function in chronic physical aggression and hyperactivity. *Journal of Abnormal psychology, 113*(4), 603-613.

Webster, D., & Kruglanski, A. W. (1996). Motivated Closing of the Mind. *Psychological Review, 103*, 263-283.

Weiss, J. L., & Schell, R. E. (1991). Estimating WAIS-R IQ from the shipley institute of living scale: A replication. *Journal of Clinical Psychology, 47*(4), 558-562.

Wilkinson, G. S. (1993). *WRAT3: The Wide Range Achievement Test Administration Manual* (3rd ed.). Wilmington, DE: Wide Range, Inc.

Zachary, R. A. (1986). *Shipley Institute of Living Scale: Revised Manual*. Los Angeles, CA: Western Psychological Services.

Supplemental Table 1

Sample Characteristics for Final Sample and Correlations among Key Variables

Variable	N	Mean	SD	Range	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1) Age	93	34.73	10.28	21-59	--																		
2) Race					.14	--																	
White	44																						
Black	49																						
3) Ethnicity					-.14	-.26*	--																
Not Hispanic	72																						
Hispanic	21																						
4) IQ	93	104.12	11.52	77-122	-.04	-.20	-.14	--															
5) Reading level	93	10.56	1.74	6-12	-.10	-.22*	-.11	.66*	--														
6) AQ Physical Aggression	93	24.62	6.66	11-38	-.23*	.22*	.24*	-.22*	-.17	--													
7) AQ Total	93	73.72	14.73	44-115	-.19	.14	.13	-.17	-.17	.76*	--												
8) RPQ Proactive Aggression	93	5.22	4.38	0-17	-.21*	.05	.27*	-.15	-.10	.66*	.54*	--											
9) RPQ Reactive Aggression	93	10.78	3.71	3-20	-.21*	.05	.23*	.03	.01	.65*	.66*	.72*	--										
10) Assault charges	93	1.65	3.23	0-20	.14	.06	.08	-.09	-.16	.17	.19	.04	.12	--									
11) PCL-R Total	93	22.83	7.15	6.3-35	.12	.26*	.12	-.09	.13	.41*	.24*	.34*	.25*	.27*	--								
12) APD symptoms	93	3.73	1.75	0-7	.08	.14	.04	-.17	.05	.38*	.29*	.35*	.30*	.14	.72*	--							
13) Lifetime SUD diagnosis					-.20	.03	.11	-.22*	-.09	.19	.00	.33*	.20	-.15	.25*	.32*	--						
No diagnosis	16																						
Diagnosis	77																						
14) NCS Total	93	169.97	16.49	129-207	-.08	.03	.03	.01	.12	.22*	.31*	.15	.27*	.12	.16	.07	.07	--					
15) Boxes opened social task	93	16.07	7.14	1-25	-.11	-.14	.05	.31*	.29*	-.23*	-.15	-.14	-.09	-.20	-.04	-.10	.09	-.01	--				
16) Boxes opened non-social task	93	17.12	7.41	1-25	-.09	-.13	-.02	.31*	.31*	-.16	-.17	-.10	-.07	-.26*	.02	-.02	.13	-.03	.85*	--			
17) Accuracy non-social task	93	.78	.16	.35-1.0	-.02	-.21*	-.04	.38*	.38*	-.15	-.19	-.09	-.06	-.19	-.11	.02	.01	.04	,51*	.63*	--		
18) Hostile judgments	93	.53	.17	.1-1.0	-.22*	-.11	.16	-.06	.13	-.12	-.10	.03	.03	-.11	-.02	.05	.25*	.08	.18	.22*	.25*	--	
19) Certainty	93	38.17	26.33	-23.45-94.90	.02	-.06	-.06	-.02	-.09	.10	.22*	.21*	.21	.16	.10	.20	-.07	.06	.02	.08	.21*	.06	--

Note. Correlations including race, ethnicity, and SUD diagnosis used Spearman's *ρ* (all other correlations used Pearson's *r*); AQ = Buss Perry Aggression Questionnaire, RPQ = Reactive Proactive Aggression Questionnaire, SUD = substance use disorder, IQ = WAIS IQ estimate from the Shipley Institute of Living Scale, Reading level = reading level from the Wide Range Achievement Test 3 Reading Subtest, PCL-R = Psychopathy Checklist—Revised, APD = Antisocial Personality Disorder, NCS = Need for Closure Scale. * $p < .05$

Supplemental Table 2

Criminal Charges for Final Sample

Crime type	N (% of sample) charged with crime type	Mean charge count	SD	Range
Violent	89 (95.7%)	6.39	7.51	0-35
Weapon	56 (60.2%)	1.81	2.73	0-14
Assault	49 (52.7%)	1.65	3.23	0-20
Murder	35 (37.6%)	1.32	0.66	0-10
Robbery	27 (29.0%)	1.15	2.84	0-15
Sex	24 (25.8%)	2.66	0.95	0-20
Kidnapping	8 (8.6%)	0.18	0.74	0-5
Non-violent	80 (86.0%)	10.75	17.48	0-105
Theft	48 (51.6%)	3.82	11.99	0-101
Escape	47 (50.5%)	1.06	1.77	0-12
Obstruction of justice	45 (48.4%)	1.22	1.90	0-10
Drug	44 (47.3%)	1.56	2.59	0-12
Negligence/driving	31 (33.3%)	0.58	0.97	0-4
Fraud	10 (10.8%)	1.02	7.07	0-66
Crimes against state	1 (1.1%)	0.02	0.21	0-2
Miscellaneous minor	44 (47.3%)	1.47	2.98	0-23

Supplemental Table 3

Institutional Infractions for Final Sample

Infraction type	N (% of sample) charged with infraction type	Mean infraction count	SD	Range
Threats to security	45 (48.4%)	1.20	1.98	0-10
Violations against persons	43 (46.2%)	1.68	3.32	0-19
Substance use violations	16 (17.2%)	0.21	0.48	0-2
Violations against property	14 (15.1%)	0.16	0.40	0-2
Other	54 (58.1%)	3.24	4.59	0-21
Total (any infraction)	70 (75.3%)	7.46	12.59	0-93

Appendix C: Study 3 Supplemental Material

Supplemental Results

Covariate Analyses

To test whether Conduct Disorder (CD) symptoms contributed to physically aggressive individuals' task performance, analyses were re-run with the addition of CD symptoms (assessed via interview using DSM-5 criteria for CD) as a covariate. When controlling for CD symptoms, all physical aggression-related effects remained significant.

To test whether symptoms of Antisocial Personality Disorder (APD) contributed to physically aggressive individuals' task performance, analyses were re-run with the addition of APD symptoms (assessed via interview using DSM-5 criteria for APD) as a covariate. When controlling for APD symptoms, all physical aggression-related effects remained significant.

To test whether psychopathic traits contributed to physically aggressive individuals' task performance, analyses were re-run with the addition of Psychopathy Checklist-Revised (Hare, 2003) total score as a covariate. When controlling for psychopathic traits, all physical aggression-related effects remained significant.

To test whether substance use disorders (SUDs) contributed to physically aggressive individuals' task performance, analyses were re-run with the addition of the total number of lifetime SUD diagnoses (assessed using the Structured Clinical Interview for DSM-5; First, Williams, Karg, & Spitzer, 2015) as a covariate. When controlling for SUD diagnoses, all physical aggression-related effects remained significant.

To test whether anxiety contributed to physically aggressive individuals' task performance, analyses were re-run with the addition of Welsh Anxiety Inventory (WAI; Welsh, 1956) total score as a covariate. When controlling for anxiety, all physical aggression-related effects remained significant.

To test whether negative emotionality contributed to physically aggressive individuals' task performance, analyses were re-run with the addition of the Negative Emotional Temperament score from the Multidimensional Personality Questionnaire-Brief (MPQ-BF; Patrick, Curtin, & Tellegen, 2002) as a covariate. When controlling for negative emotionality, one physical aggression-related effect was reduced to non-significance. Specifically, the dominant emotion × IJT × emotion decision × physical aggression interaction in the analysis of confidence was no longer significant, $F(1,71)=1.81$, $p=.183$, $\eta_p^2=0.03$, 90% CI [0.00, 0.11].

To test whether exposure to violence contributed to physically aggressive individuals' task performance, analyses were re-run with the addition of the Exposure to Violence Scale (ETV; Selner-O'Hagan, Kindlon, Buka, Raudenbush, & Earls, 1998) total score as a covariate. When controlling for exposure to violence, all physical aggression-related effects remained significant.

To test whether childhood trauma contributed to physically aggressive individuals' task performance, analyses were re-run with the addition of the Childhood Trauma Questionnaire-Short Form (CTQ-SF; Bernstein et al., 2003) total score as a covariate. When controlling for childhood trauma, all physical aggression-related effects remained significant.

Associations between Violent Institutional Infractions and Task Performance

It can be useful to examine whether the effects of self-reported physical aggression extend to non-self-reported indicators of physical aggression. The number of violent institutional infractions an individual has been cited for while incarcerated represents an alternative way of measuring physical aggression. We re-ran analyses, replacing AQ Physical Aggression as the independent variable with violent institutional infractions (recorded directly from institutional documents). We log-transformed the violent institutional infraction variable due to its right skewness and added the number of years the individual had been incarcerated as a covariate to each model, to control for the amount of time the individual had spent in the institution. Analyses revealed no significant associations between violent institutional infractions and task performance.

Supplemental Discussion

Based on the supplemental results, the physical aggression-related effects reported in the main results appear to be quite robust, with no discernable impact of CD symptoms, APD symptoms, psychopathic traits, SUDs, anxiety, exposure to violence, or childhood trauma. Negative emotionality was the only variable that impacted an association between physical aggression and task performance. Negative emotionality may, in part, account for physically aggressive individuals' tendency to lose confidence over time after they misidentify mostly angry faces as happy. This finding is consistent with research indicating that individuals who are prone to experiencing various negative emotions show enhanced processing of anger- and threat-related information (Parrott, Zeichner, & Evces, 2005; Reed & Derryberry, 1995). It is possible that negative emotionality represents an underlying predisposition toward more effective postdecisional processing of social threat. These findings highlight the importance of further research into the role of negative emotionality in postdecisional processing of emotional information.

We did not detect any associations between violent institutional infractions and task performance, representing a divergence from the significant effects observed for self-reported physical aggression. Several factors may account for the measurement-based divergence. First, whereas the self-report measure of physical aggression indexes individuals' endorsement of how characteristic certain physically aggressive behaviors are of them, violent institutional infractions represent the total number of violent acts individuals committed (and were charged with) over the course of their incarceration. There are several ways in which self-reported physical aggression may not align with violent institutional infractions. As one example, individuals who were highly aggressive when younger, were incarcerated for a long period, and then experienced a marked decline in physical aggression may have accumulated many institutional infractions throughout their sentence but self-report lower physical aggression because these behaviors are no longer seen as “characteristic" of them. Second, the self-report measure captures individuals’ perceptions of themselves, while violent institutional infractions capture an observer’s report and are influenced by a distinct set of environmental factors such as the level of monitoring on a particular unit (Steiner & Wooldredge, 2008). Finally, the self-report measure is not a count of aggressive behaviors, but rather reflects the characteristicness of certain physically aggressive behaviors in particular contexts (e.g., following provocation); however, violent institutional infractions do not account for such contextual considerations. Taken together, self-reported physical aggression and violent institutional infractions represent two different perspectives on an individual’s level of physical aggression. Thus, it is not particularly surprising that we did not detect associations between violent institutional infractions and task performance. However, future research should aim to distill valuable information from both types of measures (e.g., using latent variable approaches).

Supplemental References

Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., . . . Zule, W. (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse and Neglect, 27,* 169-190

First, M. B., Williams, J., Karg, R. S., & Spitzer, R. L. (2015). *Structured Clinical Interview for DSM-5-Research Version*. Arlington, VA: American Psychiatric Association.

Hare, R. D. (2003). *Manual for the Revised Psychopathy Checklist* (2nd ed.). Toronto, Ontario, Canada: Multi-Health Systems.

Kosterman, R., Graham, J. W., Hawkins, J. D., Catalano, R. F., & Herrenkohl, T. I. (2001). Childhood risk factors for persistence of violence in the transition to adulthood: A social development perspective. *Violence and Victims, 16*, 355-369. https://doi.org/10.1891/0886-6708.16.4.355

Parrott, D. J., Zeichner, A., & Evces, M. (2005). Effect of trait anger on cognitive processing of emotional stimuli. *Journal of General Psychology, 132*(1), 67-80. https://doi.org/10.3200/GENP.132.1.67-80

Patrick, C. J., Curtin, J. J., & Tellegen, A. (2002). Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment, 14*(2), 150-163. https://doi.org/10.1037//1040-3590.14.2.150

Reed, M. A., & Derryberry, D. (1995). Temperament and attention to positive and negative trait information. *Personality and Individual Differences, 18*(1), 135-147. https://doi.org/https://doi.org/10.1016/0191-8869(94)00121-8

Selner-O'Hagan, M., Kindlon, D., Buka, S., Raudenbush, S., and Earls, F. (1998). Assessing exposure to violence in urban youth. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*, 215-224.

Steiner, B., & Wooldredge, J. (2008). Inmate versus environmental effects on prison rule violations. *Criminal Justice and Behavior, 35*(4), 438-456. https://doi.org/10.1177/00938548073127878

Welsh, G. S. (1956). Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine* (pp. 264-281). Minneapolis: University of Minnesota Press.

Table S1

Information on Relevant Variables in Final Sample

Variable	*M*	*SD*	Range	% of Sample Meeting Criteria for Disorder	Correlation with AQ Physical Aggression	Physical Aggression Effects Reduced to Non-significance after Controlling for the Variable
Conduct Disorder symptoms	4.60	3.50	0-13	64.00%	.301*	None
Antisocial Personality Disorder symptoms	3.75	1.63	0-7	53.30%	.387*	None
PCL-R total	24.76	6.09	8-36	20.00%	.313*	None
Substance use disorder diagnoses	1.95	1.33	0-5	90.41%	.316*	None
WAI total	12.38	8.50	1-38	N/A	.146	None
MPQ-BF NEM	47.55	16.91	17-90	N/A	.498*	Dominant emotion × IJT × emotion decision × physical aggression interaction in analysis of confidence
ETV total	8.92	3.04	1-13	N/A	.413*	None
CTQ-SF total	48.14	17.89	25-103	N/A	-.039	None

Note. AQ=Buss Perry Aggression Questionnaire, PCL-R=Psychopathy Checklist–Revised , WAI=Welsh Anxiety Inventory, MPQ-BF NEM=Negative Emotional Temperament score on the Multidimensional Personality Questionnaire–Brief, ETV=Exposure to Violence scale, CTQ=Childhood Trauma Questionnaire-Short Form. *p<.05