Distribution Agreement

Hyung Seo (Caroline) Lee April 12, 2022

The Influence of Psychopathic Traits and Affective Feedback on Cooperation in the Iterated Prisoner's Dilemma

by

Hyung Seo (Caroline) Lee

Dr. Andrew Kazama

Adviser

Department of Psychology

Dr. Andrew Kazama

Adviser

Dr. Stephan Hamann

Committee Member

Dr. Umberto Mignozzetti

Committee Member

Dr. Michal Arbilly

Non-Voting Committee Member

2022

The Influence of Psychopathic Traits and Affective Feedback on Cooperation in the Iterated Prisoner's Dilemma

By

Hyung Seo (Caroline) Lee

Dr. Andrew Kazama

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Psychology

2022

Abstract

The Influence of Psychopathic Traits and Affective Feedback on Cooperation in the Iterated Prisoner's Dilemma

By Hyung Seo (Caroline) Lee

Consistent with interdependence theory (Kelley & Thibaut, 1978; Van Lange et al., 2013), a growing body of behavioral research has examined the interactive influence of structural, psychological, dynamic interaction processes on social cooperation with the Prisoner's Dilemma paradigm (PD; Luce & Raiffa, 1957). Accordingly, the PD paradigm has increasingly served as a framework for examining this interactive influence on social cooperation. Significant impairment in interpersonal-affective processes is a central characteristic of psychopathy in both clinical and non-clinical populations. However, little is known about the interaction of psychopathic traits and affective processes in the interpersonal context of social cooperation. To address this gap in the literature, the current study examined the individual and interactive influence of psychopathy, affective feedback congruence, and Stroop interference on cooperation in PD. A secondary goal was to analyze the relation between psychopathy and emotion perception, as well as the presence of a Dunning-Kruger effect (Kruger & Dunning, 1999) for emotion recognition ability. A total of 237 undergraduate and Prolific participants were recruited for this study. Consistent with previous research on reverse appraisal, the congruence of affective feedback in PD was associated with increased rate and expectation of cooperation across a 20-round iterated PD with a computerized opponent using a tit-for-tat strategy. Psychopathic Personality Inventory – Revised (PPI-R; Lilienfeld & Andrews, 1996) Machiavellian Egocentricity and Blame Externalization were negatively associated with cooperation in PD; additionally, a significant three-way interaction emerged between affective feedback congruence, Stroop Interference, and two PPI-R measures (Stress Immunity and Coldheartedness) for predicting the expectation of cooperation in PD. The relation between actual and estimated emotion recognition ability was assessed with pairwise comparisons by tercile and quartile, as well as with mixed-model regression analyses, yielding patterns that are partially consistent with previous research on the Dunning-Kruger effect. Due to a small sample size, additional research is needed to increase confidence in the validity of the findings for the current study. Future research should replicate the study with human opponents in the modified iterated PD to increase ecological validity, and control for age and gender when exploring the Dunning-Kruger effect for emotion recognition ability.

The Influence of Psychopathic Traits and Affective Feedback on Cooperation in the Iterated Prisoner’s Dilemma

By

Hyung Seo (Caroline) Lee

Dr. Andrew Kazama

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Psychology

2022

Acknowledgements

I would first like to thank Dr. Scott Lilienfeld, my late advisor, and Dr. Andrew Kazama, my current thesis advisor, for their unwavering confidence in me and their continued support in my development as an undergraduate researcher. I am beyond grateful for their support and guidance. I would also like to thank my committee members for their helpful feedback throughout this process from proposal to completion.

Next, I would like to express my deepest appreciation for and gratitude to my graduate student mentor Tom Costello. This thesis would not have been possible without his supportive mentorship. I would also like to thank my other graduate student mentor, Shauna Bowes, for her kind support and mentorship throughout the past three years.

Finally, I would like to thank my family, friends, Doldol, and Tori for their loving support.

TABLE OF CONTENTS

Figures

Appendices

Introduction

Psychopathy represents a constellation of affective (e.g., overall shallow affect, lack of remorse), interpersonal (e.g., superficial charm, egocentricity), and behavioral (e.g., irresponsibility, impulsivity) traits in both clinical and community samples (Cleckley, 1941; Lilienfeld & Andrews, 1996). Contrary to early conceptualizations of psychopathy (psychopathic personality) as a categorical construct (e.g., Harris et al., 1994), a large body of recent research has shown that psychopathy comprises an interpersonally detrimental configuration of continuously distributed trait dimensions (Edens et al., 2006; Lilienfeld et al., 2019).

Highly psychopathic individuals, who might match the archetypal "psychopath" seen in popular media (e.g., individuals with extensive and severe criminal histories) are quite uncommon (approximately 1% of the general population; Hare, 2003). However, levels of psychopathic traits can be reliably detected in non-clinical samples with dimensional measures such as the Psychopathic Personality Inventory – Revised (PPI-R; Lilienfeld & Widows, 2005) and the Levenson Self-Report Psychopathy Scale (LSRP; Levenson et al., 1995). The PPI-R comprises two higher-order dimensions (Fearless Dominance and Self-Centered Impulsivity), as well as an additional dimensional subscale (Coldheartedness) that is independent of the two dimensions. On the other hand, the LSRP comprises two factors: Factor 1 (F1), which captures the interpersonal and affective components of psychopathy, and Factor 2 (F2), which captures behavioral components of psychopathy related to social deviance. By and large, studies have shown associations between high levels of psychopathic traits as measured by PPI-R and LSRP with maladaptive outcomes, including instrumental and reactive aggression (Long et al., 2014), impaired emotional processing (Gordon et al., 2004), and atypical social reward processing (Foulkes, 2015).

Significant impairment in interpersonal relations is central to the conceptualization and maladaptive behavioral outcomes associated with personality disorders more broadly and psychopathy more specifically (see Wilson et al., 2017, for a review). Specifically, the ability to recognize and infer intentions from social signals, such as facial emotional expressions, is critical to the formation and maintenance of interpersonal relations (Blair, 2003). Hence, to better understand maladaptive behavioral outcomes associated with psychopathy, it is important to examine the relations between individual differences in emotion recognition and psychopathy in an interpersonal context.

Studies examining the relation between emotion recognition ability and psychopathy, however, have yielded mixed findings. On the one hand, some studies have found global deficits in emotion recognition in individuals high in psychopathy, supporting Cleckley's (1941) influential account of psychopathy as including pervasive affective detachment (e.g., Dawel et al., 2012; White et al., 2012). On the other hand, some studies have also shown that individuals high in psychopathy do not show global deficits in emotion recognition but exhibit abnormal selective attention (which leads to affective deficits for goal-peripheral stimuli), supporting the attention bottleneck hypothesis of psychopathy (Baskin-Sommers et al., 2011; e.g., Hiatt et al., 2004; Kranefeld & Blickle, 2022). However, studies have yet to clarify these conflicting findings by examining the relation between psychopathy and the processing of affective information in an interpersonal context.

To address this gap in the literature, the current study uses the Prisoner's Dilemma paradigm (PD; Luce & Raiffa, 1957) to investigate the individual and interactive influence of reverse appraisal (i.e., inferring intentions from affective feedback by comparing the congruence of the affective feedback to the decision in an interpersonal context; de Melo et al., 2014), Stroop interference (i.e., measured by comparing task performance in trials or conditions with congruent and incongruent stimuli; Stroop, 1935), and psychopathy on social cooperation. The PD paradigm has served as a useful framework for exploring the relation between cooperation and individual differences in social preferences (see Thielmann et al., 2020, for a review) as well as attentional and emotional processing (Bell et al., 2016; Gabay et al., 2019), and have only grown in precision and usefulness as a growing body of research has incrementally proposed improvements (see Van Dijk & De Dreu, 2021, for a review). More specifically, the paradigm allows researchers to model social cooperation in a "social decision cascade" comprising three sequential phases in each round: (1) the *decision* phase, in which two players decide to cooperate or defect, (2) the *anticipation* phase, in which players wait to view the interdependent outcome, and (3) the *feedback* phase, in which the outcome is revealed (Thompson et al., 2021). Structural influences of the paradigm (e.g., arrangement of the payoff matrix, length of the game, etc.) shape participants' behavioral responses, but so do their psychological characteristics (e.g., individual differences in personality, affect, reward responsiveness, etc.) and dynamic interaction processes (e.g., appraisal of the opponent's strategy in relation to one's own motives, etc.) (Van Lange et al., 2013).

Additionally, I will conduct auxiliary analyses on the presence of a Dunning-Kruger effect (i.e., the general tendency for lowest performers to overestimate performance and for the highest performers to underestimate performance, showing a curvilinear trend of confidence in relation to ability; Kruger & Dunning, 1999) for emotion recognition ability independent of psychopathy, and assess whether psychopathy explains significant variance in the participant's estimates of their emotion recognition ability. If psychopathy is associated with a global deficit in emotion recognition, and patterns consistent Dunning-Kruger effect emerge such that participants with greater deficits in emotion recognition ability and higher levels of psychopathy show significant overestimation of emotion recognition ability, this may influence the inference of intentions from assessing the congruence of affective feedback (i.e., engaging in reverse appraisal) for participants with higher levels of psychopathy in the iterated PD task. To date, few studies have examined the influence of individual differences in personality on the Dunning-Kruger effect (e.g., John & Robins, 1994), and only one study (Ammirati, 2013) has examined the influence of personality specifically on the Dunning-Kruger effect for emotion recognition ability. Therefore, these auxiliary analyses will be exploratory in nature.

Prisoner's Dilemma Task

Given my emphasis on the PD in the present investigation, I will now elaborate on specific elements of the paradigm. The PD paradigm, which is based on mathematical game theory, evaluates social cooperation in an interactive two-person task. In each round of the PD, both players make a simultaneous decision to either cooperate or defect; one-shot PD consists of one round, whereas iterated PD consists of more than one round. Researchers can also implement a sequential one-shot PD, wherein players would play a series of one round of PD against different players. After player A and player B settle on a decision for the round, the outcome is revealed to both players. There are four possible outcomes in PD: CC (mutual cooperation), DD (mutual defection), CD (cooperation by A and defection by B), and DC (defection by A and cooperation by D). Specific payoffs are assigned for each outcome, and may differ for each payoff matrix as long as it satisfies two conditions: S (payoff for CD) < P (payoff for DD) < R (payoff for CC) < T (payoff for DC), and 2R > S + T (Rapoport & Chammah, 1965). As T > R, a player would always receive a higher payoff by defecting when the opponent cooperates; as P > S, a player would always receive a higher payoff by defecting when the opponent defects.

Therefore, as players cannot communicate at any point during the game, participants would always do better by defecting in a one-shot PD. However, achieving the highest payoff by defecting when the opponent cooperates and receives the lowest payoff creates a social dilemma wherein the player would be "betraying" the cooperative opponent in yielding the best outcome for oneself.

This dilemma is amplified in an iterated PD (spanning multiple rounds) as a player must engage repeatedly with the same opponent (Axelrod & Hamilton, 1981). In an iterated PD, unlike the one-shot PD, the constraint of $2R > S + T$ ensures that continuous selection of R (mutual cooperation) would be preferred over the alternation between S and T (Rapoport & Chammah, 1965).

Influence of Emotions in PD

As outlined above, players do not receive or provide any form of communication throughout the entire game in a standard PD procedure. Even in the absence of communication, however, cooperative decision-making in PD against a human or computerized opponent is modulated by various factors, including emotions (Angelika-Nikita et al., 2021), reward processing (Wood et al., 2006), and individual differences in social value orientation (Pletzer et al., 2018). The modulating influence of emotions in PD is especially important to consider for studies assessing individual difference variables that may affect emotion-driven decision-making (e.g., psychopathy) in the context of repeated social interactions, such as the iterated PD.

The influence of emotions in PD has been supported by a growing body of research on the neurobiological underpinnings of decision-making processes in iterated PD. Perhaps unsurprisingly, studies probing the *decision phase* of the iterated PD (i.e., the first phase of each iterated PD round wherein players make the decision to cooperate or defect; Thompson et al.,

2021) have identified significant activation in distinct brain regions for cooperation and defection. Most notably, activation in the orbitofrontal cortex have been linked to the decision to cooperate, suggesting that cooperation is the predominant emotional response in iterated PD; the decision to defect, on the other hand, has been primarily linked to activation in the dorsolateral prefrontal cortex, suggesting that defection in iterated PD demands deliberate exertion of cognitive effort to strategically obtain rewards (Emonds et al., 2014; Rilling et al., 2007). This is consistent with a large body of literature in game theory on the evolution of cooperation, which theorizes that cooperative behavior is rewarded in iterated PD via potential benefits of reciprocal altruism over the course of the extended game (Axelrod, 1980a, 1980b; Trivers, 1971).

With regard to the following phase (i.e., the *anticipation phase*, in which players wait to view the interdependent outcome; Thompson et al., 2021), anticipation following cooperation has been commonly associated with activation of anterior insula, pointing to affective responses and exertion of cognitive control elicited by the uncertainty of the opponent either reciprocating cooperation and yielding the ideal outcome (i.e., CC) or betraying the player to reap the best payoff while subjecting the opponent to the worst payoff (i.e., CD) (e.g., Haroush & Williams, 2015). Anticipation following the decision to defect has been commonly linked to regions involved in social and emotional conflict resolution, supporting a broad consensus in the literature that choosing to defect in an iterated PD is perceived to be strategically desirable but costly in repeated social interactions such as the iterated PD, hence establishing a social dilemma (e.g., Thompson et al., 2021).

Lastly, studies examining neural activation in the *feedback phase* (i.e., the last phase of each iterated PD round wherein the outcome is revealed; Thompson et al., 2021) have identified correlates with each of the four possible outcomes in the Prisoner's Dilemma. Consistent with the general expectation that mutual cooperation should be the most optimal outcome, mutual cooperation (CC) has been associated with activation of areas linked with reward processing (Rilling et al., 2002). Receiving an outcome of unreciprocated cooperation (CD), on the other hand, has been associated with areas implicated in cognitive control, emotion regulation, and processing of aversive emotions (Rilling et al., 2008). Viewing the outcome of DC (i.e., having defected while the opponent cooperated in the previous round) has been commonly linked with activation of the dorsolateral prefrontal cortex, which has been interpreted as indicative of increased cognitive demand in regulating the negative emotion of guilt induced by having not reciprocated the opponent's cooperation (Edmiston et al., 2015; Gradin et al., 2016). As the outcome of mutual noncooperation (DD) yields a relatively low and equal payoff for both players, it has been commonly linked with low activation across a broad range of emotions (Gradin et al., 2016; Rilling et al., 2008). Studies comparing activation following reciprocated (CC/DD) and unreciprocated (CD/DC) feedback have also shown that participants find unreciprocated feedback more aversive (e.g., finding associations with heightened activity in the precuneus and striatal deactivation; Rilling et al., 2002; Thompson et al., 2021). Such results are consistent with findings in the literature that a general human tendency and preference for reciprocity in the iterated PD paradigm influences the rate of cooperation (Fehr et al., 2002; Kujala & Danielsbacka, 2019), although the prevalence and adaptiveness of strong reciprocity in cases of repeated social interactions apart from the PD is debated (Hammerstein, 2003).

Indeed, research has shown that reciprocity can significantly promote cooperation and consequently result in high overall payoff in iterated PD games, as evidenced by the robust success of tit-for-tat as a strategy. Tit-for-tat, which emerged as the most successful strategy in Axelrod's (1980a) round-robin tournament with 14 other programmed strategies in 200 repeated games, consists of beginning the game with cooperation and replicating the opponent's decision in the previous round thereafter. Subsequent analyses have attributed tit-for-tat's success to four characteristics: being nice, provocable, forgiving, and clear (Axelrod, 1980a, 1980b). Tit-for-tat is a *nice* strategy in signaling cooperative intent by always beginning with cooperation, a *provocable* strategy in not being exploited by always reciprocating opponent's defection with defection, a *forgiving* strategy in returning to cooperation by reciprocating opponent's cooperation even if they have defected previously, and a *clear* strategy in that it has a consistent pattern. Continued research on a vast array of strategies for the iterated PD have affirmed the advantage that reciprocity can provide as a strategy characteristic (Axelrod & Hamilton, 1981). However, cooperative and altruistic behavior can arise without reciprocity (i.e., competitive altruism; Roberts, 1998), and reciprocity may interact with other conditions and characteristics in affecting cooperation (e.g., preferences for smaller amounts of immediate rewards over larger amounts of rewards over a longer period of time, also known as temporal discounting; Stephens et al., 2002). Overall, the stability and level of cooperation in an iterated PD conducted between programmed strategies depend on multiple strategic characteristics.

In considering cooperative behavior in a human–human or human–computer game of iterated PD as opposed to a computer tournament, however, a variety of factors apart from strategic factors also influence decision-making. As explained previously, interdependence theory, proposed by Van Lange and colleagues (2013), suggests that a complex interplay of structural, psychological, and dynamic interaction processes contribute to decision-making in social dilemmas. To assess the interplay of such processes in interpersonal decision-making, researchers have begun to add modifications to the social dilemma paradigms, including the PD (see Van Dijk & De Dreu, 2018, for a review).

To further explore affective processes throughout the decision, anticipation, and feedback phases of PD, facial emotional feedback has been added across recent studies as a modification to the standard PD paradigm. Although there is a wealth of literature supporting the communicative function of facial emotional feedback in social interactions (e.g., Xu et al., 2013), the Emotion as Social Information (EASI) model (Van Kleef et al., 2009) stands out as a theoretical framework via which researchers can systematically parse the interpersonal function of emotional expressions.

According to the EASI model, observing another person's emotional expression in a social interaction can affect the observer's behavior via two pathways: (1) the *inferential pathway*, in which the emotional expression of the other individual allows the observer to infer their mental states, including beliefs, desires, and intentions (i.e., engage in mentalizing) and (2) the *affective pathway*, in which the emotional expression of the other individual elicits affective reactions in the observer.

In the context of modified social dilemma paradigms in which affective feedback (i.e., facial emotional expressions) is presented alongside the outcome in the feedback phase (i.e., each player's previous decision and the resulting payoff), there is an additional layer to this inferential pathway. In addition to the other player's emotional expression, their decision in contrast to the player's decision is presented in conjunction with one another as information from which to infer their mental state. This specific process has been termed *reverse appraisal* (de Melo et al., 2014; Hareli & Hess, 2010; Scherer & Grandjean, 2008).

In a recent study, de Melo and colleagues (2014) examined reverse appraisal in the iterated PD paradigm by testing if the affective feedback paired with the CC and CD outcomes would enable the players to infer their mental state (specifically, their cooperative intent) in informing their decision-making via reverse appraisal. Accordingly, as the CC outcome allows both players to take the second-highest payoff while the CD outcome results in the highest payoff for the opponent while subjecting the player to the lowest payoff, de Melo and colleagues (2014) hypothesized that players would be able to infer cooperative intent from receiving the affective feedback of a happy facial expression paired with the outcome of CC, whereas they would infer competitive intent when a regretful/guilty expression is paired with the outcome. For the outcome of CD, it was hypothesized that the reverse would be true (i.e., competitive intent would be inferred from a happy facial expression, whereas cooperative intent would be inferred from a regretful/guilty expression). Indeed, in a cooperative context of mutual cooperation (i.e., CC), the opponent's facial expression of joy led to increased expectations of cooperation, whereas the expression of regret led to decreased expectations of cooperation. As hypothesized, the opponent's facial expression of regret in a competitive context (i.e., CD) led to increased expectations of cooperation; however, the opponent's expression of joy in this context did not lead to decreased expectations of cooperation (de Melo et al., 2014). One possible explanation for this finding suggested by the authors is that the expression of joy in the competitive context may not have led the participants to infer a mental state reflective of an intentionally non-cooperative or malicious intent (e.g., experiencing schadenfreude in the opponent having received a worse payoff).

Psychopathy

Psychopathic personality is characterized by a constellation of interpersonal, affective, antisocial, and lifestyle features. Although varying conceptualizations of psychopathy exist, the construct operationalized by the Psychopathic Personality Inventory – Revised (PPI-R; Lilienfeld & Widows, 2005) has been widely adopted by personality researchers to assess psychopathy in both community and prison samples. The PPI-R, developed to capture affective and interpersonal features central to psychopathy, comprises three higher-order dimensions: Fearless Dominance, Self-Centered Impulsivity, and Coldheartedness (Lilienfeld & Widows, 2005). Interpersonal features (e.g., manipulation, dishonesty; Patrick, 2018) and the associated affective features (e.g., fearlessness, guiltlessness; Lilienfeld et al., 2014) are especially important in conceptualizing psychopathy, as these features distinguish psychopathy from related constructs. Broadly, psychopathy can be understood as an emergent interpersonal syndrome (Lilienfeld et al., 2019), characterized by prototypes of interpersonal disturbances (Viding, 2019). Studies exploring neural correlates of psychopathy have supported these findings, with psychopathic traits showing associations with hypoactivity in areas implicated in interpersonal decision making, including but not limited to the orbitofrontal cortex (Yang & Raine, 2008), dorsolateral prefrontal cortex (Koenigs, 2014), and the amygdala (Blair, 2007).

Psychopathy and Emotion

Both historical and modern accounts of psychopathy have included affective features as central components of conceptualizing the construct. Following Cleckley's seminal (1976) work, which includes "general poverty in major affective reactions" (p. 348) as a defining feature of psychopathy, studies have found associations between psychopathy and more specific indices of emotional functioning, including emotion recognition ability and emotional reactivity (see Nentjes et al., 2022, for a review).

However, there is ongoing debate in the literature regarding a vast array of these indices, which attests to the complexity of conceptualizing psychopathy. With regard to emotion recognition ability, studies have largely focused on the link between deficits in distress (i.e., fear and sadness) recognition and psychopathy (e.g., Marsh & Blair, 2008); however, meta-analytic findings have suggested that deficits in emotion recognition ability for individuals high in psychopathy may be more pervasive, including facial and vocal expressions of all six basic emotions (i.e., anger, disgust, fear, happiness, sadness, and surprise; Dawel et al., 2012). Research on the relation between psychopathy and emotional reactivity has also resulted in mixed findings, with studies pointing to specific deficits in negative emotional reactivity (e.g., Lykken, 1957), deficits in both positive and negative emotional reactivity (e.g., Blair, 2006), or neither (e.g., Shane & Groat, 2018).

With the emergence of the response modulation hypothesis (RMH; Gorenstein & Newman, 1980) and the attention bottleneck model of psychopathy (Baskin-Sommers et al., 2011), studies have also explored the interplay of attentional mechanisms and emotional functioning in relation to psychopathic traits. According to the RMH, emotional and self-regulatory deficits associated with psychopathy are moderated by response modulation deficits (i.e., deficits in the ability to shift attention to peripheral affective information or take such information into account while engaging in goal-directed behavior; Patterson & Newman, 1993). Specifying the RMH, the attention bottleneck model of psychopathy argues that individuals with higher levels of psychopathy display reduced processing of goal-peripheral information at an early perceptual processing stage, hence experiencing an early attentional "bottleneck" (Baskin-Sommers et al., 2011).

Psychopathy and Cooperation

In recent years, a growing body of research has assessed the influence of psychopathic traits on social cooperation using the PD paradigm, as it is well-suited for assessing interpersonal-affective features central to the conceptualization of psychopathy in a dynamic interpersonal context. The PD task has also been utilized to assess cooperation in relation to various psychological disorders, including antisocial personality disorder (Rada et al., 2003), borderline personality disorder (Bartz et al., 2011), social anxiety disorder (Rodebaugh et al., 2013), autism spectrum disorder (Kaartinen et al., 2019), and depression (Gradin et al., 2016).

The first study to assess the impact of psychopathy on social cooperation with PD was conducted by Widom (1976) in a maximum security hospital in England. The analysis revealed no significant differences in cooperative responses in a 30-round iterated PD between the control group and groups of individuals identified as primary or secondary psychopaths based on the criteria outlined by Cleckley (1941) and Hare (1970). Unexpectedly, the primary psychopathy group had the highest average frequency of mutually cooperative strings in the PD; the secondary psychopathy group had a significantly lower average frequency, but only marginally lower than that of the control group.

Subsequently, studies have utilized PD games of varying lengths to examine the relation between psychopathy and social cooperation, using the Levenson Self-Report Psychopathy Scale (LSRP; Levenson et al., 1995) and versions of the Psychopathy Personality Inventory (PPI; Lilienfeld, 1990; Lilienfeld & Andrews, 1996) to operationalize psychopathy. In male undergraduates separately analyzed in a mixed-gender sample of undergraduate participants, the overall frequency of cooperation in a 20-round iterated PD game against a forgiving-tit-for-tat opponent correlated negatively with LSRP Total and Factor 1 scores, but not LSRP Factor 2 or PPI scores (Rilling et al., 2007). Consistent with Rilling et al. (2007), Baggio & Benning (2022) found no significant associations between PPI-R subscale scores and the rate of cooperation or defection across nine 30-round iterated PD games conducted with a computerized opponent using tit-for-tat and eight other strategies with varying degrees of leniency.

Supporting previous work on the impact of affective feedback in PD (e.g., Reed et al., 2012), Johnston and colleagues (2014) found a negative correlation between LSRP Factor 1 scores and cooperation when PD was accompanied by affective feedback, but not in the absence of affective feedback. In a study by Gervais and colleagues (2013), undergraduate participants participated in one unannounced one-shot PD game with human opponents who they had a chance to have unstructured conversations with for 10 minutes. The results showed that participants with higher LSRP Factor 1 scores defected more frequently with opponents who interrupted them more often during the conversation (Gervais et al., 2013).

Contrary to findings in Rilling et al. (2007), several subsequent studies have found significant associations between PPI-R measures and cooperation in iterated and sequential one-shot PD games. In a sample of high-security psychiatric patients, PPI-R Impulsive Nonconformity and Machiavellian Egocentricity—subscales of PPI-R Self-Centered Impulsivity—were positively associated with the overall rate of defection in a 40-round iterated PD against a tit-for-two tats opponent (Mokros et al., 2008). In sequential one-shot PD games, PPI-R Machiavellian Egocentricity was also negatively associated with the initiation and reciprocation of cooperation (Curry et al., 2011). Additionally, in a sample of undergraduate participants, the overall frequency of defection in a 10-round iterated PD against a tit-for-tat opponent was not only associated with LSRP (Total, Factor 1, and Factor 2) scores, but also PPI-R Self-Centered Impulsivity and Coldheartedness (Berg et al., 2013). However, in a study by Testori et al. (2019), the higher-order dimension of PPI-R Fearless Dominance was negatively associated with the rate of cooperation in a 30-round iterated PD against a tit-for-two-tats opponent, whereas no significant association was found between cooperation PPI-R Self-Centered Impulsivity or Coldheartedness.

Current Study

The primary purpose of the present study was to examine the individual and interactive influence of psychopathy, affective feedback congruence, and Stroop interference on cooperation in PD. A secondary goal for the present study was to conduct exploratory analyses on the relation between psychopathy and emotion perception, as well as the influence of psychopathy on the Dunning-Kruger effect for emotion recognition ability should the effect emerge. In the following subsections, the specific aims and hypotheses for the present study are outlined in greater detail.

Psychopathy and Emotion Perception

Despite a wealth of research on affective features as central characteristics in operationalizing psychopathy, findings regarding the relation between psychopathy and emotion recognition ability remain opaque. This is due to a continued debate regarding whether psychopathy is associated with deficits specifically in distress (i.e., fear and sadness) recognition (Marsh & Blair, 2008), pervasive deficits in recognition of vocal and facial expressions of all six basic emotions (Dawel et al., 2012), or deficits in emotion recognition specifically when the affective information provided is not included in the goal-relevant attentional set (Baskin-Sommers & Newman, 2014). In one study analyzing personality dimensions in relation to self-reported estimates of confidence in one’s emotion recognition skill and the presence of a Dunning-Kruger effect (i.e., the general tendency for lowest performers to overestimate performance and for the highest performers to underestimate performance, showing a curvilinear trend of confidence in relation to ability; Kruger & Dunning, 1999), narcissism and neuroticism were significantly associated with higher and lower comparative estimates of emotion recognition ability respectively (Ammirati, 2013). However, no study to date has assessed the relation between psychopathy and self-reported estimates of emotion recognition ability.

Hypothesis 1a: I predict that actual and estimated emotion recognition ability will be weakly correlated, but consistent with the Dunning-Kruger effect.

Hypothesis 1b: Furthermore, I predict that psychopathy will explain significant variance in the participant's estimates of their emotion recognition ability. Specifically, higher levels of psychopathy will be positively associated with estimates of emotion recognition ability.

Individual Difference and Interpersonal Influences on Cooperation in PD

In the second part of the present study, the individual and interactive influence of psychopathy, affective feedback congruence, and Stroop interference on cooperation in PD were examined. The specific aims and hypotheses for the present study are outlined below.

Affective Feedback Congruence. Emerging research on the influence of emotions in decision-making in PD have shown that receiving affective feedback from the opponent can allow the participant to (1) infer the opponent's appraisal of the outcome for each round (i.e., engage in reverse appraisal) and (2) adjust one's own decision and affective feedback according to an updated expectation of the opponent's cooperation in future rounds (de Melo et al., 2014). In a recent study by de Melo and colleagues, players receiving cooperative affective feedback showed increased rates and expectations of cooperation on average, whereas players receiving competitive affective feedback showed decreased rates and expectations of cooperation on average (de Melo & Terada, 2020). This finding suggests that players were able to engage in reverse appraisal, wherein they inferred the opponent's intentions by comparatively assessing whether their decision to cooperate or defect in each round were congruent with their affective feedback (de Melo & Terada, 2020).

Hypothesis 2a: I predict that the rate and expectation of cooperation in PD for participants assigned to the cooperative/congruent affective feedback conditions (wherein the computerized opponent expresses joy following mutual cooperation and regret following exploitation) will be higher compared to those of participants assigned to competitive/incongruent affective feedback conditions (wherein the computerized opponent expresses regret following mutual cooperation and joy following exploitation).

Psychopathy. Across a small body of studies that have examined the influence of psychopathic traits on the rate of cooperation or defection in PD, measures of psychopathy have shown negative associations with the rate of cooperation or positive associations with the rate of defection. More specifically, mixed findings have separately shown significant negative associations between the rate of cooperation in PD and PPI-R Fearless Dominance (Testori et al., 2019), PPI-R Self-Centered Impulsivity (Berg et al., 2013; Curry et al., 2011; Mokros et al., 2008), PPI-R Coldheartedness (Berg et al., 2013), LSRP Total (Berg et al., 2013; Rilling et al., 2007), and LSRP Factor 1 (Berg et al., 2013; Gervais et al., 2013; Johnston et al., 2014; Rilling et al., 2007), and LSRP Factor 2 (Berg et al., 2013) scores.

Hypothesis 2b: Given mixed findings across studies examining the link between psychopathy and the rate of cooperation in PD with variations in measures of psychopathy, strategy of the opponent, and length of the PD games, my hypothesis is exploratory in nature. As I will be using the 40-item PPI-R for measuring the level of psychopathy, and since multiple studies have found a significant negative relation between cooperation in PD and scores on subscales subsumed under the dimension of PPI-R Self-Centered Impulsivity, I predict that the rate and expectation of cooperation in PD will be negatively associated with PPI-R Self-Centered Impulsivity and its subscales.

Stroop Interference. The Stroop effect (Stroop interference; Stroop, 1935) is a well-validated attentional phenomenon measured by comparing task performance in trials or conditions with Stroop-incongruent stimuli (i.e., simultaneous presentation of one type of stimuli presenting task-relevant information and another type of stimuli presenting task-irrelevant information, creating conflict) and Stroop-congruent stimuli (i.e., simultaneous presentation of two types of stimuli presenting task-relevant information). Therefore, the verbal-facial Stroop effect for the affective feedback employed in this study may interfere with the processing of information from the affective feedback congruence to infer the opponent's intentions (i.e., by engaging in reverse appraisal).

Hypotheses 2c: I predict that the rate and expectation of cooperation in PD for participants assigned to the cooperative x Stroop-incongruent affective feedback condition will be lower compared to those of participants assigned to the cooperative x Stroop-congruent affective feedback condition. Furthermore, I predict that the rate and expectation of cooperation in PD for participants assigned to the competitive x Stroop-incongruent affective feedback condition will be higher compared to those of participants assigned to the competitive x Stroop-congruent affective feedback condition.

Interactive Influence of Psychopathy, Affective Feedback Congruence, and Stroop Interference. Since reverse appraisal is a form of Theory of Mind (ToM) reasoning (Gratch & de Melo, 2019), and as studies have shown associations between ToM impairment and psychopathy (e.g., Crick & Dodge, 1994; but see Richell et al., 2003), individuals with higher levels of psychopathy may experience more difficulty utilizing information regarding the opponent's intention via reverse appraisal of cooperative/congruent vs. competitive/incongruent affective feedback provided following mutual cooperation and exploitation in the absence of

Stroop interference. Consistent with the attentional bottleneck hypothesis (Baskin-Sommers et al., 2011), studies utilizing variations of the Stroop task have also shown that psychopathy is associated with reduced Stroop interference (Hiatt et al., 2004; Strohmaier, 2015). These results support the possibility that participants assigned to cooperative x Stroop-incongruent affective feedback and competitive x Stroop-incongruent affective feedback conditions with higher levels of psychopathy may experience less Stroop interference compared to participants in these conditions with lower levels of psychopathy.

Hypothesis 2d: I predict that there will be a 3-way interaction between psychopathy, affective feedback congruence, and Stroop interference, such that the interaction of affective feedback congruence and Stroop interference will differ based on the level of psychopathy. More specifically, I predict that participants receiving cooperative and competitive feedback will show decreasing rates and expectations of cooperation at higher levels of psychopathy, with a larger interaction effect in conditions with Stroop interference.

Method

Participants

Participants were recruited from an undergraduate sample (*n*=169) and a community sample (*n*=100) recruited via Prolific Academic, an online crowdsourcing research platform for participant recruitment and data collection. In both samples, participants were required to be at least 18 years of age to be eligible for participation.

Undergraduate participants enrolled in introductory psychology courses for the Spring 2022 semester were recruited via the Emory University Psychology Department's Student Subject Pool via the SONA system, which allows undergraduate participants to identify and access research opportunities for course credit compensation. Repeat submissions were not allowed, and completing other studies or complete a writing assignment were offered as alternative options for completing the class requirement. Data from 21 participants who failed an implicit attention check item adopted from Oppenheimer et al. (2009) were excluded from analyses. However, all participants received a compensation of 1 class credit for completion of the study.

The community sample was recruited via Prolific Academic, an online crowdsourcing research platform for participant recruitment and data collection. Prolific Academic was selected for this study, as studies have shown higher ratings on multiple indices of data quality (e.g., response rate, naivety, diversity) for Prolific when compared to other major crowdsourcing research platforms (e.g., Amazon Mechanical Turk, CrowdFlower; Palan & Schitter, 2018; Peer et al., 2017). Data from 11 participants who failed an implicit attention check item adopted from Oppenheimer et al. (2009) were excluded from analyses. However, all participants were financially compensated for participation of the study via the platform ($10.41/hour, *M* = 21.32 minutes).

Data from the two aforementioned participant pools were combined for all analyses. The resulting final sample (*n*=237) was predominantly female (68.35%), with an age range of 18 to 56 years (*M* = 21.5, *SD* = 5.52). Participants' ethnic and racial background were primarily White (38.4%), followed by Asian (25.74%), Hispanic (15.19%), Mixed (10.55%), Black or African-American (7.59%), Middle Eastern or North African (0.84%), and other (0.84%), with 2 (0.84%) participants declining to answer. The sample size for the final sample was slightly below the lower end of sample size recommended for achieving stable estimates of correlations (*n*=250; Schönbrodt & Perugini, 2013), and much below the sample size estimated by an a priori power analysis with the *InteractionPoweR* package in R (Baranger, 2021) for detecting effects in interaction models with a power of 80% (approximately *n*=600). However, the final sample size was ultimately determined based on funding and undergraduate participant availability for this study.

Procedure

All procedures were approved by Emory University's Institutional Review Board as compliant with ethical research standards. The study was programmed using jsPsych, a JavaScript framework for behavioral tasks (de Leeuw, 2015). Upon successful registration, participants accessed the study via Cognition, a platform for hosting jsPsych experiments. Prior to beginning the experimental tasks, participants were asked to confirm their understanding of the informed consent and filled out a brief demographic form including basic biographical information on age, gender, race, and ethnicity. Participants were fully debriefed upon completing the experiment, exiting the experiment, or declining to proceed after viewing informed consent.

Measures

Facial Emotion Recognition Task. Before completing the Facial Emotion Recognition Task (FERT; Passarelli et al., 2018), participants were asked to complete the non-verbal Self-Assessment Manikin (SAM; Bradley & Lang, 1994)—a picture-oriented, self-report assessment of one's affective state in which participants rate their current state of affective valence (on a five-point scale ranging from *1-happy* to *5-unhappy*) and arousal (on a five-point scale ranging from *1-excited* to *5-calm*)—prior to and after watching a 1-minute clip of a suburban street view. This video is coded in the Open Library for Affective Videos (OpenLAV; Israel et al., 2021) database as a neutral, non-emotional video, and was shown to the participants to settle distracting emotions before the start of an emotion recognition task, as recommended by Wingenbach et al. (2016).

The 36-item pool of stimuli for the FERT comprises 6 basic emotions (anger, disgust, fear, happiness, sadness, surprise) displayed by 6 Caucasian actors (Passarelli et al., 2018). Following the procedure validated by Passarelli and colleagues (2018), each of the 36 emotional stimuli were randomly shown to participants in conjunction with a neutral picture of the same actor as a reference. For each pair of stimuli presented, participants were asked to select the emotion displayed by the actor (with the options being the six basic emotions: anger, disgust, fear, happiness, sadness, surprise). As the two-parameter logistic (2PL) model for the FERT (test score reliability ρ = .92; Raykov et al., 2010) assigned specific item parameters for each of the 36 items, the composite FERT facial emotion recognition ability score for each participant was computed using a Hamiltonian Monte Carlo sampler in the package *rstan* (Guo et al., 2021) for R software version 4.1.2. The R script for this analysis was provided in the Supplementary Materials for Passarelli et al. (2018), and modified to be compatible with the format of the data input.

Iterated Prisoner's Dilemma. After completing the FERT, participants were randomly assigned by the Cognition platform to one of the four between-subjects conditions for the type of affective feedback used in a 20-round iterated PD task. Deception was involved, as participants were informed that they would be competing with another participant in real-time; previous studies have found that the perception of playing against a human opponent as opposed to a computer opponent elicits higher engagement in interactive games such as the iterated PD (Kätsyri et al., 2013). To aid in this deception, a loading screen was presented throughout the task to imply delays via the online connection. Before starting the task, participants were informed that an ID (“Anonymous16”) and a facial character of another individual have been assigned to them to protect their anonymity throughout the task. Static and high-intensity facial emotional expressions by the M06 encoder from the Amsterdam Dynamic Facial Expression Set – Bath Intensity Variations (ADFES-BIV; Wingenbach et al., 2016) were used to represent the participant’s facial character in conveying affective feedback. The opponent’s ID was always “Anonymous35”, with static pictures of facial emotional expressions by the M08 encoder in the ADFES-BIV serving as the facial character.

Similar to previous studies (e.g., de Melo et al., 2020; Kulms et al., 2014), the iterated PD task was recast as a two-person investment game in which investing in ‘project green’ represented cooperation, and investing in the ‘project blue’ represented defection. Thorough instructions and explanations were provided prior to the task without time limits, and the payoff matrix (see Figure 1) remained visible throughout the duration of the task. The instructions noted that the payoff matrix would remain on the screen, and participants proceeded by confirming their understanding of the task.

In each round, the participant began by either pressing G on the keyboard to invest in ‘project green’ or B to invest in ‘project blue’. The loading screen appeared before the opponent’s decision was displayed. The opponent’s decision was programmed to follow the tit-for-tat strategy, wherein cooperation is met with cooperation and defection is met with defection (Axelrod, 1980; McClure et al., 2007). The opponent always cooperated (i.e., chose project green) in the first round regardless of whether the participant chose to cooperate or defect (i.e., chose project blue). In rounds 2-18, the opponent replicated the participant’s decision in the previous round. In the last two rounds, the opponent always defected, as this is most consistent with human behavior (Rilling et al., 2007).

After the participant made their decision for the round and the loading screen was shown, the outcome (including each player's decision and the resulting payoff) was displayed. Immediately after the outcome was shown for the round, the opponent's affective feedback was shown to the participants. Participants assigned to Conditions 1 and 3 always viewed cooperative affective feedback, whereas participants assigned to Conditions 2 and 4 always reviewed competitive affective feedback. Cooperative affective feedback consisted of the opponent always responding with joy (i.e., a happy facial expression or the verbal label of "happy") for the outcome of CC, regret for CD, anger for DC, and with a neutral expression for DD. Competitive affective feedback consisted of the opponent always responding with regret for CC, joy for CD, anger for DC, and with a neutral expression for DD. Facial expression provided in Conditions 1 and 2 had red verbal labels with incongruent emotions superimposed to test for the influence of Stroop interference. Following the receipt of the opponent's affective feedback, the participants were asked to convey their affective feedback by selecting a neutral, angry, happy, or regretful expression of their assigned character (i.e., facial expressions of the M06 encoder from the ADFES-BIV database). After the exchange of affective feedback, the participants rated their expectation of the opponent's cooperation in the following round.

Psychopathy. At the end of the experiment, participants were asked to complete the 40-item short-form of the Psychopathic Personality Inventory-Revised (PPI-R-40; Eisenbarth et al., 2015). The PPI-R-40 comprises abbreviated versions of eight subscales in the Personality Personality Inventory–Revised (PPI-R; Lilienfeld & Widows, 2005): Blame Externalization (e.g., "When I'm with people who do something wrong, I usually get the blame"), Rebellious Nonconformity (e.g., "I have always seen myself as something of a rebel"), Coldheartedness (e.g., "If someone is hurt by something I say or do, I usually consider that to be their problem"), Social Influence (e.g., “I feel sure of myself when I’m around other people”), Carefree Nonplanfulness (e.g., “I generally prefer to act first and think later”), Fearlessness (e.g., “I am a daredevil”), Machiavellian Egocentricity (e.g., “If I can’t change the rules, I try to get others to bend them for me”), and Stress Immunity (e.g., “I can remain calm in situations that would make many other people panic”). The PPI-R measure showed high internal consistency reliability overall (α = .80), with PPI-R higher-order dimensional subscales of Fearless Dominance and Self-Centered Impulsivity also showing good internal consistency reliability estimates (αs = .76 – .79). However, PPI-R Coldheartedness showed low internal consistency, with Cronbach’s alpha of .59.

Results

The descriptive statistics and intercorrelations between psychopathy and facial emotion recognition measures are presented in Tables 1 and 2. Comparisons of SAM ratings for arousal and valence indicated a statistically significant change in affective state prior to and after watching the 1-minute clip, with average ratings shifting to be closer to neutral (3) after watching the video for both arousal (before: $M = 2.59$, $SD = 0.99$, after: $M = 2.8$, $SD = 0.95$, $t(236) = -3.88$, $p < .001$) and valence of affective state (before: $M = 1.47$, $SD = 0.98$, after: $M = 1.58$, $SD = 0.87$, $t(236) = -2.07$, $p = 0.040$).

Psychopathy and emotional perception.

Overall, participants’ estimated and actual performance on the FERT as measured by the proportion of correct responses on the FERT were not significantly associated, $t(235) = 1.53$, $p = 0.13$, $r = 0.10$ (see Figure 4). This finding contrasted with Hypothesis 1a, which predicted that a small correlation would be found.

Consistency with the Dunning-Kruger effect was assessed by (1) conducting paired t-tests to compare average estimated vs. actual proportion of correct responses for FERT in the overall sample, terciles (lower, middle, upper), and quartiles (bottom, second, third, top) (see Table 4 and Figure 2), and (2) conducting multi-model regression analyses with linear, quadratic, and cubic components to test for curvilinearity (see Table 5 and Figure 3). Partial support was found for Hypothesis 1a, as explained in further detail below. However, participants' estimate of accuracy on the FERT exceeded actual performance on the FERT as measured by the proportion of correct responses (*M*s = 0.65 and 0.69, respectively, *t*(236) = 3.33, *p* = .001, η_p^2 = -0.216), and by the composite score calculated via the Bayesian 2PL procedure (*M*s = 0.65 and -0.92, respectively, *t*(236) = -32.44, *p* < 0.001, η_p^2 =2.11).

Before conducting paired t-tests, average proportions of correct responses for the FERT by tercile and quartile were computed. Pairwise comparisons for terciles revealed that participants in the lower tercile did not overestimate their performance on the FERT (*t*(94) = 1.73, *p* = 0.086), contrary to the predictions for the Dunning-Kruger effect. However, statistically significant overestimation was found in both middle and upper terciles (2nd tercile: *t(*65) = -0.06, *p* = 0.003; 3rd tercile: *t*(75) = -0.11, *p* < 0.001), consistent with trends of a Dunning-Kruger effect. Pairwise comparisons for quartiles revealed similar patterns, with participants in the bottom quartile not showing statistically significant overestimation of their performance on the FERT (*t*(72) = 1.61, *p* = .111) and participants in the second quartile also not showing statistically significant underestimation of their performance on the FERT (t(60) = -1.79, p = 0.078). However, consistent with predictions for the Dunning-Kruger effect, participants in upper quartiles underestimated their performance (third quartile: *t*(50) = -3.32, *p* < 0.001; top quartile: *t*(51) = -6.01, *p* < 0.001).

Mixed-model regression analyses were subsequently conducted to test for curvilinearity with linear, quadratic, and cubic components, as recommended by Sanchez & Dunning (2018). Standardized parameters were obtained by fitting the model on a standardized version of the dataset, and 95% Confidence Intervals (CIs) and p-values were computed using the Wald approximation. With the composite score calculated by the Bayesian 2PL procedure as the measure of participants' performance on FERT, a cubic model produced the best fit (R^2 = .043). In this cubic model, the linear and cubic trends were both found to be statistically significant (linear: *b* = 0.35, 95% CI [0.13, 0.56], *t*(233) = 3.20, *p* = 0.002; *β* = 0.35, 95% CI [0.13, 0.56]; cubic: *b* = -0.08, 95% CI [-0.14, -0.02], *t*(233) = -2.76, *p* = 0.006; *β* = -0.31, 95% CI [-0.53, -0.09])).

To assess relations between psychopathic traits and facial emotion recognition measures, zero-order correlations were computed (Table 3). The composite FERT score calculated with the Bayesian 2PL procedure was negatively related to PPI-R Machiavellian Egocentricity only (*r* = -0.18, *p* = .0097). Self-report estimates of accuracy for FERT was moderately positively related to PPI-R Fearless Dominance (*r* = 0.17, *p* = 0.0164), Stress Immunity (*r* = 0.18, *p* = 0.0078), and Rebellious Nonconformity (*r* = 0.14, *p* = 0.04). FERT subscales for Fear, Happiness, and Surprise showed moderate negative relations with PPI-R measures (*r*s ranged from -0.14 to -0.19), whereas FERT Anger showed moderate positive relations with PPI-R measures (*r*s ranged from 0.14 to 0.15).

As the composite FERT score calculated with the Bayesian 2PL procedure was negatively related to PPI-R Machiavellian Egocentricity only (*r* = -0.18, *p* = .0097), fixed-effects ANOVAs were conducted to assess whether the level of Machiavellian Egocentricity significantly differ between the terciles and quartiles for emotion recognition ability. Results indicated that the level of Machiavellian Egocentricity did not vary significantly between terciles (F(2, 205) = 1.57, p = 0.211; η_p^2 = 0.02, 95% CI [0.00, 1.00]) or quartiles (F(3, 204) = 1.67, p = 0.174; η_p^2 = 0.02, 95% CI [0.00, 1.00]), not supporting Hypothesis 1b. Therefore, the relation between Machiavellian Egocentricity and the Dunning-Kruger effect for emotion recognition ability was not examined.

Individual difference and interpersonal influences on cooperation in PD.

As Bartlett's test for homogeneity of group variances across the four between-group conditions for the iterated PD did not reveal significant heteroscedasticity for the rate of cooperation (i.e., proportion of rounds wherein the participant cooperated across 20 rounds; K-squared = 1.88, df = 3, p = 0.598) or the expectation of cooperation (K-squared = 2.4585, df = 3, p = 0.483), Fisher's one-way ANOVAs were used to examine participants' rate and expectation of cooperation across the PD conditions. Results from the ANOVA demonstrated differences across the four between-subjects conditions for the rate of cooperation in PD (F_{fisher}(3, 233) = 6.00, p < 0.001; see Figure 5A), as well as the expectation of cooperation in PD (F_{fisher}(3, 233) = 7.76, p < 0.001; see Figure 5B). The effect sizes for both ANOVAs ($\widehat{\omega}_p^2$ = 0.06 and 0.08) were medium, as per Murphy and Myors's (2004) criteria.

Supporting Hypothesis 2a, Holms-Sidak post-hoc pairwise multiple comparisons revealed that (1) participants in Condition 1 (i.e., participants who received Stroop-incongruent and cooperative affective feedback) cooperated more and expected more cooperation by the opponent in PD compared to participants in Condition 4 (i.e., participants who received Stroop-congruent and competitive affective feedback) (p < .05), and (2) participants in Condition 3 (i.e., participants who received Stroop-congruent and cooperative affective feedback) cooperated more and expected more cooperation by the opponent in PD compared to participants in

Condition 4 (i.e., participants who received Stroop-congruent and competitive affective feedback) (*p* < 0.001). As hypothesized, participants in Condition 2 (i.e., participants who received Stroop-incongruent and competitive affective feedback) cooperated more in PD compared to participants in Condition 4 (i.e., participants who received Stroop-congruent and competitive affective feedback) (*p* = 0.04). However, the pairwise comparison for the expectation of cooperation was not statistically significant.

In line with Hypothesis 2b, the rate of cooperation in PD was moderately negatively associated with two subscales of PPI-R Self-Centered Impulsivity: Machiavellian Egocentricity (*r* = -0.15, *p* = 0.033) and Blame Externalization (*r* = -0.17, *p* = 0.014). However, the expectation of cooperation in PD was moderately negatively associated with Machiavellian Egocentricity only (*r* = -0.14, *p* = 0.038) (see Table 6).

Subsequently, a 2 (Stroop-incongruent vs. Stroop-congruent) x 2 (Cooperative vs. Competitive) Factorial ANOVA was conducted to test for the interaction effect between the two conditional manipulations for affective feedback in PD on the rate of cooperation in PD (see Table 7, Figure 6A, and Figure 6B). The main effect of affective feedback congruence was found to be statistically significant and small, *F*(1, 233) = 9.90, *p* = 0.002; η_p^2 = 0.04, 95% CI [9.32e-03, 1.00]. However, the main effect of Stroop interference was found to be statistically not significant and very small, *F*(1, 233) = 1.35, *p* = 0.247, η_p^2 = 5.75e-03, 95% CI [0.00, 1.00]. These main effects were qualified by a statistically significant and small interaction between Stroop interference and affective feedback, *F(*1, 233) = 6.76, *p* = 0.010; η_p^2 = 0.03, 95% CI [3.77e-03, 1.00]. Tukey's Honest Significant Differences (HSD) post-hoc comparisons replicated the results from Fisher's one-way ANOVA. The comparisons indicated that (1) the rate of cooperation in PD by participants who received Stroop-incongruent and competitive affective feedback was 0.155 higher than participants who received Stroop-congruent and competitive affective feedback (in line with Hypothesis 2c; p = 0.043, 95% CI of the difference = 0.003 to 0.307), (2) rate of cooperation in PD by participants who received Stroop-congruent and cooperative affective feedback was 0.231 higher than participants who received Stroop-congruent and competitive affective feedback (p < 0.001, 95% CI of the difference = 0.084 to 0.379), and (3) the rate of cooperation in PD by participants who received Stroop-incongruent and cooperative affective feedback was 0.172 higher than participants who received Stroop-congruent and competitive affective feedback (p = 0.016, 95% CI of the difference = 0.024 to 0.322).

A 2 (Stroop-incongruent vs. Stroop-congruent) x 2 (Cooperative vs. Competitive) Factorial ANOVA was conducted to test for the interaction effect between the two conditional manipulations for affective feedback in PD on the expectation of cooperation in PD (see Table 8, Figure 6C, and Figure 6D). The main effect of affective feedback congruence was found to be statistically significant and medium, $F(1, 233) = 15.46, p < .001; \eta_p^2 = 0.06$, 95% CI [0.02, 1.00]. However, the main effect of Stroop interference was found to be statistically not significant and very small, $F(1, 233) = 0.10, p = 0.749, \eta_p^2 = 4.40\text{e-}04$, 95% CI [0.00, 1.00]. These main effects were qualified by a statistically significant and small interaction between Stroop interference and affective feedback, $F(1, 233) = 7.71, p = 0.006; \eta_p^2 = 0.03$, 95% CI [5.30e-03, 1.00]. Tukey's Honest Significant Differences (HSD) post-hoc comparisons replicated the results from Fisher's one-way ANOVA. The comparisons indicated that (1) the expectation of cooperation in PD by participants who received Stroop-congruent and cooperative affective feedback was 0.22 higher than participants who received Stroop-congruent and competitive affective feedback (p < 0.001, 95% *CI* of the difference = 0.1 to 0.34), and (2) the expectation of cooperation in PD by participants who received Stroop-incongruent and cooperative affective feedback was 0.14 higher than participants who received Stroop-congruent and competitive affective feedback (*p* = 0.018, 95% CI of the difference = 0.017 to 0.26). Statistically significant and negative 3-way interactions were found between affective feedback congruence, Stroop interference, and two PPI-R subscales (Stress Immunity and Coldheartedness) in predicting the rate and expectation of cooperation in PD (see Figures 7, Figure 8, Supplemental 6, Supplemental 14; Stress Immunity: *b* = -0.60, 95% CI [-1.13, -0.08], *t*(200) = -2.25, *p* = 0.025; Coldheartedness: *b* = -0.61, 95% CI [-1.13, -0.08], *t*(200) = -2.28, *p* = 0.023). This finding did not support Hypothesis 2d, which predicted that participants receiving cooperative and competitive feedback will show decreasing rates and expectations of cooperation at higher levels of psychopathy, with a larger interaction effect in conditions with Stroop interference.

Discussion

The primary aim of the current study was to examine the individual and interactive influence of psychopathy, affective feedback congruence, and Stroop interference on cooperation in PD. Furthermore, exploratory analyses were conducted to examine the relation between psychopathy and emotion perception, as well as the consistency of the relation between estimated and measured emotion recognition ability with the Dunning-Kruger effect. The findings were consistent with previous research on the influence of reverse appraisal on social cooperation, and provided novel insights into the interactive influence of interpersonal-affective processes associated with psychopathy in the context of social cooperation.

As recent studies have shown that adults with typical facial expression recognition ability have only modest cognitive insight into their ability (see Palermo et al., 2017, for a review), it was hypothesized that facial emotion recognition ability as measured by the proportion of correct responses on the FERT would be weakly correlated with estimated measures of performance. The relation between measured and estimated emotion recognition ability has been debated; some studies have found small or non-significant associations (Bowles et al., 2009; Hall et al., 1999), whereas other studies have found robust associations that demonstrate increased cognitive awareness of emotion recognition ability in comparison to other cognitive abilities (Arizpe et al., 2019; Zell & Krizan, 2014). Consistent with previous research (e.g., Hall et al., 1999), the present study found a statistically insignificant relation between actual and estimated emotion recognition ability.

The two-part analyses for Dunning-Kruger effect of facial emotion recognition ability yielded mixed findings. Although participants in upper quartiles (i.e., third and top quartiles) underestimated their performance on the task, no statistically significant differences were found at lower quartiles (i.e., bottom and second quartiles) despite marginal, non-significant overestimation and underestimation at the bottom and second quartiles, respectively. This is partially consistent with the Dunning-Kruger effect, which expects underestimation of performance at the top quartile and overestimation of sequentially decreasing magnitudes in the lower three quartiles (Kruger & Dunning, 1999). Similar patterns were found when the sample was divided into terciles for emotion recognition ability, with underestimation of performance at the upper two terciles and marginal, non-significant overestimation of performance at the lower tercile. In conducting a mixed-model regression analyses to explore the fit of linear, quadratic, and cubic trends for the relation between estimated and actual emotion recognition ability as measured by the composite FERT score, the cubic model yielded the best fit with both linear and cubic trends showing statistical significance. This is consistent with previous studies that have found a significant cubic trend for the relation between actual and estimated ability (Hood, 2015;

Sanchez & Dunning, 2018; Sanchez & Dunning, 2021). There are notable differences, however, as the cubic model for the present study included a significant linear trend and showed a tail-end decrease in the estimate of performance as opposed to an increase as shown in some studies (McKenzie et al., 2008; Sanchez & Dunning, 2018; but see Ericsson & Smith, 1991; Wallsten & Budescu, 1983).

Machiavellian Egocentricity was the only PPI-R measure to be significantly correlated with the actual measures of emotion recognition ability (the composite score as well as the proportion of correct responses for the FERT), suggesting that this subscale may partially explain variance in participants' estimates of their emotion recognition ability. The correlation was negative in direction, supporting the prediction that higher levels of psychopathy would be associated with lower estimates of emotion recognition ability. However, as the level of Machiavellian Egocentricity did not significantly differ between the terciles and quartiles for emotion recognition ability, and as prior analyses in this study only partially supported the hypothesis that a Dunning-Kruger effect would emerge for emotion recognition ability, the relation between Machiavellian Egocentricity and the Dunning-Kruger effect for emotion recognition ability was not examined.

As hypothesized, there was a significant main effect of affective feedback congruence on the rate and expectation of cooperation in PD such that participants who received Stroop-congruent x cooperative affective feedback showed higher rate and expectation of cooperation in PD compared to participants who received Stroop-congruent x competitive affective feedback. Contrary to predictions, the main effect of Stroop interference on the rate and expectation of cooperation in PD were both statistically not significant, qualified by a small but significant interaction between Stroop interference and affective feedback. Demonstrating the interaction effect, Stroop interference only had a significant influence on the rate of cooperation in PD for competitive affective feedback.

In analyzing relations between PPI-R measures and rate and expectation of cooperation in PD, the rate of cooperation was found to be moderately negatively correlated with both Machiavellian Egocentricity and Blame Externalization, whereas the expectation of cooperation was found to be moderately negatively correlated with Machiavellian Egocentricity only. This finding is partially consistent with previous studies that have found significant relations between PPI-R Self-Centered Impulsivity and its subscales (Berg et al., 2013; Curry et al., 2011; Mokros et al., 2008), as the higher-order dimension of Self-Centered Impulsivity (which includes dimension includes Machiavellian Egocentricity and Blame Externalization as its subscales) was not related to either rate or expectation of cooperation in this study.

However, the current study's finding of a moderate negative relation between Machiavellian Egocentricity and the rate of cooperation is consistent with two previous studies that have hypothesized that Machiavellian Egocentricity would be negatively related to cooperation in PD (Curry et al., 2011; Mokros et al., 2008), as Machiavellian Egocentricity reflects an orientation of interpersonal manipulativeness, and prioritizing one's interests over others' with "ruthless practicality" (Lilienfeld & Andrews, 1996). This hypothesis was supported in both studies, with Mokros et al. (2008) finding a moderate negative association between Machiavellian Egocentricity and cooperation in an iterated PD, and Curry et al. (2011) finding a negative association between Machiavellian Egocentricity and initiation and reciprocation of cooperation in a sequential one-shot PD. No study to date has examined the relation between PPI-R measures and the expectation of cooperation in PD. Therefore, this study presented a novel finding that Machiavellian Egocentricity is negatively related to the expectation of cooperation in an iterated PD.

The current study's finding of a negative relation between Blame Externalization and the rate of cooperation in PD is consistent with Mokros et al. (2008), which found a positive relation between Blame Externalization and the rate of defection in an iterated PD. This is also consistent with the operationalization of Blame Externalization in the PPI-R (i.e., an inclination to blame others as the cause of one's hardship and rationalizing one's own behavior; Lilienfeld & Andrews, 1996), and also with studies that have shown that attributions of blame mediate the relationship between anger and the inferring non-cooperative intentions, possibly contributing to increased rates of defection (Quigley & Tedeschi, 1996; Seip et al., 2014). Accordingly, studies have hypothesized that Blame Externalization should be negatively associated with the rate of cooperation, as cooperation depends on continuous positive interpersonal interactions in an iterated PD (e.g., Curry et al., 2011). However, it should be noted that some studies have found a positive relation between Blame Externalization and the rate of cooperation in PD (e.g., Baggio & Benning, 2011).

Finally, three-way interactions between affective feedback congruence, Stroop interference, and two PPI-R measures were detected in predicting the expectation of cooperation in PD: Stress Immunity (a subscale of Fearless Dominance) and Coldheartedness. Parallel trends emerged for the two three-way interactions between affective feedback congruence, Stroop interference, and the two PPI-R measures (Stress Immunity and Coldheartedness), with inverse trends emerging for the influence of affective feedback on cooperation in PD for higher levels of Stress Immunity and Coldheartedness depending on the presence or absence of Stroop interference (see Figure 9C and Figure 9K). In the absence of Stroop interference, participants who received cooperative affective feedback showed increasing expectations of cooperation in PD at higher levels of Stress Immunity and Coldheartedness, whereas participants who received competitive affective feedback showed decreasing expectations of cooperation in PD at higher levels of Stress Immunity and Coldheartedness. With Stroop interference, participants who received cooperative affective feedback showed decreasing expectations of cooperation in PD at higher levels of Stress Immunity and Coldheartedness, whereas participants who received competitive affective feedback showed increasing expectations of cooperation in PD at higher levels of Stress Immunity and Coldheartedness.

Limitations and Future Directions

Several limitations are to be noted in interpreting the results for the present study. First, the total sample size was below the sample size required to reliably estimate correlations and detect effects, as indicated by an a priori power analysis. As this limitation qualifies the generalizability of the results for the present study with a high likelihood of making a Type II error, subsequent replications would need to be conducted to increase confidence in the validity of the findings for the current study. Subsequent replications may also benefit from utilizing a longer version of the PPI-R, as the reliability of the data for the current study may have been affected by the low internal consistency (α=.59) for PPI-R Coldheartedness with the 40-item PPI-R (PPI-R-40). Although the internal consistency of the overall measure and the two other higher-order dimensional measures were in the acceptable range (α=.76–.80), analyses specifically conducted for relations with PPI-R Coldheartedness may have resulted in low validity, as the reliability of an instrument is closely associated with its validity (Tavakol & Dennik, 2011). As PPI-R Coldheartedness comprises fewer items compared to PPI-R Fearless Dominance and Self-Centered Impulsivity (Lilienfeld & Andrews, 1996), and as Cronbach's alpha is affected by the length of the measure (Nunnally & Bernstein, 1994), internal consistency may increase with longer versions of the PPI-R wherein more items are used to derive the subscale score.

Second, the ecological validity of the study may be improved by replicating the study with a human opponent with software such as z-Tree Unleashed (Zurich Toolbox for Ready-made Economic Experiments - Unleashed; Duch et al., 2020). Although the credibility of the deception (i.e., believing that the opponent is a real-life participant as stated rather than a computerized opponent) was not assessed in this study, it is possible for this deception to have had low credibility. This may have had a confounding influence on the participants' rate and expectation of cooperation in PD, as previous studies have shown that (the belief of) playing against a human opponent results in more positive affective responses and higher engagement (Kätsyri et al., 2013; Ravaja et al., 2004). As this study also relied on the assumption that participants would not be able to predict the pattern of the opponent's tit-for-tat strategy throughout the 20 rounds of the iterated PD or look past the deception implemented in this study to realize that the opponent is computerized due to the consistency of the strategy, it is possible that this may have contributed as a confounding variable. Future studies could also investigate the presence of a Dunning-Kruger effect for emotion recognition ability by controlling for the participants' age and gender, as a recent study by DeGutis and colleagues have shown that human awareness of emotion recognition ability varies significantly by age and gender (DeGutis et al., 2021).

Overall, the present study adds to the growing body of literature on the influence of personality and emotion on social cooperation more broadly, and the influence of interpersonal-affective processes of psychopathy on social cooperation more specifically. This study also further explored the relationship between psychopathy and emotion recognition ability, given the ongoing debate in the field regarding this relation. Exploratory analyses conducted on the Dunning-Kruger effect of emotion recognition ability and the influence of individual differences in personality on the variation of the relation between actual and estimated emotion recognition ability also adds to the sparse literature on the relationship between personality and cognitive awareness. Further advancing research on individual difference and interpersonal influences on cooperation will help in clarifying the interplay of structural, psychological, and dynamic interaction processes in influencing social cooperation, and should strive to model these interactive processes in more ecologically valid contexts.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Ammirati, R. J. (2013). *Self-Assessed Emotion Recognition Skill and Social Adjustment among College Students*. Emory University.

Anderson, N. E., Steele, V. R., Maurer, J. M., Rao, V., Koenigs, M. R., Decety, J., Kosson, D. S., Calhoun, V. D., & Kiehl, K. A. (2017). Differentiating emotional processing and attention in psychopathy with functional neuroimaging. *Cognitive, Affective & Behavioral Neuroscience*, *17*(3), 491–515. https://doi.org/10.3758/s13415-016-0493-5

Angelika-Nikita, M., de Melo, C. M., Terada, K., Lucas, G., & Gratch, J. (2021). The Impact of Partner Expressions on Felt Emotion in the Iterated Prisoner's Dilemma: An Event-level Analysis. *Proceedings of the Ninth Annual Conference on Advances in Cognitive Systems*, 18.

Arizpe, J. M., Saad, E., Douglas, A. O., Germine, L., Wilmer, J. B., & DeGutis, J. M. (2019). Self-reported face recognition is highly valid, but alone is not highly discriminative of prosopagnosia-level performance on objective assessments. *Behavior Research Methods*, *51*(3), 1102–1116. https://doi.org/10.3758/s13428-018-01195-w

Axelrod, R. (1980a). Effective Choice in the Prisoner's Dilemma. *The Journal of Conflict Resolution*, *24*(1), 3–25.

Axelrod, R. (1980b). More Effective Choice in the Prisoner's Dilemma. *The Journal of Conflict Resolution*, *24*(3), 379–403.

Axelrod, R., & Hamilton, W. D. (1981). The Evolution of Cooperation. *Science*, *211*(4489), 1390–1396.

Baggio, M. C., & Benning, S. D. (2022). The influence of psychopathic traits and strategic harshness on point gain and cooperation rate in the Prisoner's Dilemma. *Personality and Individual Differences*, *186*, 111344. https://doi.org/10.1016/j.paid.2021.111344

Baranger, D. (2021). *InteractionPoweR: Power analysis for interactions via simulation (Version R package version 0.1.0.3).* https://github.com/dbaranger/InteractionPoweR

Bartz, J., Simeon, D., Hamilton, H., Kim, S., Crystal, S., Braun, A., Vicens, V., & Hollander, E. (2011). Oxytocin can hinder trust and cooperation in borderline personality disorder. *Social Cognitive and Affective Neuroscience*, *6*(5), 556–563. https://doi.org/10.1093/scan/nsq085

Baskin-Sommers, A. R., Curtin, J. J., & Newman, J. P. (2011). Specifying the attentional selection that moderates the fearlessness of psychopathic offenders. *Psychological Science*, *22*(2), 226–234. https://doi.org/10.1177/0956797610396227

Baskin-Sommers, A. R., & Newman, J. P. (2014). Psychopathic and externalizing offenders display dissociable dysfunctions when responding to facial affect. *Personality Disorders: Theory, Research, and Treatment*, *5*(4), 369–379. https://doi.org/10.1037/per0000077

Bell, R., Sasse, J., Möller, M., Czernochowski, D., Mayr, S., & Buchner, A. (2016). Event-related potentials in response to cheating and cooperation in a social dilemma game. *Psychophysiology*, *53*(2), 216–228. https://doi.org/10.1111/psyp.12561

Berg, J. M., Lilienfeld, S. O., & Waldman, I. D. (2013). Bargaining with the devil: Using economic decision-making tasks to examine the heterogeneity of psychopathic traits. *Journal of Research in Personality*, *47*(5), 472–482. https://doi.org/10.1016/j.jrp.2013.04.003

Blair, J., Mitchell, D., & Blair, K. (2005). *The psychopath: Emotion and the brain.* (pp. ix, 201). Blackwell Publishing.

Blair, R. J. R. (2003). Facial expressions, their communicatory functions and neuro–cognitive substrates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 561–572. https://doi.org/10.1098/rstb.2002.1220

Blair, R. J. R. (2006). Subcortical Brain Systems in Psychopathy: The Amygdala and Associated Structures. In *Handbook of psychopathy* (pp. 296–312). The Guilford Press.

Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, *11*(9), 387–392. https://doi.org/10.1016/j.tics.2007.07.003

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., Rivolta, D., Wilson, C. E., & Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant-stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology*, *26*(5), 423–455. https://doi.org/10.1080/02643290903343149

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Breitmoser, Y. (2015). Cooperation, but No Reciprocity: Individual Strategies in the Repeated Prisoner's Dilemma. *American Economic Review*, *105*(9), 2882–2910. https://doi.org/10.1257/aer.20130675

Cleckley, H. (1941). *The mask of sanity; an attempt to reinterpret the so-called psychopathic personality* (p. 298). Mosby.

Cleckley, H. (1976). *The Mask of Sanity* (5th ed.). C.V. Mosby Co.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115*(1), 74–101. https://doi.org/10.1037/0033-2909.115.1.74

Curry, O., Chesters, M. J., & Viding, E. (2011). The psychopath's dilemma: The effects of psychopathic personality traits in one-shot games. *Personality and Individual Differences*, *50*(6), 804–809. https://doi.org/10.1016/j.paid.2010.12.036

Dawel, A., O'Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience and Biobehavioral Reviews*, *36*(10), 2288–2304. https://doi.org/10.1016/j.neubiorev.2012.08.006

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

de Melo, C. M., Carnevale, P. J., Read, S. J., & Gratch, J. (2014). Reading people's minds from emotion expressions in interdependent decision making. *Journal of Personality and Social Psychology*, *106*(1), 73–88. https://doi.org/10.1037/a0034251

DeGutis, J., Yosef, B., Lee, E., Saad, E., Arizpe, J., Song, J., Wilmer, J., Germine, L., & Esterman, M. (2021). The rise and fall of face recognition awareness across the lifespan. *PsyArXiv*. https://doi.org/10.31234/osf.io/rqkvx

Duch, M., Grossmann, M. R. P., & Lauer, T. (2020). *Z-Tree Unleashed: A Novel Client-Integrating Architecture for Conducting Z-Tree Experiments Over the Internet* (SSRN Scholarly Paper ID 3564274). Social Science Research Network. https://doi.org/10.2139/ssrn.3564274

Edens, J. F., Marcus, D. K., Lilienfeld, S. O., & Poythress, N. G. (2006). Psychopathic, not psychopath: Taxometric evidence for the dimensional structure of psychopathy. *Journal of Abnormal Psychology*, *115*(1), 131–144. https://doi.org/10.1037/0021-843X.115.1.131

Edmiston, E. K., Merkle, K., & Corbett, B. A. (2015). Neural and cortisol responses during play with human and computer partners in children with autism. *Social Cognitive and Affective Neuroscience*, *10*(8), 1074–1083. https://doi.org/10.1093/scan/nsu159

Eisenbarth, H., Lilienfeld, S. O., & Yarkoni, T. (2015). Using a genetic algorithm to abbreviate the Psychopathic Personality Inventory-Revised (PPI-R). *Psychological Assessment*, *27*(1), 194–202. https://doi.org/10.1037/pas0000032

Emonds, G., Declerck, C. H., Boone, C., Seurinck, R., & Achten, R. (2014). Establishing cooperation in a mixed-motive social dilemma. An fMRI study investigating the role of social value orientation and dispositional trust. *Social Neuroscience*, *9*(1), 10–22. https://doi.org/10.1080/17470919.2013.858080

Ericsson, K. A., & Smith, J. (Eds.). (1991). *Toward a general theory of expertise: Prospects and limits* (pp. x, 344). Cambridge University Press.

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, *13*(1), 1–25. https://doi.org/10.1007/s12110-002-1012-7

Foulkes, L. (2015). *Psychopathic traits and social reward*. University College London.

Gabay, A. S., Kempton, M. J., Gilleen, J., & Mehta, M. A. (2019). MDMA Increases Cooperation and Recruitment of Social Brain Areas When Playing Trustworthy Players in an Iterated Prisoner's Dilemma. *Journal of Neuroscience*, *39*(2), 307–320. https://doi.org/10.1523/JNEUROSCI.1276-18.2018

Gervais, M. M., Kline, M., Ludmer, M., George, R., & Manson, J. H. (2013). The strategy of psychopathy: Primary psychopathic traits predict defection on low-value relationships. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1757), 20122773. https://doi.org/10.1098/rspb.2012.2773

Glenn, A. L., Raine, A., & Schug, R. A. (2009). The neural correlates of moral decision-making in psychopathy. *Molecular Psychiatry*, *14*(1), 5–6. https://doi.org/10.1038/mp.2008.104

Gordon, H. L., Baird, A. A., & End, A. (2004). Functional differences among those high and low on a trait measure of psychopathy. *Biological Psychiatry*, *56*(7), 516–521. https://doi.org/10.1016/j.biopsych.2004.06.030

Gorenstein, E. E., & Newman, J. P. (1980). Disinhibitory psychopathology: A new perspective and a model for research. *Psychological Review*, *87*(3), 301–315. https://doi.org/10.1037/0033-295X.87.3.301

Gradin, V., Pérez, A., Macfarlane, J., Cavin, I., Waiter, G., Tone, E., Dritschel, B., Maiche, A., & Steele, J. (2016). Neural correlates of social exchanges during the Prisoner's Dilemma game in depression. *Psychological Medicine*, *1*, 1–12. https://doi.org/10.1017/S0033291715002834

Gratch, J., & Melo, C. (2019). *Inferring Intentions from Emotion Expressions in Social Decision Making* (pp. 141–160). https://doi.org/10.1007/978-3-030-32968-6_8

Guo, J., Gabry, J., Goodrich, B., Weber, S., Lee, D., Sakrejda, K., Martin, M., University, T. of C., Sklyar (R/cxxfunplus.R), O., Team (R/pairs.R, T. R. C., R/dynGet.R), Oehlschlaegel-Akiyoshi (R/pairs.R), J., Maddock (gamma.hpp), J., Bristow (gamma.hpp), P., Agrawal (gamma.hpp), N., Kormanyos (gamma.hpp), C., & Steve, B. (2021). *rstan: R Interface to Stan* (2.21.3) [Computer software]. https://CRAN.R-project.org/package=rstan

Hall, C. W., Gaul, L., & Kent, M. (1999). College students' perception of facial expressions. *Perceptual and Motor Skills*, *89*(3 Pt 1), 763–770. https://doi.org/10.2466/pms.1999.89.3.763

Hammerstein, P. (2003). Why is reciprocity so rare in social animals? A protestant appeal. In *Genetic and cultural evolution of cooperation* (pp. 83–93). MIT Press.

Hare, R. D. (1970). *Psychopathy: Theory and research* (pp. x, 138). John Wiley.

Hare, R. D. (2003). *The Hare Psychopathy Checklist Revised (2nd ed.)*. Multi-Health Systems.

Hareli, S., & Hess, U. (2010). What emotional reactions can tell us about the nature of others: An appraisal perspective on person perception. *Cognition and Emotion*, *24*(1), 128–140. https://doi.org/10.1080/02699930802613828

Haroush, K., & Williams, Z. M. (2015). Neuronal Prediction of Opponent's Behavior during Cooperative Social Interchange in Primates. *Cell*, *160*(6), 1233–1245. https://doi.org/10.1016/j.cell.2015.01.045

Harris, G., Rice, M., & Quinsey, V. (1994). Psychopathy as a Taxon: Evidence That Psychopaths Are a Discrete Class. *Journal of Consulting and Clinical Psychology*, *62*, 387–397. https://doi.org/10.1037/0022-006X.62.2.387

Hiatt, K. D., Schmitt, W. A., & Newman, J. P. (2004). Stroop tasks reveal abnormal selective attention among psychopathic offenders. *Neuropsychology*, *18*(1), 50–59. https://doi.org/10.1037/0894-4105.18.1.50

Hood, J. (2015). *Deconstructing The Unskilled-And-Unaware Problem: Examining The Effect of Feedback on Misestimation While Disentangling Cognitive Bias From Statistical Artifact*. https://www.semanticscholar.org/paper/Deconstructing-The-Unskilled-And-Unaware-Problem%3A-Hood/7bca1eff26f0a5ad5c28b66ff3d9c92d0a1c9ded

Israel, L., Paukner, P., Schiestel, L., Diepold, K., & Schönbrodt, F. (2021). *Open Library for Affective Videos (OpenLAV)*. https://www.psycharchives.org/en/item/18779e98-c04b-4299-8311-dc442dc89bcd

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology*, *66*(1), 206–219. https://doi.org/10.1037/0022-3514.66.1.206

Johnston, L., Hawes, D. J., & Straiton, M. (2014). Psychopathic Traits and Social Cooperation in the Context of Emotional Feedback. *Psychiatry, Psychology and Law*, *21*(5), 767–778. https://doi.org/10.1080/13218719.2014.893550

Kaartinen, M., Puura, K., Pispa, P., Helminen, M., Salmelin, R., Pelkonen, E., Juujärvi, P., Kessler, E. B., & Skuse, D. H. (2019). Associations between cooperation, reactive aggression and social impairments among boys with autism spectrum disorder. *Autism*, *23*(1), 154–166. https://doi.org/10.1177/1362361317726417

Kätsyri, J., Hari, R., Ravaja, N., & Nummenmaa, L. (2013). The opponent matters: Elevated FMRI reward responses to winning against a human versus a computer opponent during interactive video game playing. *Cerebral Cortex (New York, N.Y.: 1991)*, *23*(12), 2829–2839. https://doi.org/10.1093/cercor/bhs259

Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal relations: A theory of interdependence.* New York: Wiley.

King-Casas, B., & Chiu, P. H. (2012). Understanding Interpersonal Function in Psychiatric Illness Through Multiplayer Economic Games. *Biological Psychiatry*, *72*(2), 119–125. https://doi.org/10.1016/j.biopsych.2012.03.033

Koenigs, M. (2012). The role of prefrontal cortex in psychopathy. *Reviews in the Neurosciences*, *23*(3), 253–262. https://doi.org/10.1515/revneuro-2012-0036

Kranefeld, I., & Blickle, G. (2022). Disentangling the relation between psychopathy and emotion recognition ability: A key to reduced workplace aggression? *Personality and Individual Differences*, *184*, 111232. https://doi.org/10.1016/j.paid.2021.111232

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Kujala, A., & Danielsbacka, M. (2019). *Reciprocity in human societies: From ancient times to the modern welfare state*. Palgrave Macmillan.

Kulms, P., Kopp, S., & Krämer, N. C. (2014). Let's Be Serious and Have a Laugh: Can Humor Support Cooperation with a Virtual Agent? In T. Bickmore, S. Marsella, & C. Sidner (Eds.), *Intelligent Virtual Agents* (pp. 250–259). Springer International Publishing. https://doi.org/10.1007/978-3-319-09767-1_32

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and Decision Making. *Annual Review of Psychology*, *66*(1), 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Levenson, M., Kiehl, K., & Fitzpatrick, C. (1995). Assessing Psychopathic Attributes in a Noninstitutionalized Population. *Journal of Personality and Social Psychology*, *68*, 151–158. https://doi.org/10.1037//0022-3514.68.1.151

Lilienfeld, S. O. (1990). *Development and preliminary validation of a self-report measure of psychopathic personality.* University of Minnesota.

Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *Journal of Personality Assessment*, *66*(3), 488–524. https://doi.org/10.1207/s15327752jpa6603_3

Lilienfeld, S. O., Latzman, R. D., Watts, A. L., Smith, S. F., & Dutton, K. (2014). Correlates of psychopathic personality traits in everyday life: Results from a large community survey. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/article/10.3389/fpsyg.2014.00740

Lilienfeld, S. O., Watts, A. L., Murphy, B., Costello, T. H., Bowes, S. M., Smith, S. F., Latzman, R. D., Haslam, N., & Tabb, K. (2019). Psychopathy as an Emergent Interpersonal Syndrome: Further Reflections and Future Directions. *Journal of Personality Disorders*, *33*(5), 645–652. https://doi.org/10.1521/pedi.2019.33.5.645

Lilienfeld, S. O., & Widows, M. R. (2005). Psychopathic Personality Inventory-Revised: Professional Manual. *Lutz, FL: Psychological Assessment Resources Inc.*, 6.

Long, K., Felton, J. W., Lilienfeld, S. O., & Lejuez, C. W. (2014). The Role of Emotion Regulation in the Relations between Psychopathy Factors and Impulsive and Premeditated Aggression. *Personality Disorders*, *5*(4), 390–396. https://doi.org/10.1037/per0000085

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey* (pp. xix, 509). Wiley.

Lykken, D. T. (1957). A study of anxiety in the sociopathic personality. *The Journal of Abnormal and Social Psychology*, *55*(1), 6–10. https://doi.org/10.1037/h0047232

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, *32*(3), 454–465. PubMed. https://doi.org/10.1016/j.neubiorev.2007.08.003

McClure, E. B., Parrish, J. M., Nelson, E. E., Easter, J., Thorne, J. F., Rilling, J. K., Ernst, M., & Pine, D. S. (2007). Responses to conflict and cooperation in adolescents with anxiety and mood disorders. *Journal of Abnormal Child Psychology*, *35*(4), 567–577. https://doi.org/10.1007/s10802-007-9113-8

McKenzie, C. R. M., Liersch, M. J., & Yaniv, I. (2008). Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Processes*, *107*(2), 179–191. https://doi.org/10.1016/j.obhdp.2008.02.007

Mokros, A., Menner, B., Eisenbarth, H., Alpers, G. W., Lange, K. W., & Osterheider, M. (2008). Diminished cooperativeness of psychopaths in a prisoner's dilemma game yields higher rewards. *Journal of Abnormal Psychology*, *117*(2), 406–413. https://doi.org/10.1037/0021-843X.117.2.406

Montañes Rada, F., de Lucas Taracena, M. T., & Martín Rodríguez, M. A. (2003). Antisocial personality disorder evaluation with the prisoner's dilemma. *Actas Espanolas De Psiquiatria*, *31*(6), 307–314.

Murphy, K. R., & Myors, B. (2004). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests, 2nd ed* (pp. ix, 160). Lawrence Erlbaum Associates Publishers.

Nentjes, L., Garofalo, C., & Kosson, D. S. (2022). 4 - Emotional functioning in psychopathy: A critical review and integration with general emotion theories. In P. B. Marques, M. Paulino, & L. Alho (Eds.), *Psychopathy and Criminal Behavior* (pp. 75–103). Academic Press. https://doi.org/10.1016/B978-0-12-811419-3.00006-6

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. McGraw-Hill Higher, INC.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Palermo, R., Rossion, B., Rhodes, G., Laguesse, R., Tez, T., Hall, B., Albonico, A., Malaspina, M., Daini, R., Irons, J., Al-Janabi, S., Taylor, L. C., Rivolta, D., & McKone, E. (2017). Do People Have Insight into their Face Recognition Abilities? *Quarterly Journal of Experimental Psychology*, *70*(2), 218–233. https://doi.org/10.1080/17470218.2016.1161058

Passarelli, M., Masini, M., Bracco, F., Petrosino, M., & Chiorri, C. (2018). Development and validation of the Facial Expression Recognition Test (FERT). *Psychological Assessment*, *30*(11), 1479–1490. https://doi.org/10.1037/pas0000595

Patrick, C. J. (Ed.). (2018). *Handbook of psychopathy, 2nd ed* (pp. xx, 828). The Guilford Press.

Patterson, C. M., & Newman, J. P. (1993). Reflectivity and learning from aversive events: Toward a psychological mechanism for the syndromes of disinhibition. *Psychological Review*, *100*(4), 716–736. https://doi.org/10.1037/0033-295X.100.4.716

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Pletzer, J. L., Balliet, D., Joireman, J., Kuhlman, D. M., Voelpel, S. C., & Van Lange, P. A. M. (2018). Social Value Orientation, Expectations, and Cooperation in Social Dilemmas: A

Meta–Analysis. *European Journal of Personality*, *32*(1), 62–83. https://doi.org/10.1002/per.2139

Quigley, B. M., & Tedeschi, J. T. (1996). Mediating Effects of Blame Attributions on Feelings of Anger. *Personality and Social Psychology Bulletin*, *22*(12), 1280–1288. https://doi.org/10.1177/01461672962212008

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's Dilemma: A Study in Conflict and Cooperation*. University of Michigan Press.

Ravaja, N., Saari, T., Turpeinen, M., Laarni, J., Salminen, M., & Kivikangas, M. (2006). Spatial Presence and Emotions during Video Game Playing: Does It Matter with Whom You Play? *Presence: Teleoperators and Virtual Environments*, *15*(4), 381–392. https://doi.org/10.1162/pres.15.4.381

Raykov, T., Dimitrov, D. M., & Asparouhov, T. (2010). Evaluation of Scale Reliability With Binary Measures Using Latent Variable Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(2), 265–279. https://doi.org/10.1080/10705511003659417

Reed, L. I., Zeglen, K. N., & Schmidt, K. L. (2012). Facial expressions as honest signals of cooperative intent in a one-shot anonymous Prisoner's Dilemma game. *Evolution and Human Behavior*, *33*(3), 200–209. https://doi.org/10.1016/j.evolhumbehav.2011.09.003

Richell, R. A., Mitchell, D. G. V., Newman, C., Leonard, A., Baron-Cohen, S., & Blair, R. J. R. (2003). Theory of mind and psychopathy: Can psychopathic individuals read the 'language of the eyes'? *Neuropsychologia*, *41*(5), 523–526. https://doi.org/10.1016/S0028-3932(02)00175-6

Rilling, J., Goldsmith, D., Glenn, A., Jairam, M., Elfenbein, H., Dagenais, J., Murdock, C., & Pagnoni, G. (2008). The Neural Correlates of the Affective Response to Unreciprocated

Cooperation. *Neuropsychologia*, *46*, 1256–1266. https://doi.org/10.1016/j.neuropsychologia.2007.11.033

Rilling, J. K., Glenn, A. L., Jairam, M. R., Pagnoni, G., Goldsmith, D. R., Elfenbein, H. A., & Lilienfeld, S. O. (2007). Neural correlates of social cooperation and non-cooperation as a function of psychopathy. *Biological Psychiatry*, *61*(11), 1260–1271. https://doi.org/10.1016/j.biopsych.2006.07.021

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A Neural Basis for Social Cooperation. *Neuron*, *35*(2), 395–405. https://doi.org/10.1016/S0896-6273(02)00755-9

Roberts, G. (1998). Competitive altruism: From reciprocity to the handicap principle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1394), 427–431. https://doi.org/10.1098/rspb.1998.0312

Rodebaugh, T. L., Shumaker, E. A., Levinson, C. A., Fernandez, K. C., Langer, J. K., Lim, M. H., & Yarkoni, T. (2013). Interpersonal Constraint Conferred by Generalized Social Anxiety Disorder is Evident on a Behavioral Economics Task. *Journal of Abnormal Psychology*, *122*(1), 39–44. https://doi.org/10.1037/a0030975

Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology*, *114*(1), 10–28. https://doi.org/10.1037/pspa0000102

Sanchez, C., & Dunning, D. (2021). Jumping to conclusions: Implications for reasoning errors, false belief, knowledge corruption, and impeded learning. *Journal of Personality and Social Psychology*, *120*(3), 789–815. https://doi.org/10.1037/pspp0000375

Scherer, K. R., & Grandjean, D. (2008). Facial expressions allow inference of both emotions and their components. *Cognition and Emotion*, *22*(5), 789–801. https://doi.org/10.1080/02699930701516791

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Seip, E. C., Van Dijk, W. W., & Rotteveel, M. (2014). Anger motivates costly punishment of unfair behavior. *Motivation and Emotion*, *38*(4), 578–588. https://doi.org/10.1007/s11031-014-9395-4

Shane, M. S., & Groat, L. L. (2018). Capacity for upregulation of emotional processing in psychopathy: All you have to do is ask. *Social Cognitive and Affective Neuroscience*, *13*(11), 1163–1176. https://doi.org/10.1093/scan/nsy088

Singer, T., & Fehr, E. (2005). The Neuroeconomics of Mind Reading and Empathy. *American Economic Review*, *95*(2), 340–345. https://doi.org/10.1257/000282805774670103

Stephens, D. W., McLinn, C. M., & Stevens, J. R. (2002). Discounting and Reciprocity in an Iterated Prisoner's Dilemma. *Science*, *298*(5601), 2216–2218. https://doi.org/10.1126/science.1078498

Strohmaier, H. (2015). *Successful Psychopathy: Do Abnormal Selective Attention Processes Observed in Criminal Psychopaths Replicate Among Non-Criminal Psychopaths?* [Doctor of Philosophy, Drexel University]. https://doi.org/10.17918/etd-6390

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Testori, M., Harris, T. O. A., Hoyle, R. B., & Eisenbarth, H. (2019). The effect of psychopathy on cooperative strategies in an iterated Prisoner's Dilemma experiment with emotional feedback. *Scientific Reports*, *9*(1), 2299. https://doi.org/10.1038/s41598-019-38796-0

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, *146*(1), 30–90. https://doi.org/10.1037/bul0000217

Thompson, K., Nahmias, E., Fani, N., Kvaran, T., Turner, J., & Tone, E. (2021). The Prisoner's Dilemma paradigm provides a neurobiological framework for the social decision cascade. *PLOS ONE*, *16*(3), e0248006. https://doi.org/10.1371/journal.pone.0248006

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, *46*(1), 35–57.

van Dijk, E., & De Dreu, C. K. W. (2021). Experimental Games and Social Decision Making. *Annual Review of Psychology*, *72*(1), 415–438. https://doi.org/10.1146/annurev-psych-081420-110718

Van Kleef, G. A. (2009). How Emotions Regulate Social Life: The Emotions as Social Information (EASI) Model. *Current Directions in Psychological Science*, *18*(3), 184–188. https://doi.org/10.1111/j.1467-8721.2009.01633.x

Van Lange, P. A. M., Joireman, J., Parks, C. D., & Van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes*, *120*(2), 125–141. https://doi.org/10.1016/j.obhdp.2012.11.003

Viding, E. (2019). We need to talk about development and victims. *Journal of Personality Disorders*, *33*(5), 640–644. https://doi.org/10.1521/pedi.2019.33.5.640

Wallsten, T. S., & Budescu, D. V. (1983). State of the Art—Encoding Subjective Probabilities: A Psychological and Psychometric Review. *Management Science*, *29*(2), 151–173. https://doi.org/10.1287/mnsc.29.2.151

Widom, C. S. (1976). Interpersonal conflict and cooperation in psychopaths. *Journal of Abnormal Psychology*, *85*(3), 330–334. https://doi.org/10.1037/0021-843X.85.3.330

Wilson, S., Stroud, C. B., & Durbin, C. E. (2017). Interpersonal dysfunction in personality disorders: A meta-analytic review. *Psychological Bulletin*, *143*(7), 677–734. https://doi.org/10.1037/bul0000101

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2016). Validation of the Amsterdam Dynamic Facial Expression Set – Bath Intensity Variations (ADFES-BIV): A Set of Videos Expressing Low, Intermediate, and High Intensity Emotions. *PLoS ONE*, *11*(1), e0147112. https://doi.org/10.1371/journal.pone.0147112

Wood, R. M., Rilling, J. K., Sanfey, A. G., Bhagwagar, Z., & Rogers, R. D. (2006). Effects of tryptophan depletion on the performance of an iterated Prisoner's Dilemma game in healthy adults. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *31*(5), 1075–1084. https://doi.org/10.1038/sj.npp.1300932

Wright, A. G. C., & Hopwood, C. J. (2022). Integrating and distinguishing personality and psychopathology. *Journal of Personality*, *90*(1), 5–19. https://doi.org/10.1111/jopy.12671

Xu, Y., Kelly, A., & Smillie, C. (2013). Emotional expressions as communicative signals. In S. Hancil & D. Hirst (Eds.), *Iconicity in Language and Literature* (Vol. 13, pp. 33–60). John Benjamins Publishing Company. https://doi.org/10.1075/ill.13.02xu

Yang, Y., & Raine, A. (2008). Functional neuroanatomy of psychopathy. *Psychiatry*, *7*(3), 133–136. https://doi.org/10.1016/j.mppsy.2008.01.001

Zell, E., & Krizan, Z. (2014). Do People Have Insight Into Their Abilities? A Metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125. https://doi.org/10.1177/1745691613518075

Tables

Table 1. *Means, standard deviations, and correlations with confidence intervals for psychopathy measures.*

Variable	*M*	*SD*	1	2	3	4
1. PPI-R Total	85.25	11.61	(.80)			
2. PPI-R Fearless Dominance	35.00	6.40	.69*** [.13, .76]	(.76)		
3. PPI-R Self-Centered Impulsivity	40.63	7.67	.80*** [.75, .84]	.17* [.03, .30]	(.79)	
4. PPI-R Coldheartedness	9.42	2.45	.43*** [.31, .53]	.13 [-.01, .26]	.13** [-.01, .26]	(.59)

Note. * is *p*<.05, ** is *p*<.01, and *** is *p*<.001. PPI-R = Psychopathic Personality Inventory – Revised. *M* and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence intervals for each correlation. The confidence interval is a plausible range of populations correlations that could have caused the sample correlation (Cumming, 2014). Internal consistencies (Cronbach’s alpha) are reported along the diagonal.

Table 2. *Means, standard deviations, and correlations with confidence intervals for facial emotion recognition measures.*

Variable	*M*	*SD*	1	2	3	4	5	6	7
1. FERT Score	-0.92	0.74							
2. FERT Estimate	0.65	0.17	.10 [-.02, .23]						
3. FERT Anger	0.73	0.22	.69*** [.62, .75]	.13* [.00, .25]					
4. FERT Disgust	0.65	0.20	.43*** [.32, .53]	.02 [-.10, .15]	.11 [-.02, .24]				
5. FERT Fear	0.48	0.23	.59*** [.50, .66]	-.00 [-.13, .13]	.16* [.03, .28]	.17** [.04, .29]			
6. FERT Happiness	0.79	0.10	.39*** [.28, .50]	.03 [-.10, .15]	.20** [.08, .32]	.19** [.06, .31]	.04 [-.09, .17]		
7. FERT Sadness	0.63	0.16	.35*** [.23, .46]	.02 [-.11, .15]	.11 [-.02, .23]	-.04 [-.17, .09]	.07 [-.06, .19]	-.05 [-.18, .08]	
8. FERT Surprise	0.88	0.15	.38*** [.26, .48]	.11 [-.02, .23]	.17** [.04, .29]	.24*** [.12, .36]	-.01 [-.14, .12]	.28*** [.16, .39]	.14* [.02, .27]

Note. * is *p*<.05, ** is *p*<.01, and *** is *p*<.001. FERT = Facial Expression Recognition Task. *M* and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence intervals for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014).

Table 3. *Correlations between psychopathic traits and facial emotion recognition measures.*

Variables	FERT Composite Score	FERT Correct Responses	FERT Estimate	FERT Anger	FERT Disgust	FERT Fear	FERT Happiness	FERT Sadness	FERT Surprise
PPI-R Total	-0.02	-0.05	0.13	0.14*	-0.03	-0.12	-0.14*	-0.05	-0.01
PPI-R Fearless Dominance	0.02	-0.03	0.17*	0.13	-0.05	-0.06	-0.05	-0.08	-0.02
Stress Immunity	-0.02	-0.05	0.18**	0.07	-0.11	-0.02	-0.03	-0.11	0.02
Social Influence	0.02	-0.02	0.07	0.03	-0.02	-0.03	0.04	0.06	-0.14*
Fearlessness	0.03	-0.01	0.11	0.15*	0.01	-0.07	-0.11	-0.10	0.05
PPI-R Self-Centered Impulsivity	-0.08	-0.07	0.03	0.06	-0.01	-0.13	-0.18**	-0.01	-0.01
Carefree Nonplanfulness	0.10	0.12	-0.04	0.10	0.11	0.06	-0.02	0.04	0.04
Machiavellian Egocentricity	-0.18**	-0.18**	0.02	-0.05	-0.09	-0.18**	-0.18*	0.03	-0.10
Blame Externalization	-0.09	-0.08	-0.06	0.00	-0.03	-0.08	-0.19**	0.03	-0.05
Rebellious Nonconformity	-0.02	-0.03	0.14*	0.13	0.01	-0.15*	-0.09	-0.11	0.08
PPI-R Coldheartedness	0.10	0.08	0.11	0.14*	0.01	0.00	0.04	0.00	0.07

Note. * is *p*<.05, ** is *p*<.01, and *** is *p*<.001. PPI-R = Psychopathic Personality Inventory – Revised. FERT = Facial Expression Recognition Task. Values in square brackets indicate the 95% confidence intervals for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014).

Table 4. *Paired t-tests of actual vs. estimated proportion of correct responses on the Facial Expression Recognition Task.*

Variable	*n*	*M (SD)*	*t*	Estimate
Lower Tercile Performance Group				
Actual	95	0.60 (0.06)	1.73	Over (*ns*)
Estimated	--	0.63 (0.19)	--	
Middle Tercile Performance Group				
Actual	66	0.71 (0.01)	-3.13 **	Under
Estimated	--	0.64 (0.17)	--	
Upper Tercile Performance Group				
Actual	76	0.79 (0.04)	-6.42 ***	Under
Estimated	--	0.68 (0.15)	--	
Bottom Quartile Performance Group				
Actual	73	0.58 (0.06)	1.61	Over (*ns*)
Estimated	--	0.63 (0.21)		
Second Quartile Performance Group				
Actual	61	0.68 (0.01)	-1.79	Under (*ns*)
Estimated	--	0.65 (0.17)		
Third Quartile Performance Group				
Actual	51	0.74 (0.01)	-3.32 **	Under
Estimated	--	0.67 (0.14)		
Top Quartile Performance Group				
Actual	52	0.81 (0.01)	-6.01 ***	Under
Estimated	--	0.15 (0.15)		

Note. * is *p*<.05, ** is *p*<.01, and *** is *p*<.001. *ns* = Not significant. Over = Overestimate. Under = Underestimate.

Table 5. *Regression results using estimated facial expression recognition ability as the criterion.*

Predictor	b	b 95% CI [LL, UL]	β	β 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	r	Fit	Difference
(Intercept)	0.00	[-0.13, 0.13]							
Linear	0.10	[-0.02, 0.23]	0.10	[-0.02, 0.23]	.01	[.00, .05]	.10		
								R^2 = .011	
								95% CI[.00,.05]	
(Intercept)	-0.01	[-0.17, 0.15]							
Linear	0.11	[-0.02, 0.23]	0.11	[-0.02, 0.23]	.01	[-.02, .04]	.10		
Quadratic	0.01	[-0.08, 0.10]	0.01	[-0.11, 0.14]	.00	[-.00, .00]	.00		
								R^2 = .011	ΔR^2 = .000
								95% CI[.00,.05]	95% CI[-.00, .00]
(Intercept)	0.02	[-0.13, 0.18]							
Linear	0.35**	[0.13, 0.56]	0.35	[0.13, 0.56]	.04	[-.01, .09]	.10		
Quadratic	-0.03	[-0.13, 0.06]	-0.05	[-0.18, 0.09]	.00	[-.01, .01]	.00		
Cubic	-0.08**	[-0.14, -0.02]	-0.31	[-0.53, -0.09]	.03	[-.01, .07]	-.02		
								R^2 = .043*	ΔR^2 = .031**
								95% CI[.00,.09]	95% CI[-.01, .07]

Note. * is p<.05, ** is p<.01, and *** is p<.001. Linear = Linear trend for the Facial Expression Recognition Task score. Quadratic = Quadratic trend for the Facial Expression Recognition Task score. Cubic = Cubic trend for the Facial Expression Recognition Test score. A significant b-weight indicates the beta-weight and semi-partial correlation are also significant. b represents unstandardized regression weights. β indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. r represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

Table 6. *Correlations between psychopathic traits and indices for the modified Prisoner's Dilemma.*

	Cooperation	Expectation of Cooperation
PPI-R Total	-0.01	-0.01
PPI-R Fearless Dominance	0.08	0.08
Stress Immunity	0.08	0.04
Social Influence	0.01	-0.01
Fearlessness	0.07	0.12
PPI-R Self-Centered Impulsivity	-0.09	-0.08
Carefree Nonplanfulness	0.09	0.03
Machiavellian Egocentricity	-0.15*	-0.14*
Blame Externalization	-0.17*	-0.12
Rebellious Nonconformity	0.00	0.02
PPI-R Coldheartedness	0.05	-0.01

Note. * is *p*<.05, ** is *p*<.01, and *** is *p*<.001. PPI-R = Psychopathic Personality Inventory – Revised.

Table 7. *Fixed-Effects ANOVA results using rate of cooperation in PD as the criterion.*

Predictor	Sum of Squares	*df*	Mean Square	*F*	*p*	partial η^2	partial η^2 90% CI [LL, UL]
(Intercept)	0.00	1	0.00	0.00	.989		
Stroop	1.30	1	1.30	1.38	.241	.01	[.00, .03]
Affective Feedback	8.71	1	8.71	9.26	.003 **	.04	[.01, .09]
Stroop x Affective Feedback	6.36	1	6.36	6.76	.010*	.03	[.00, .07]
Error	219.07	233	0.94				

Note. * is *p*<.05, ** is *p*<.01, and *** is *p*<.001. LL and UL represent the lower-limit and upper-limit of the partial η^2 confidence interval, respectively.

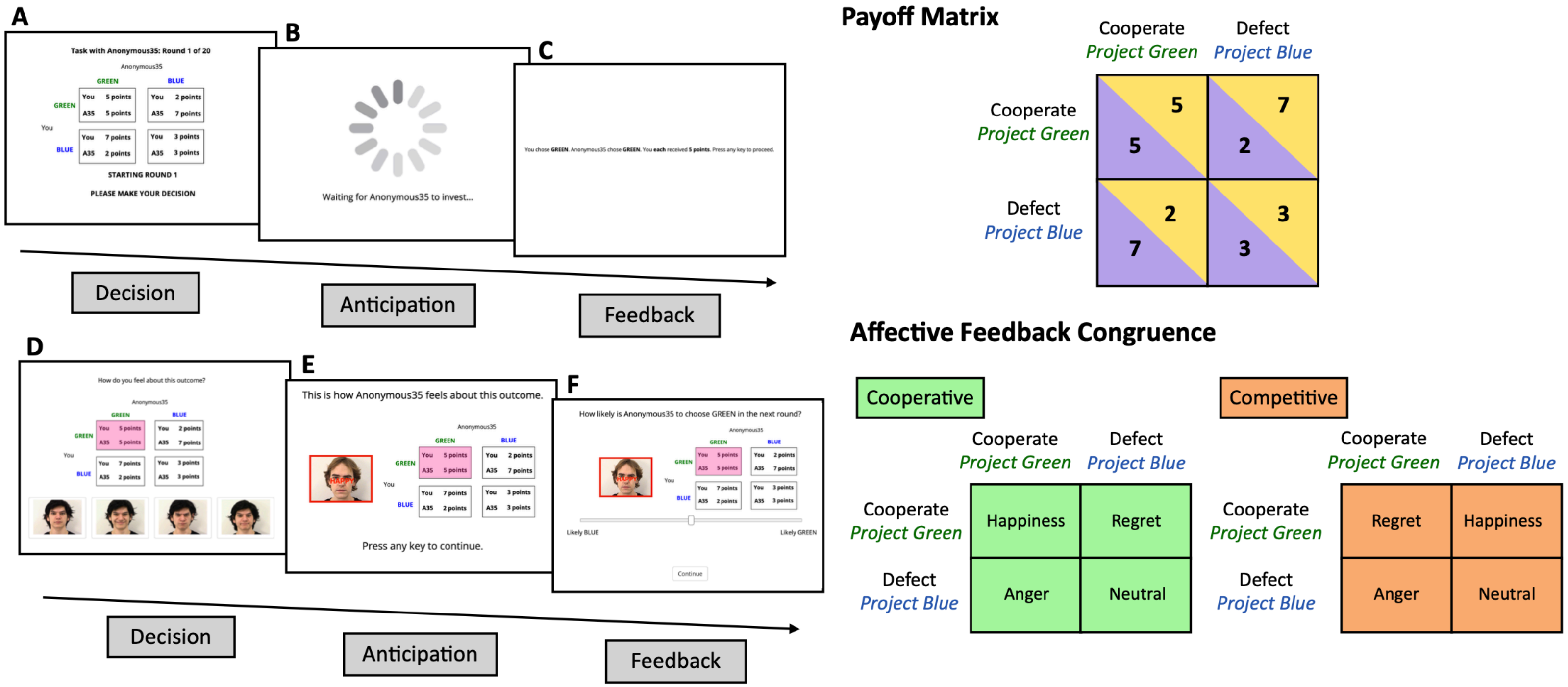
Table 8. *Fixed-Effects ANOVA results using expectation of cooperation in PD as the criterion.*

Predictor	Sum of Squares	*df*	Mean Square	*F*	*p*	partial η^2	partial η^2 90% CI [LL, UL]
(Intercept)	0.00	1	0.00	0.00	.974		
Stroop	0.10	1	0.10	0.11	.743	.00	[.00, .01]
Affective Feedback	13.45	1	13.45	14.61	.000 ***	.06	[.02, .11]
Stroop x Affective Feedback	7.10	1	7.10	7.71	.006 **	.03	[.01, .08]
Error	214.57	233	0.92				

Note. * is *p*<.05, ** is *p*<.01, and *** is *p*<.001. LL and UL represent the lower-limit and upper-limit of the partial η^2 confidence interval, respectively.

Figures

Figure 1. *Experimental procedure for one round of the PD (left, A-F), payoff matrix (upper right), and affective feedback congruence for each outcome (lower right).*


A
Task with Anonymous35: Round 1 of 20
Anonymous35
GREEN
BLUE
You
You 5 points
A35 5 points
You 2 points
A35 7 points
You 7 points
A35 2 points
You 3 points
A35 3 points
STARTING ROUND 1
PLEASE MAKE YOUR DECISION
B
Waiting for Anonymous35 to invest...
C
You chose GREEN. Anonymous35 chose GREEN. You each received 5 points. Press any key to proceed.
Decision
Anticipation
Feedback
D
How do you feel about this outcome?
E
This is how Anonymous35 feels about this outcome.
Press any key to continue.
F
How likely is Anonymous35 to choose GREEN in the next round?
Likely BLUE
Likely GREEN
Continue
Decision
Anticipation
Feedback
Payoff Matrix
Affective Feedback Congruence
Cooperative
Competitive

Payoff Matrix

	Cooperate *Project Green*	Defect *Project Blue*
Cooperate *Project Green*	5 5	7 2
Defect *Project Blue*	2 7	3 3

Affective Feedback Congruence

Cooperative

	Cooperate *Project Green*	Defect *Project Blue*
Cooperate *Project Green*	Happiness	Regret
Defect *Project Blue*	Anger	Neutral

Competitive

	Cooperate *Project Green*	Defect *Project Blue*
Cooperate *Project Green*	Regret	Happiness
Defect *Project Blue*	Anger	Neutral

Figure 2. *Average proportion of correct responses for the Facial Expression Recognition Task by tercile and quartile.*

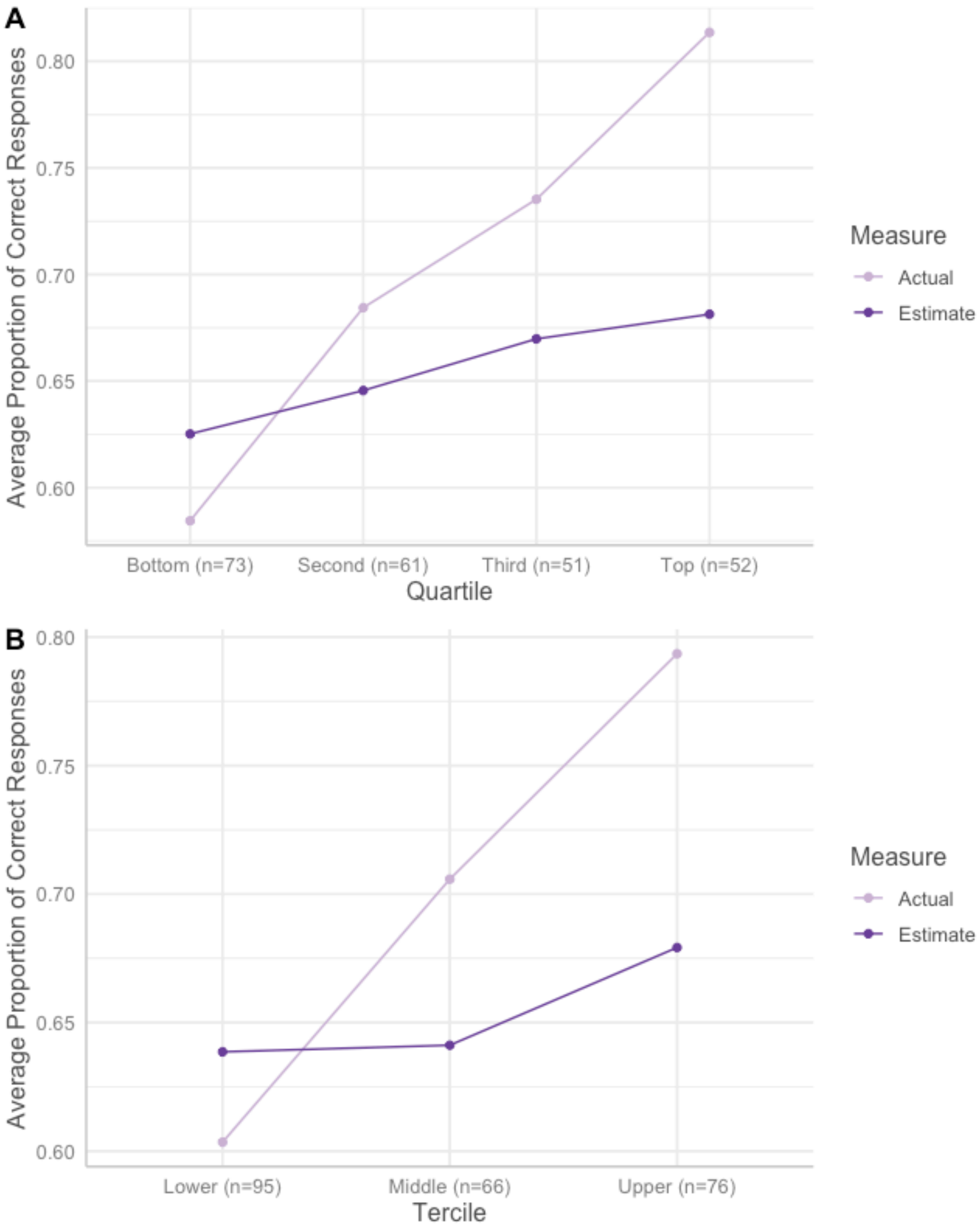


Figure 3. *Linear, quadratic, and cubic trends for the relation between estimated and actual emotion recognition ability as measured by the proportion of correct FERT responses (left panel) and the composite FERT score (right panel).*

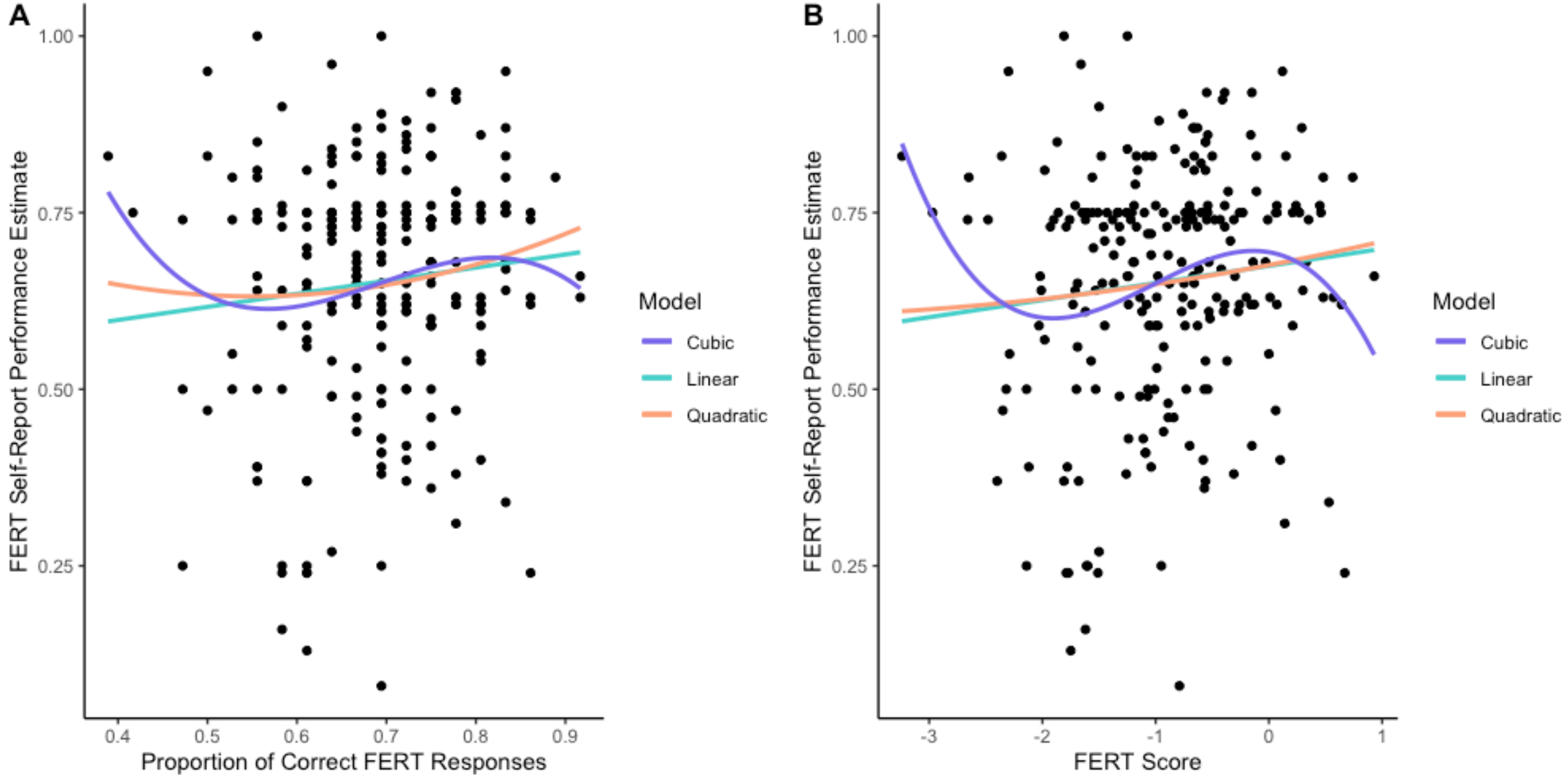


Figure 4. *Estimated vs. actual proportion of correct responses on the Facial Expression Recognition Task.*

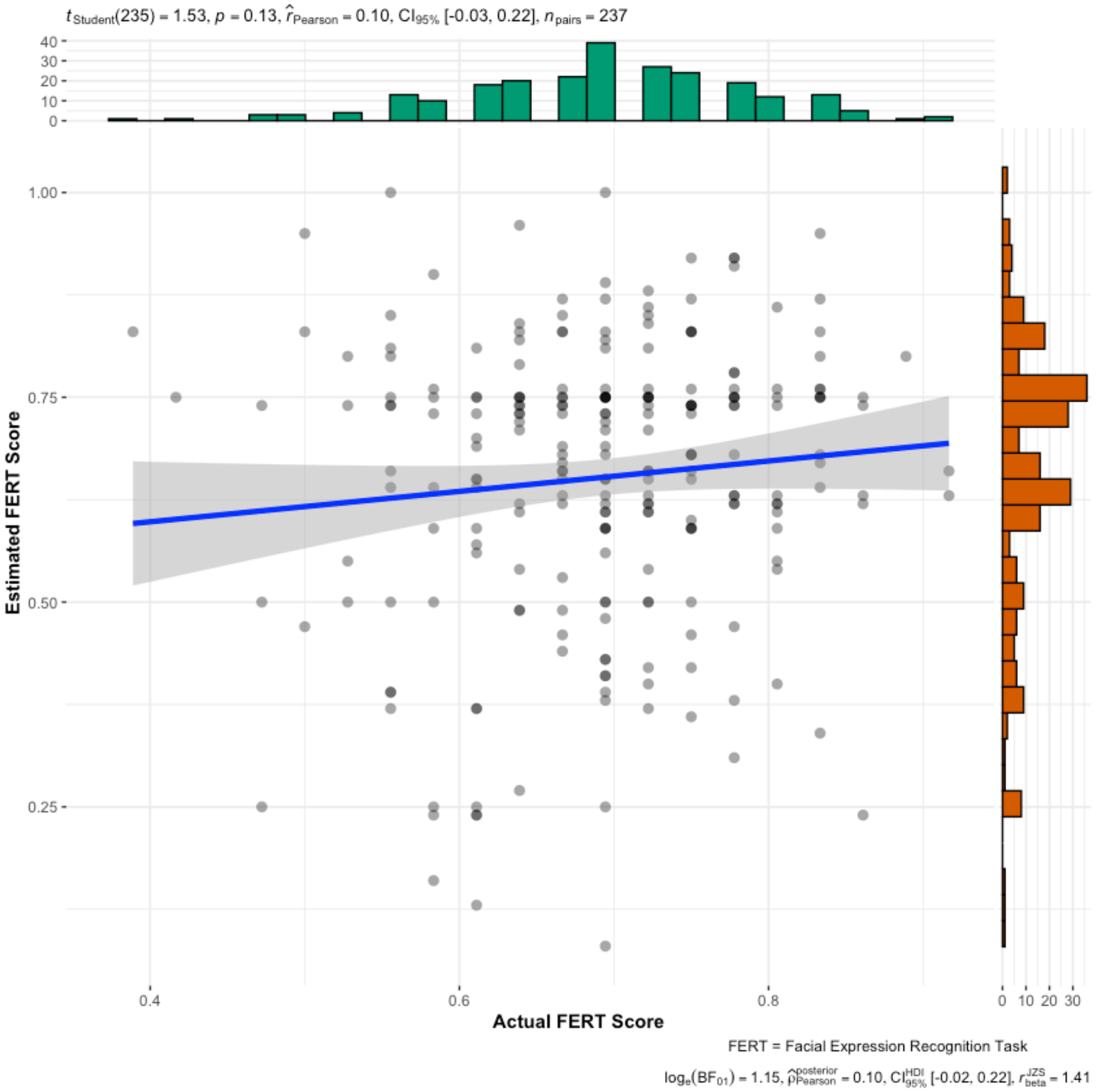


Figure 5. *Box-violin plots of the rate and expectation of cooperation in PD across conditions.*

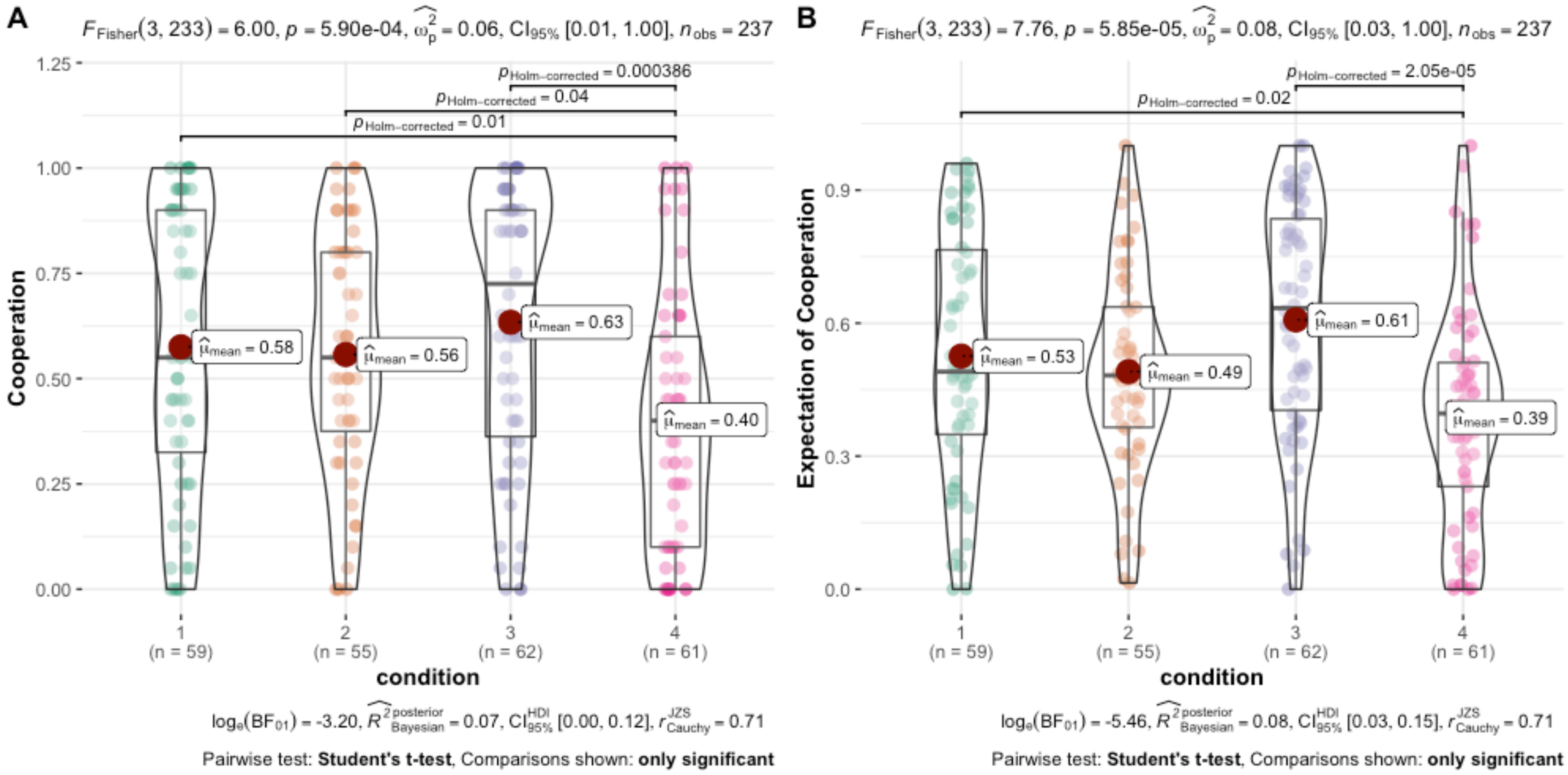


Figure 6. *Interaction plots of the rate and expectation of cooperation in PD across conditions.*

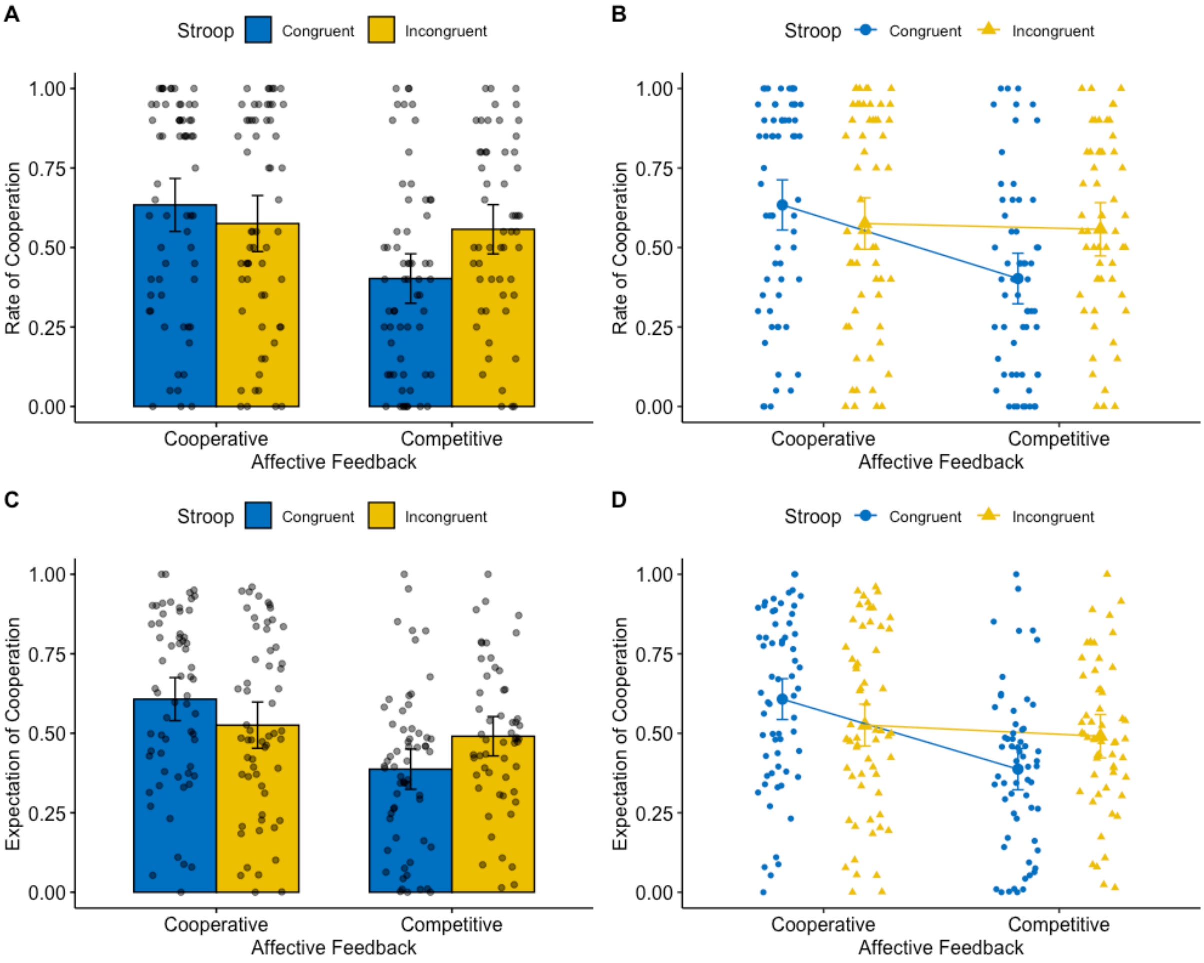


Figure 7. *Three-way interactions between affective feedback congruence, Stroop interference, and PPI-R measures on the rate of cooperation in PD.*

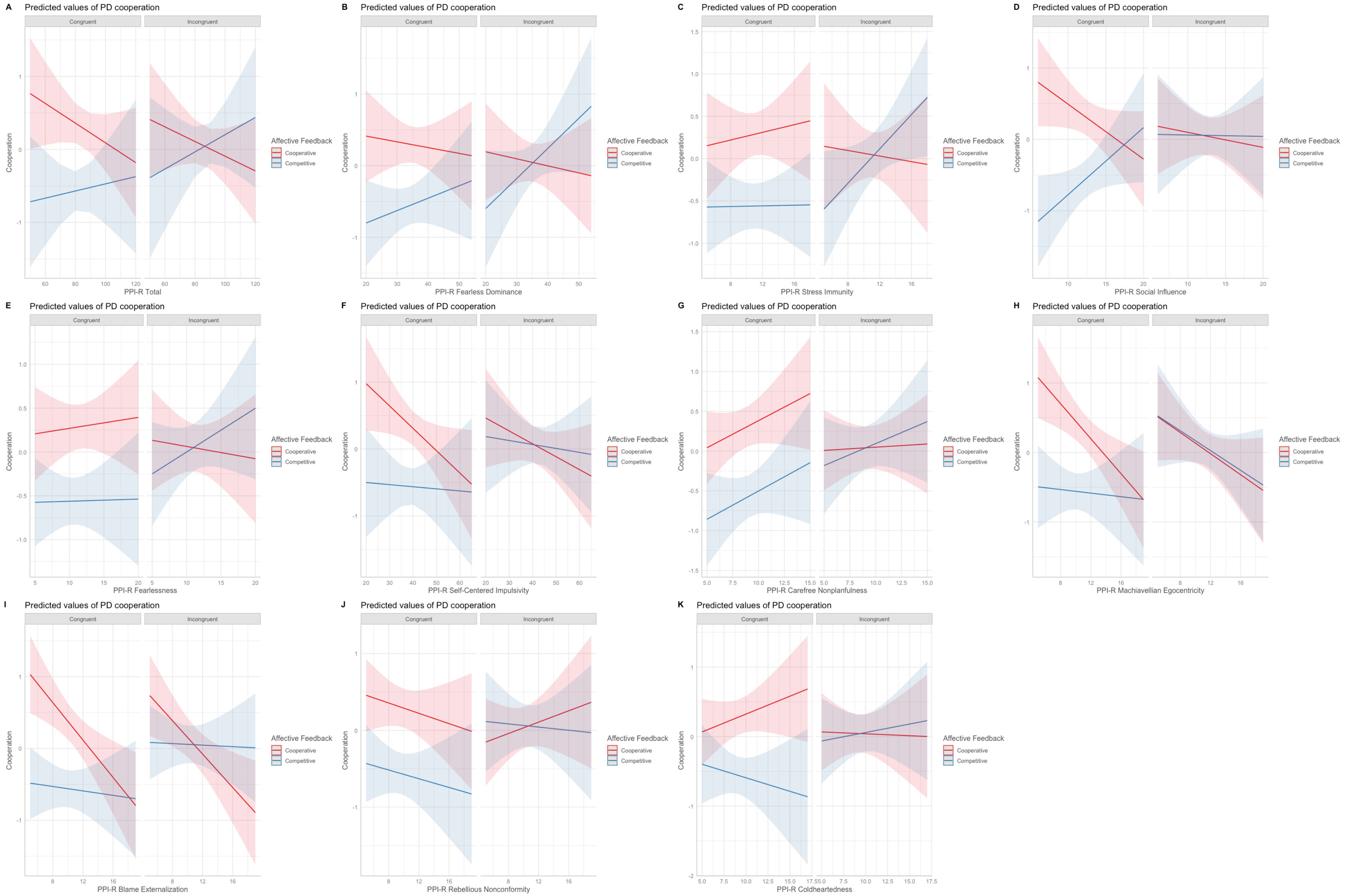
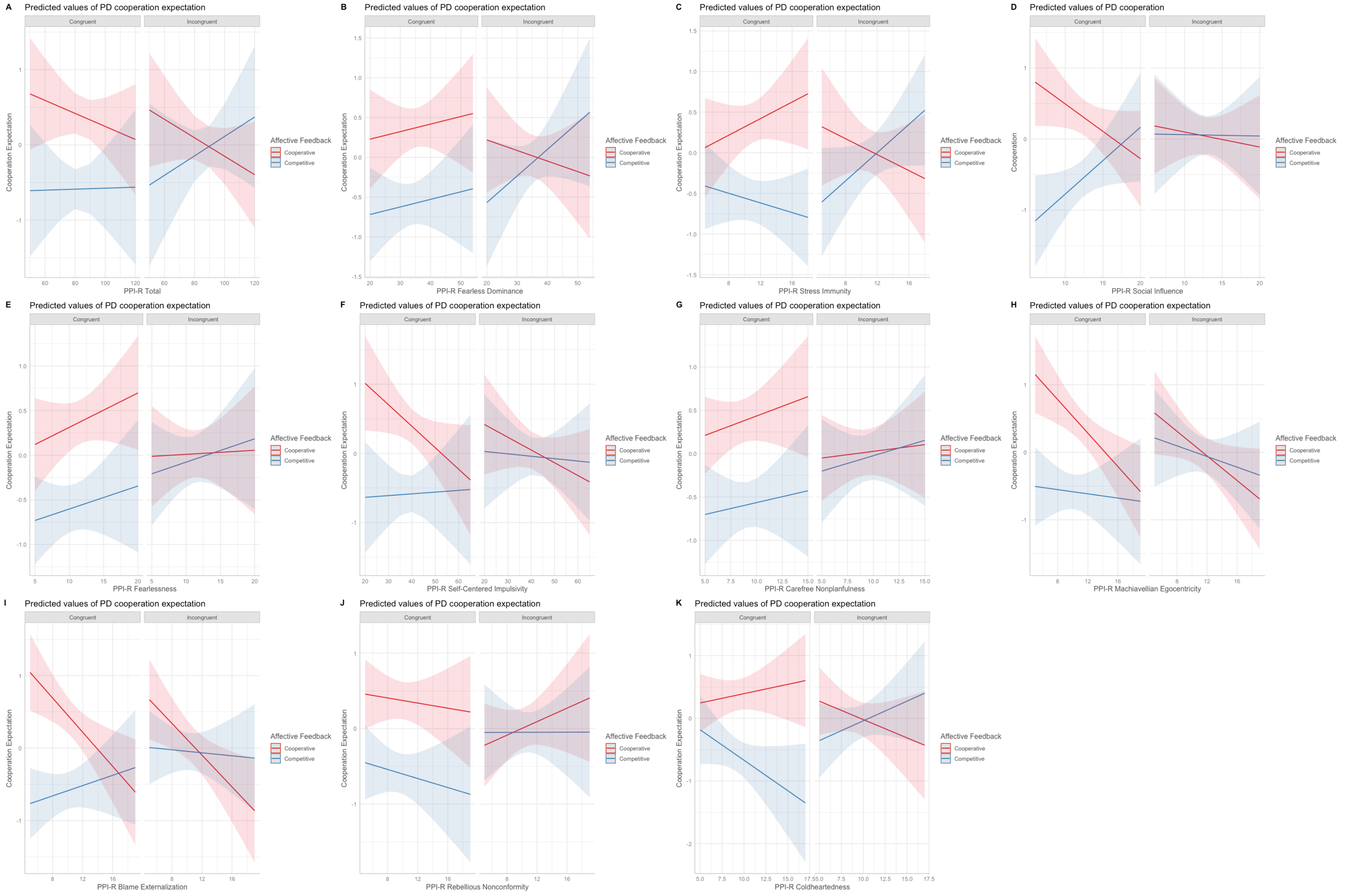


Figure 8. *Three-way interactions between affective feedback congruence, Stroop interference, and PPI-R measures on expectation of cooperation in PD.*



Appendices

S1. *Fixed-Effects ANOVA results using PPI-R Machiavellian Egocentricity as the criterion.*

Predictor	Sum of Squares	*df*	Mean Square	*F*	*p*	partial η^2	partial η^2 90% CI [LL, UL]
(Intercept)	10534.05	1	10534.05	1395.55	.000 ***		
Tercile	23.65	2	11.82	1.57	.211	.02	[.00, .05]
Error	1547.41	205	7.55				

Note. * is p<.05, ** is p<.01, and *** is p<.001. LL and UL represent the lower-limit and upper-limit of the partial η^2 confidence interval, respectively. PPI-R = Psychopathic Personality Inventory – Revised. Tercile = Tercile for average proportion of correct responses for the Facial Expression Recognition Task.

S2. *Fixed-Effects ANOVA results using PPI-R Machiavellian Egocentricity as the criterion.*

Predictor	Sum of Squares	*df*	Mean Square	*F*	*p*	partial η^2	partial η^2 90% CI [LL, UL]
(Intercept)	8120.07	1	8120.07	1080.34	.000 ***		
Quartile	37.76	3	12.59	1.67	.174	.02	[.00, .06]
Error	1533.30	204	7.52				

Note. * is *p*<.05, ** is *p*<.01, and *** is *p*<.001. LL and UL represent the lower-limit and upper-limit of the partial η^2 confidence interval, respectively. Quartile = Quartile for average proportion of correct responses for the Facial Expression Recognition Task.

S3. *Means and standard deviations for cooperation in PD as a function of a 2 (Cooperative) X 2 (Stroop) design.*

	Stroop					
	Congruent		Incongruent		Marginal	
Affective Feedback	*M*	*SD*	*M*	*SD*	*M*	*SD*
Competitive	0.40	0.30	0.56	0.29	0.48	0.30
Cooperative	0.63	0.33	0.58	0.34	0.61	0.33
Marginal	0.52	0.34	0.57	0.31		

Note. *M* and *SD* represent mean and standard deviation, respectively.

S4. *Three-way interaction between PPI-R total score, affective feedback congruence, and Stroop interference on the rate and expectation of cooperation in PD.*

	Cooperation					**Expectation of Cooperation**				
Predictors	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*
(Intercept)	-0.04	-0.01	[-0.18, 0.09]	[-0.14, 0.13]	0.529	-0.07	-0.02	[-0.20, 0.07]	[-0.15, 0.12]	0.332
PPI-R Total	-0.02	-0.02	[-0.16, 0.12]	[-0.16, 0.12]	0.781	-0.02	-0.02	[-0.16, 0.12]	[-0.16, 0.12]	0.761
Cooperative	-0.21	-0.21	[-0.35, -0.08]	[-0.35, -0.08]	**0.002**	-0.27	-0.27	[-0.40, -0.13]	[-0.40, -0.14]	**<0.001**
Stroop	-0.08	-0.08	[-0.22, 0.55]	[-0.22, 0.05]	0.217	-0.04	-0.04	[-0.17, 0.09]	[-0.17, 0.09]	0.540
PPI-R Total * Cooperative	0.12	0.12	[-0.02, 0.26]	[-0.02, 0.26]	0.104	0.10	0.10	[-0.04, 0.24]	[-0.04, 0.24]	0.154
PPI-R Total * Stroop	-0.03	-0.03	[-0.17, 0.11]	[-0.17, 0.11]	0.676	-0.03	-0.03	[-0.16, 0.11]	[-0.16, 0.11]	0.721
Cooperative * Stroop	-0.20	-0.20	[-0.34, 0.07]	[-0.33, -0.07]	**0.004**	-0.21	-0.21	[-0.34, -0.08]	[-0.35, -0.08]	**0.002**
PPI-R Total * Cooperative * Stroop	-0.01	-0.01	[-0.15, 0.13]	[-0.15, 0.13]	0.889	-0.05	-0.05	[-0.18, 0.09]	[-0.18, 0.09]	0.511
Observations	208					208				
R^2 / R^2 adjusted	0.115 / 0.084					0.136 / 0.106				

Note. Bolded is *p*<.05. *b* represents unstandardized regression weights. *β* indicates the standardized regression weights. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. PPI-R Total = Psychopathic Personality Inventory – Revised Total. Cooperative = Cooperative affective feedback. Stroop = Stroop-incongruent affective feedback.

S5. *Three-way interaction between PPI-R Fearless Dominance, affective feedback congruence, and Stroop interference on the rate and expectation of cooperation in PD.*

	Cooperation					**Expectation of Cooperation**				
Predictors	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*
(Intercept)	-0.04	-0.01	[-0.18, 0.09]	[-0.14, 0.12]	0.522	-0.06	-0.01	[-0.19, 0.07]	[-0.14, 0.12]	0.335
PPI-R FD	0.07	0.07	[-0.07, 0.20]	[-0.07, 0.20]	0.335	0.06	0.06	[-0.07, 0.20]	[-0.07, 0.20]	0.353
Cooperative	-0.21	-0.21	[-0.34, -0.08]	[-0.34, -0.08]	**0.002**	-0.26	-0.26	[-0.39, -0.13]	[-0.39, -0.13]	**<0.001**
Stroop	-0.08	-0.08	[-0.21, 0.05]	[-0.21, 0.05]	0.266	-0.04	-0.04	[-0.17, 0.09]	[-0.17, 0.09]	0.550
PPI-R FD * Cooperative	0.12	0.12	[-0.01, 0.26]	[-0.01, 0.26]	0.074	0.07	0.08	[-0.06, 0.21]	[-0.06, 0.21]	0.268
PPI-R FD * Stroop	-0.04	-0.04	[-0.17, 0.10]	[-0.17, 0.10]	0.594	-0.00	-0.00	[-0.13, 0.13]	[-0.14, 0.13]	0.978
Cooperative * Stroop	-0.20	-0.20	[-0.34, -0.07]	[-0.34, -0.07]	**0.003**	-0.21	-0.22	[-0.34, -0.09]	[-0.35, -0.09]	**0.001**
PPI-R FD * Cooperative * Stroop	-0.04	-0.04	[-0.18, 0.09]	[-0.18, 0.09]	0.538	-0.07	-0.07	[-0.21, 0.06]	[-0.21, 0.06]	0.269
Observations	208					208				
R² / R² adjusted	0.117 / 0.086					0.134 / 0.104				

Note. Bolded is *p*<.05. *b* represents unstandardized regression weights. *β* indicates the standardized regression weights. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. PPI-R FD = Psychopathic Personality Inventory – Revised Fearless Dominance. Cooperative = Cooperative affective feedback. Stroop = Stroop-incongruent affective feedback.

S6. *Three-way interaction between PPI-R Stress Immunity, affective feedback congruence, and Stroop interference on the rate and expectation of cooperation in PD.*

	Cooperation					**Expectation of Cooperation**				
Predictors	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*
(Intercept)	-0.05	-0.01	[-0.18, 0.08]	[-0.15, 0.12]	0.465	-0.07	-0.02	[-0.20, 0.06]	[-0.15, 0.11]	0.305
PPI-R STI	0.08	0.08	[-0.06, 0.21]	[-0.06, 0.21]	0.268	0.04	0.04	[-0.09, 0.17]	[-0.09, 0.17]	0.539
Cooperative	-0.22	-0.22	[-0.35, -0.08]	[-0.35, -0.08]	**0.001**	-0.26	-0.27	[-0.39, -0.14]	[-0.39, -0.14]	**<0.001**
Stroop	-0.09	-0.09	[-0.22, 0.05]	[-0.22, 0.05]	0.202	-0.04	-0.04	[-0.17, 0.09]	[-0.17, 0.09]	0.535
PPI-R STI * Cooperative	0.07	0.07	[-0.07, 0.20]	[-0.07, 0.20]	0.325	0.04	0.04	[-0.09, 0.17]	[-0.09, 0.17]	0.566
PPI-R STI * Stroop	-0.04	-0.04	[-0.18, 0.09]	[-0.18, 0.09]	0.541	-0.01	-0.01	[-0.14, 0.12]	[-0.14, 0.12]	0.864
Cooperative * Stroop	-0.21	-0.21	[-0.34, -0.08]	[-0.34, -0.08]	**0.002**	-0.22	-0.22	[-0.35, -0.09]	[-0.35, -0.09]	**0.001**
PPI-R STI * Cooperative * Stroop	-0.10	-0.10	[-0.23, 0.04]	[-0.23, 0.04]	0.162	-0.15	-0.15	[-0.28, -0.02]	[-0.28, -0.02]	**0.025**
Observations	208					208				
R^2 / R^2 adjusted	0.119 / 0.088					0.147 / 0.117				

Note. Bolded is *p*<.05. *b* represents unstandardized regression weights. *β* indicates the standardized regression weights. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. PPI-R STI = Psychopathic Personality Inventory – Revised Stress Immunity. Cooperative = Cooperative affective feedback. Stroop = Stroop-incongruent affective feedback.

S7. *Three-way interaction between PPI-R Social Influence, affective feedback congruence, and Stroop interference on the rate and expectation of cooperation in PD.*

	Cooperation					Expectation of Cooperation				
Predictors	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*
(Intercept)	-0.03	0.00	[-0.17, 0.10]	[-0.17, 0.10]	0.608	-0.06	-0.01	[-0.19, 0.07]	[-0.14, 0.12]	0.379
PPI-R SOI	-0.00	-0.00	[-0.14, 0.13]	[-0.14, 0.13]	0.951	-0.02	-0.02	[-0.15, 0.12]	[-0.15, 0.12]	0.819
Cooperative	-0.20	-0.20	[-0.33, -0.07]	[-0.33, -0.07]	**0.003**	-0.25	-0.25	[-0.38, -0.12]	[-0.38, -0.12]	**<0.001**
Stroop	-0.08	-0.08	[-0.22, 0.05]	[-0.22, 0.05]	0.210	-0.04	-0.04	[-0.17, 0.09]	[-0.17, 0.09]	0.557
PPI-R SOI * Cooperative	0.13	0.13	[-0.00, 0.27]	[-0.00, 0.27]	0.055	0.11	0.11	[-0.02, 0.25]	[-0.02, 0.25]	0.092
PPI-R SOI * Stroop	0.03	0.03	[-0.11, 0.16]	[-0.11, 0.16]	0.684	0.02	0.02	[-0.11, 0.16]	[-0.11, 0.16]	0.724
Cooperative * Stroop	-0.21	-0.21	[-0.34, -0.08]	[-0.34, -0.08]	**0.002**	-0.22	-0.22	[-0.35, -0.09]	[-0.35, -0.09]	**0.001**
PPI-R SOI * Cooperative * Stroop	0.10	0.10	[-0.03, 0.24]	[-0.03, 0.24]	0.125	0.03	0.03	[-0.10, 0.17]	[-0.10, 0.17]	0.606
Observations	208					208				
R² / R² adjusted	0.131 / 0.101					0.139 / 0.109				

Note. Bolded is *p*<.05. *b* represents unstandardized regression weights. *β* indicates the standardized regression weights. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. PPI-R SOI = Psychopathic Personality Inventory – Revised Social Influence. Cooperative = Cooperative affective feedback. Stroop = Stroop-incongruent affective feedback.

S8. *Three-way interaction between PPI-R Fearlessness, affective feedback congruence, and Stroop interference on the rate and expectation of cooperation in PD.*

	Cooperation					**Expectation of Cooperation**				
Predictors	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*
(Intercept)	-0.04	-0.01	[-0.18, 0.09]	[-0.14, 0.12]	0.531	-0.06	-0.01	[-0.19, 0.07]	[-0.14, 0.12]	0.353
PPI-R F	0.04	0.04	[-0.09, -0.18]	[-0.09, 0.18]	0.526	0.08	0.08	[-0.05, 0.21]	[-0.05, 0.21]	0.226
Cooperative	-0.21	-0.21	[-0.34, -0.07]	[-0.34, -0.07]	**0.002**	-0.25	-0.25	[-0.38, -0.12]	[-0.38, -0.12]	**<0.001**
Stroop	-0.09	-0.09	[-0.23, 0.04]	[-0.23, 0.04]	0.161	-0.05	-0.05	[-0.18, 0.08]	[-0.18, 0.08]	0.490
PPI-R F * Cooperative	0.05	0.05	[-0.09, 0.18]	[-0.09, 0.18]	0.502	0.01	0.01	[-0.12, 0.14]	[-0.12, 0.14]	0.910
PPI-R F * Stroop	-0.02	-0.02	[-0.15, 0.12]	[-0.15, 0.12]	0.793	0.03	0.03	[-0.10, 0.16]	[-0.10, 0.16]	0.668
Cooperative * Stroop	-0.21	-0.21	[-0.35, -0.08]	[-0.35, -0.08]	**0.002**	-0.22	-0.22	[-0.35, -0.09]	[-0.35, -0.09]	**0.001**
PPI-R F * Cooperative * Stroop	-0.06	-0.06	[-0.20, 0.07]	[-0.20, 0.07]	0.357	-0.03	-0.03	[-0.16, 0.10]	[-0.16, 0.10]	0.660
Observations	208					208				
R^2 / R^2 adjusted	0.106 / 0.075					0.131 / 0.101				

Note. Bolded is *p*<.05. *b* represents unstandardized regression weights. *β* indicates the standardized regression weights. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. PPI-R F = Psychopathic Personality Inventory – Revised Fearlessness. Cooperative = Cooperative affective feedback. Stroop = Stroop-incongruent affective feedback.

S9. *Three-way interaction between PPI-R Self-Centered Impulsivity, affective feedback congruence, and Stroop interference on the rate and expectation of cooperation in PD.*

	Cooperation					**Expectation of Cooperation**				
Predictors	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*
(Intercept)	-0.04	-0.01	[-0.17, 0.09]	[-0.14, 0.13]	0.553	-0.06	-0.01	[-0.19, 0.07]	[-0.14, 0.12]	0.393
PPI-R SCI	-0.12	-0.12	[-0.25, 0.02]	[-0.25, 0.02]	0.088	-0.10	-0.10	[-0.23, 0.04]	[-0.23, 0.04]	0.154
Cooperative	-0.21	-0.21	[-0.34, -0.08]	[-0.34, -0.08]	**0.002**	-0.26	-0.26	[-0.39, -0.13]	[-0.39, -0.13]	**<0.001**
Stroop	-0.10	-0.10	[-0.24, 0.03]	[-0.23, 0.03]	0.131	-0.05	-0.05	[-0.18, 0.08]	[-0.18, 0.08]	0.437
PPI-R SCI * Cooperative	0.08	0.08	[-0.05, 0.22]	[-0.05, 0.22]	0.227	0.09	0.09	[-0.04, 0.23]	[-0.04, 0.23]	0.170
PPI-R SCI * Stroop	-0.02	-0.02	[-0.16, 0.11]	[-0.16, 0.11]	0.752	-0.01	-0.01	[-0.15, 0.12]	[-0.15, 0.12]	0.851
Cooperative * Stroop	-0.21	-0.21	[-0.35, -0.08]	[-0.34, -0.08]	**0.002**	-0.22	-0.22	[-0.35, -0.09]	[-0.35, -0.09]	**0.001**
PPI-R SCI * Cooperative * Stroop	0.03	0.03	[-0.10, 0.17]	[-0.10, 0.17]	0.639	0.04	0.04	[-0.10, 0.17]	[-0.10, 0.17]	0.599
Observations	208					208				
R^2 / R^2 adjusted	0.125 / 0.094					0.145 / 0.115				

Note. Bolded is *p*<.05. *b* represents unstandardized regression weights. *β* indicates the standardized regression weights. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. PPI-R SCI = Psychopathic Personality Inventory – Revised Self-Centered Impulsivity. Cooperative = Cooperative affective feedback. Stroop = Stroop-incongruent affective feedback.

S10. *Three-way interaction between PPI-R Carefree Nonplanfulness, affective feedback congruence, and Stroop interference on the rate and expectation of cooperation in PD.*

	Cooperation					Expectation of Cooperation				
Predictors	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*
(Intercept)	-0.04	-0.01	[-0.17, 0.09]	[-0.14, 0.13]	0.550	-0.06	-0.01	[-0.19, 0.07]	[-0.14, 0.12]	0.367
PPI-R CN	0.12	0.12	[-0.02, 0.26]	[-0.02, 0.25]	0.083	0.07	0.07	[-0.06, 0.21]	[-0.06, 0.21]	0.285
Cooperative	-0.22	-0.22	[-0.35, -0.09]	[-0.35, -0.09]	**0.001**	-0.26	-0.26	[-0.39, -0.13]	[-0.39, -0.13]	**<0.001**
Stroop	-0.08	-0.08	[-0.22, 0.05]	[-0.21, 0.05]	0.220	-0.04	-0.04	[-0.17, 0.09]	[-0.17, 0.09]	0.565
PPI-R CN * Cooperative	0.03	0.03	[-0.11, 0.17]	[-0.11, 0.16]	0.667	0.00	0.00	[-0.13, 0.14]	[-0.13, 0.14]	0.981
PPI-R CN * Stroop	0.04	0.04	[-0.09, 0.18]	[-0.09, 0.18]	0.515	0.01	0.01	[-0.12, 0.15]	[-0.12, 0.15]	0.854
Cooperative * Stroop	-0.22	-0.22	[-0.35, -0.09]	[-0.35, -0.09]	**0.001**	-0.23	-0.23	[-0.36, -0.10]	[-0.36, -0.10]	**0.001**
PPI-R CN * Cooperative * Stroop	-0.03	-0.03	[-0.16, 0.11]	[-0.16, 0.11]	0.705	-0.02	-0.02	[-0.16, 0.11]	[-0.16, 0.11]	0.744
Observations	208					208				
R^2 / R^2 adjusted	0.116 / 0.085					0.128 / 0.098				

Note. Bolded is *p*<.05. *b* represents unstandardized regression weights. *β* indicates the standardized regression weights. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. PPI-R CN = Psychopathic Personality Inventory – Revised Carefree Nonplanfulness. Cooperative = Cooperative affective feedback. Stroop = Stroop-incongruent affective feedback.

S11. *Three-way interaction between PPI-R Machiavellian Egocentricity, affective feedback congruence, and Stroop interference on the rate and expectation of cooperation in PD.*

	Cooperation					**Expectation of Cooperation**				
Predictors	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*
(Intercept)	-0.03	0.01	[-0.16, 0.11]	[-0.12, 0.14]	0.690	-0.05	-0.00	[-0.18, 0.08]	[-0.13, 0.13]	0.448
PPI-R ME	-0.20	-0.20	[-0.33, -0.06]	[-0.33, -0.06]	**0.004**	-0.19	-0.19	[-0.32, -0.06]	[-0.32, -0.06]	**0.005**
Cooperative	-0.21	-0.21	[-0.34, -0.08]	[-0.34, -0.08]	**0.002**	-0.26	-0.27	[-0.39, -0.14]	[-0.40, -0.14]	**<0.001**
Stroop	-0.10	-0.10	[-0.23, 0.03]	[-0.23, 0.03]	0.131	-0.05	-0.05	[-0.18, 0.08]	[-0.18, 0.08]	0.433
PPI-R ME * Cooperative	0.08	0.08	[-0.05, 0.21]	[-0.05, 0.21]	0.234	0.11	0.11	[-0.02, 0.24]	[-0.02, 0.24]	0.098
PPI-R ME * Stroop	0.01	0.01	[-0.13, 0.14]	[-0.13, 0.14]	0.931	-0.01	-0.01	[-0.14, 0.12]	[-0.14, 0.13]	0.926
Cooperative * Stroop	-0.23	-0.23	[-0.36, -0.10]	[-0.36, -0.10]	**0.001**	-0.24	-0.24	[-0.37, -0.11]	[-0.37, -0.11]	**<0.001**
PPI-R ME * Cooperative * Stroop	0.07	0.07	[-0.06, 0.21]	[-0.06, 0.21]	0.273	0.04	0.04	[-0.09, 0.17]	[-0.09, 0.17]	0.557
Observations	208					208				
R^2 / R^2 adjusted	0.154 / 0.125					0.177 / 0.148				

Note. Bolded is *p*<.05. *b* represents unstandardized regression weights. *β* indicates the standardized regression weights. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. PPI-R ME = Psychopathic Personality Inventory – Revised Machiavellian Egocentricity. Cooperative = Cooperative affective feedback. Stroop = Stroop-incongruent affective feedback.

S12. *Three-way interaction between PPI-R Blame Externalization, affective feedback congruence, and Stroop interference on the rate and expectation of cooperation in PD.*

	Cooperation					**Expectation of Cooperation**				
Predictors	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*
(Intercept)	-0.03	0.01	[-0.16, 0.10]	[-0.12, 0.14]	0.678	-0.04	0.01	[-0.17, 0.09]	[-0.12, 0.14]	0.526
PPI-R BE	-0.20	-0.20	[-0.33, -0.08]	[-0.33, -0.07]	**0.002**	-0.15	-0.15	[-0.28, -0.03]	[-0.28, -0.03]	**0.017**
Cooperative	-0.23	-0.23	[-0.36, -0.10]	[-0.36, -0.10]	**0.001**	-0.27	-0.27	[-0.40, -0.14]	[-0.40, -0.14]	**<0.001**
Stroop	-0.10	-0.10	[-0.23, 0.03]	[-0.23, 0.03]	0.120	-0.05	-0.05	[-0.17, 0.08]	[-0.17, 0.08]	0.455
PPI-R BE * Cooperative	0.17	0.17	[0.04, 0.30]	[0.04, 0.30]	**0.009**	0.19	0.19	[0.07, 0.32]	[0.07, 0.32]	**0.003**
PPI-R BE * Stroop	-0.02	-0.02	[-0.15, 0.11]	[-0.15, 0.11]	0.779	0.03	0.03	[-0.10, 0.15]	[-0.10, 0.16]	0.657
Cooperative * Stroop	-0.21	-0.21	[-0.34, -0.08]	[-0.34, -0.08]	**0.002**	-0.21	-0.21	[-0.34, -0.09]	[-0.34, -0.09]	**0.001**
PPI-R BE * Cooperative * Stroop	0.00	0.00	[-0.13, 0.13]	[-0.13, 0.13]	0.964	0.04	0.04	[-0.08, 0.17]	[-0.09, 0.17]	0.519
Observations	208					208				
R^2 / R^2 adjusted	0.169 / 0.140					0.185 / 0.156				

Note. Bolded is *p*<.05. *b* represents unstandardized regression weights. *β* indicates the standardized regression weights. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. PPI-R BE = Psychopathic Personality Inventory – Revised Blame Externalization. Cooperative = Cooperative affective feedback. Stroop = Stroop-incongruent affective feedback.

S13. *Three-way interaction between PPI-R Rebellious Nonconformity, affective feedback congruence, and Stroop interference on the rate and expectation of cooperation in PD.*

	Cooperation					**Expectation of Cooperation**				
Predictors	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*
(Intercept)	-0.05	-0.01	[-0.18, 0.09]	[-0.15, 0.12]	0.479	-0.07	-0.02	[-0.20, 0.06]	[-0.15, 0.11]	0.306
PPI-R RN	-0.03	-0.03	[-0.16, 0.11]	[-0.16, 0.11]	0.710	-0.00	-0.00	[-0.13, 0.13]	[-0.14, 0.13]	0.985
Cooperative	-0.21	-0.21	[-0.34, -0.08]	[-0.34, -0.07]	**0.002**	-0.26	-0.26	[-0.39, -0.13]	[-0.39, -0.13]	**<0.001**
Stroop	-0.10	-0.10	[-0.23, 0.04]	[-0.23, 0.04]	0.150	-0.05	-0.05	[-0.18, 0.08]	[-0.18, 0.08]	0.462
PPI-R RN * Cooperative	-0.03	-0.03	[-0.17, 0.11]	[-0.17, 0.10]	0.656	-0.04	-0.04	[-0.18, 0.09]	[-0.18, 0.09]	0.535
PPI-R RN * Stroop	-0.06	-0.06	[-0.20, 0.07]	[-0.20, 0.07]	0.350	-0.07	-0.07	[-0.20, 0.07]	[-0.20, 0.07]	0.321
Cooperative * Stroop	-0.22	-0.22	[-0.36, -0.09]	[-0.35, -0.09]	**0.001**	-0.23	-0.23	[-0.36, -0.10]	[-0.36, -0.10]	**0.001**
PPI-R RN * Cooperative * Stroop	0.04	0.04	[-0.10, 0.17]	[-0.10, 0.17]	0.579	0.02	0.02	[-0.11, 0.16]	[-0.11, 0.16]	0.735
Observations	208					208				
R^2 / R^2 adjusted	0.107 / 0.076					0.129 / 0.099				

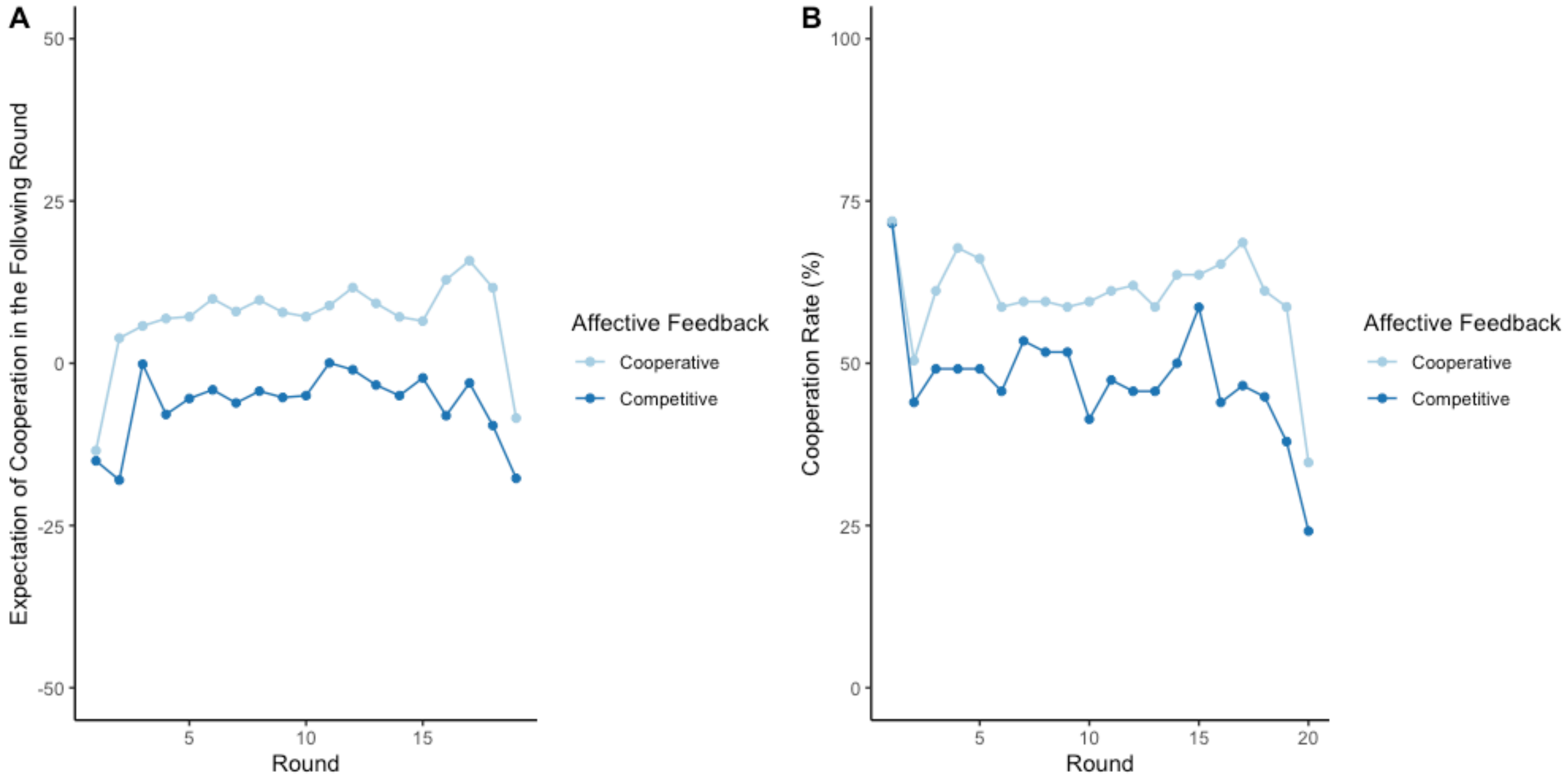
Note. Bolded is p<.05. *b* represents unstandardized regression weights. *β* indicates the standardized regression weights. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. PPI-R RN = Psychopathic Personality Inventory – Revised Rebellious Nonconformity. Cooperative = Cooperative affective feedback. Stroop = Stroop-incongruent affective feedback.

S14. *Three-way interaction between PPI-R Coldheartedness, affective feedback congruence, and Stroop interference on the rate and expectation of cooperation in PD.*

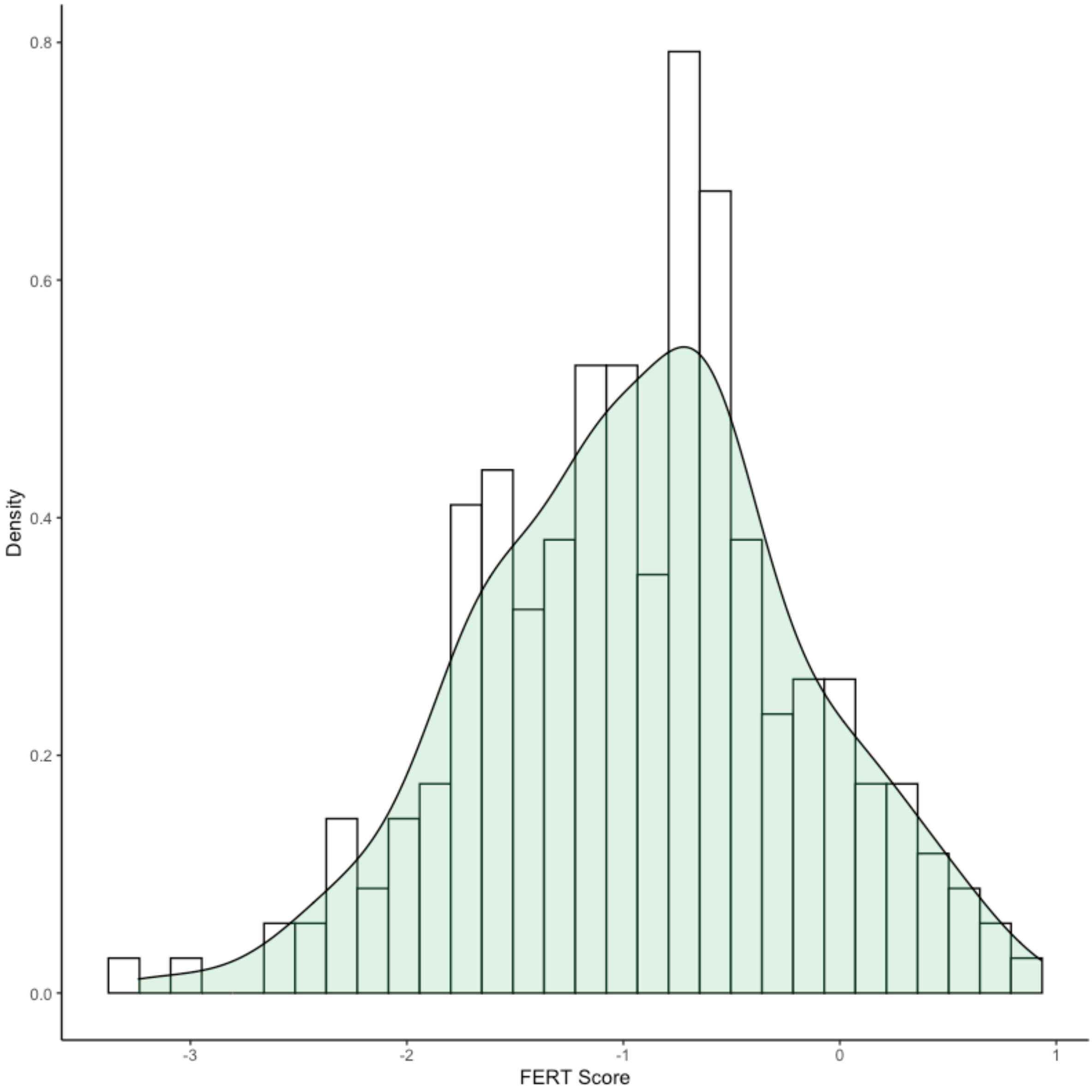
	Cooperation					**Expectation of Cooperation**				
Predictors	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*	*b*	*β*	*b* 95% CI [LL, UL]	*β* 95% CI [LL, UL]	*p*
(Intercept)	-0.05	-0.01	[-0.18, 0.09]	[-0.15, 0.12]	0.486	-0.07	-0.03	[-0.20, 0.05]	[-0.15, 0.10]	0.254
PPI-R C	0.02	0.02	[-0.12, 0.15]	[-0.12, 0.15]	0.780	-0.04	-0.04	[-0.17, 0.09]	[-0.17, 0.09]	0.562
Cooperative	-0.22	-0.21	[-0.35, -0.08]	[-0.35, -0.08]	**0.002**	-0.27	-0.27	[-0.40, -0.14]	[-0.40, -0.14]	**<0.001**
Stroop	-0.09	-0.09	[-0.22, 0.04]	[-0.22, 0.04]	0.180	-0.04	-0.04	[-0.17, 0.09]	[-0.17, 0.09]	0.506
PPI-R C * Cooperative	-0.04	-0.04	[-0.17, 0.10]	[-0.17, 0.10]	0.589	-0.00	-0.00	[-0.13, 0.13]	[-0.14, 0.13]	0.959
PPI-R C * Stroop	-0.00	-0.00	[-0.14, 0.13]	[-0.14, 0.13]	0.954	-0.04	-0.04	[-0.18, 0.09]	[-0.18, 0.09]	0.509
Cooperative * Stroop	-0.22	-0.22	[-0.35, -0.08]	[-0.35, -0.08]	**0.002**	-0.22	-0.23	[-0.35, -0.10]	[-0.36, -0.10]	**0.001**
PPI-R C * Cooperative * Stroop	-0.07	-0.07	[-0.21, 0.06]	[-0.21, 0.06]	0.285	-0.15	-0.15	[-0.28, -0.02]	[-0.28, -0.02]	**0.023**
Observations	208					208				
R^2 / R^2 adjusted	0.107 / 0.076					0.146 / 0.116				

Note. Bolded is *p*<.05. *b* represents unstandardized regression weights. *β* indicates the standardized regression weights. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. PPI-R C = Psychopathic Personality Inventory – Revised Coldheartedness. Cooperative = Cooperative affective feedback. Stroop = Stroop-incongruent affective feedback.

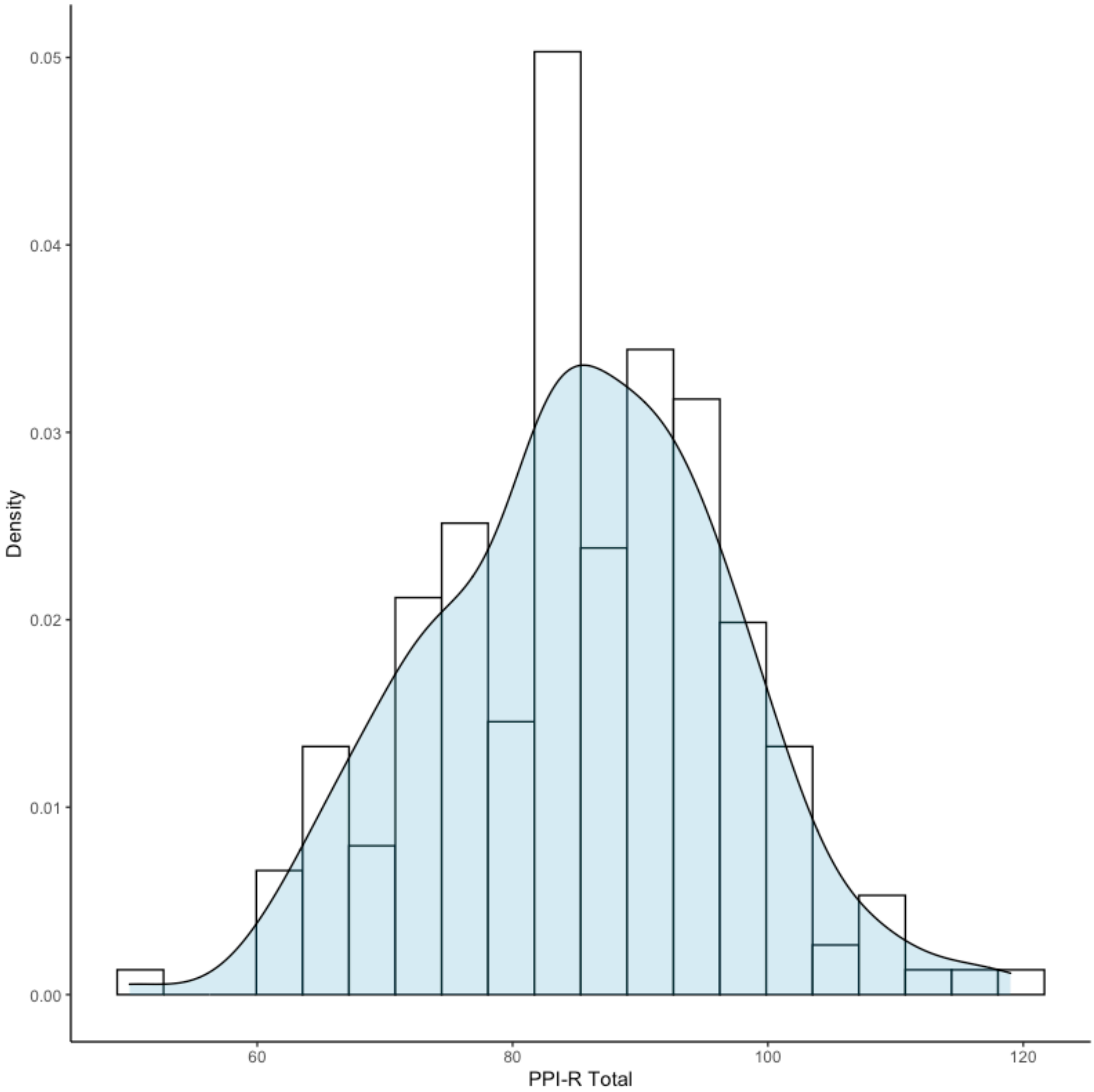
S15. *Rate and expectation of cooperation across the 20-round iterated PD.*



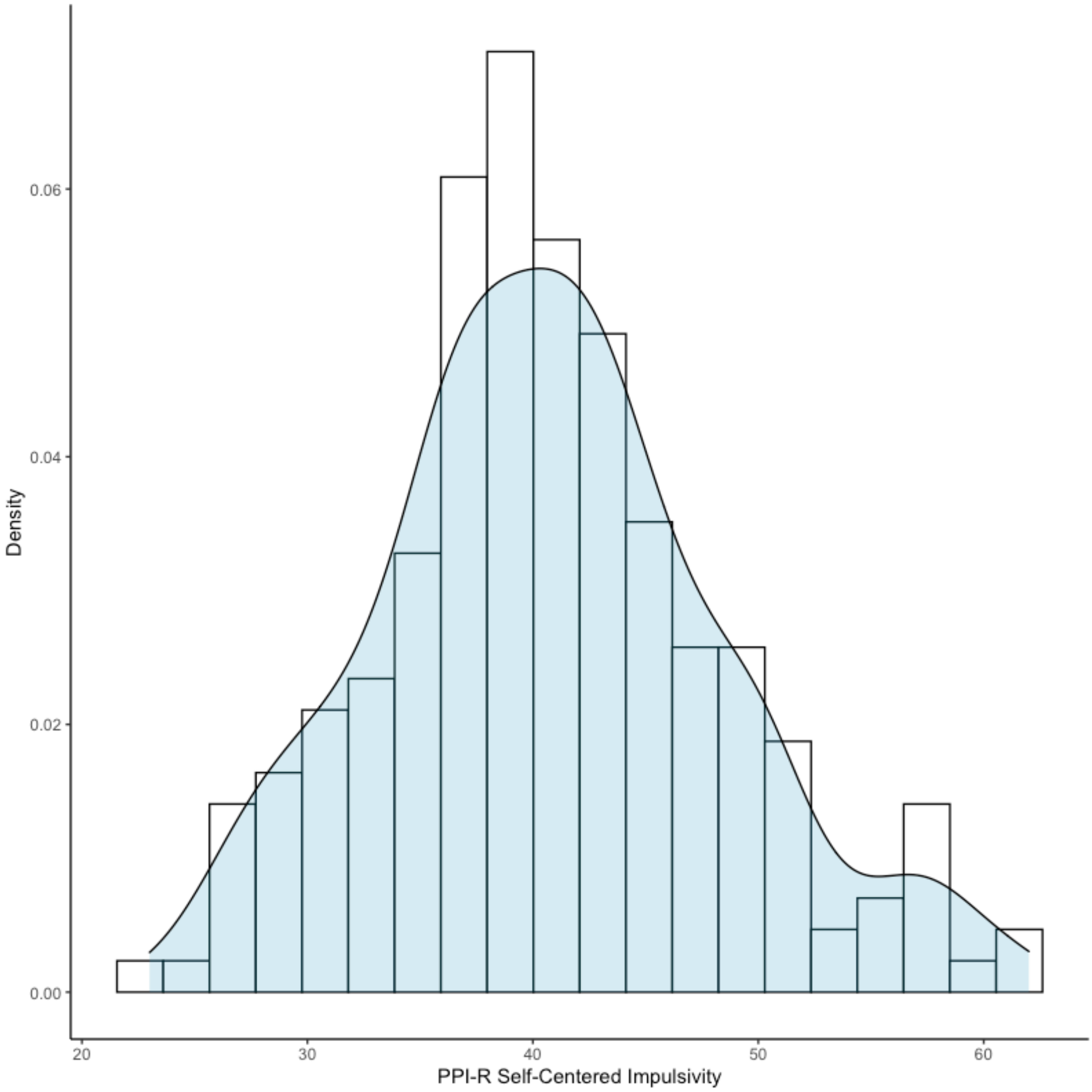
S16. *Distribution of the Facial Expression Recognition Task scores.*



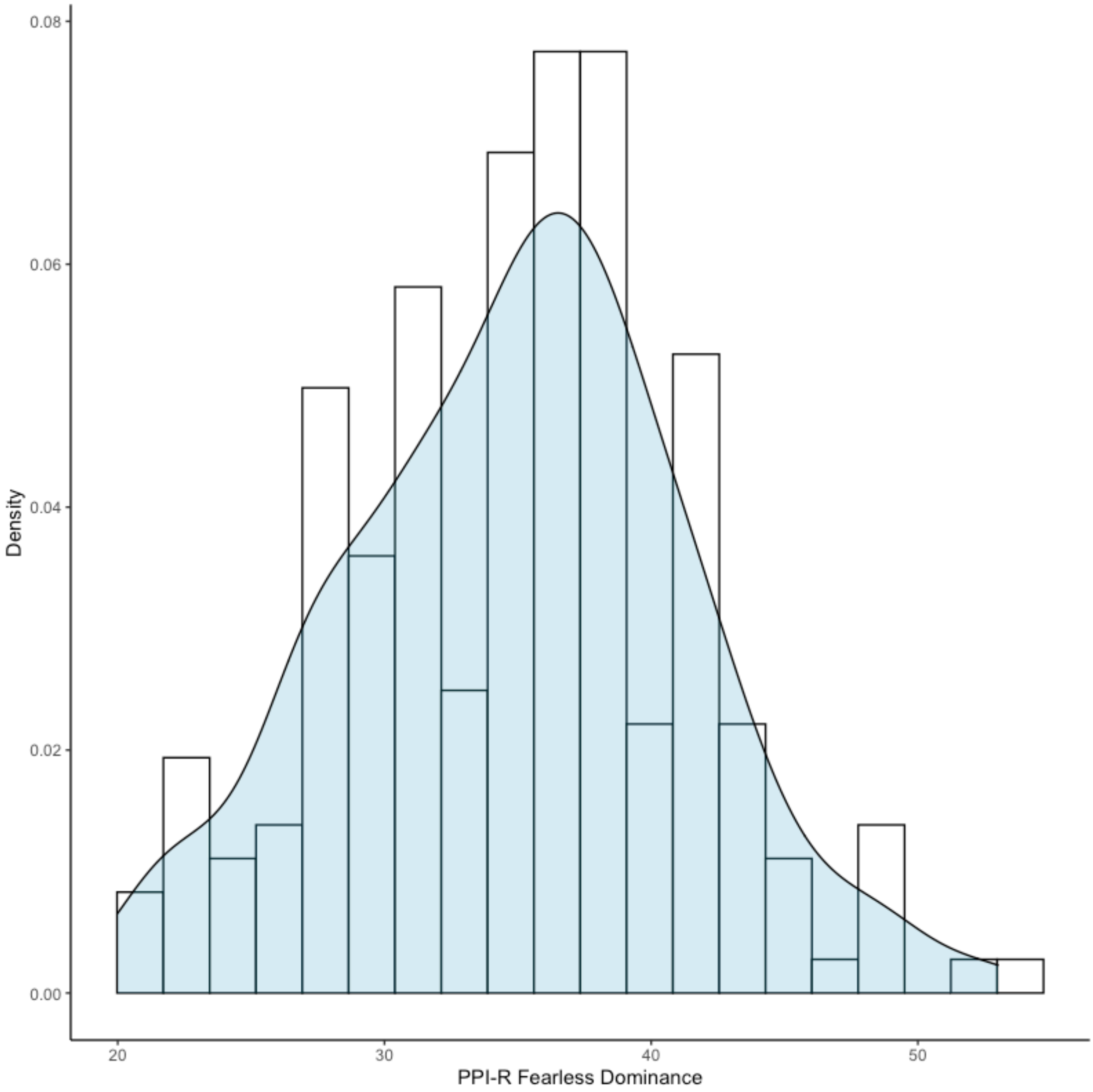
S17. *Distribution of the Psychopathic Personality Inventory – Revised total scores.*



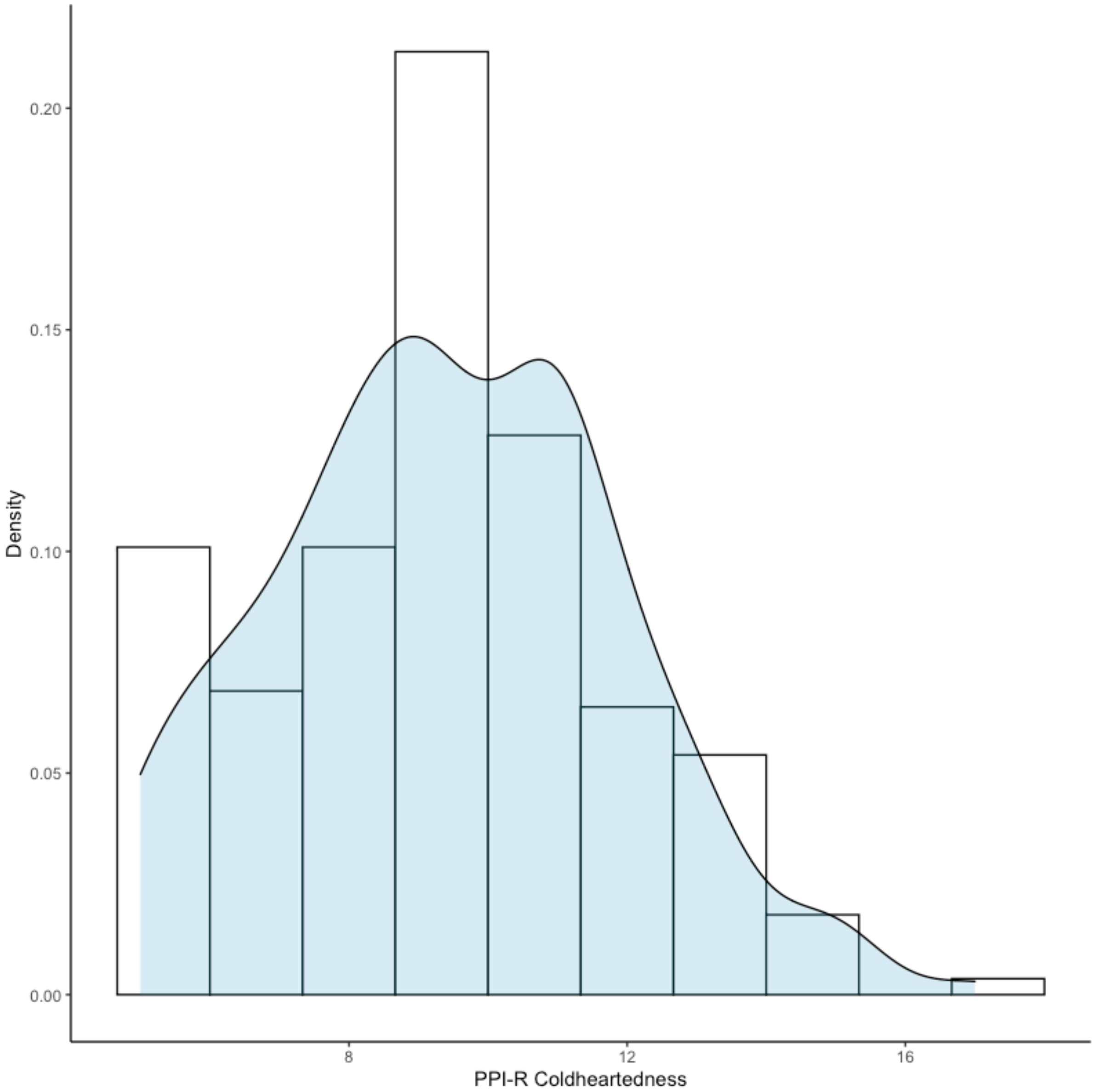
S18. *Distribution of the Psychopathic Personality Inventory – Revised Self-Centered Impulsivity Scores.*



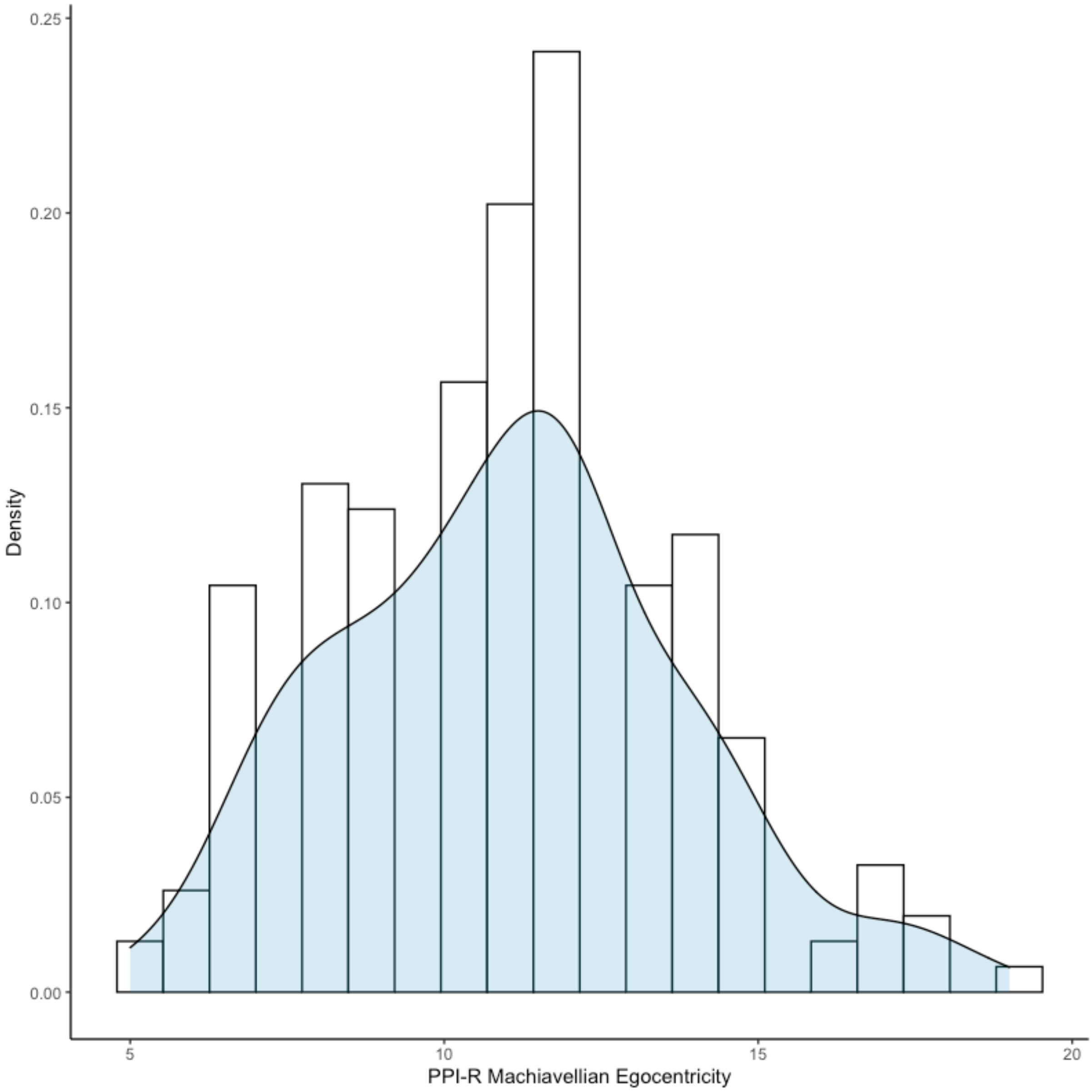
S19. *Distribution of the Psychopathic Personality Inventory – Revised Fearless Dominance Scores.*



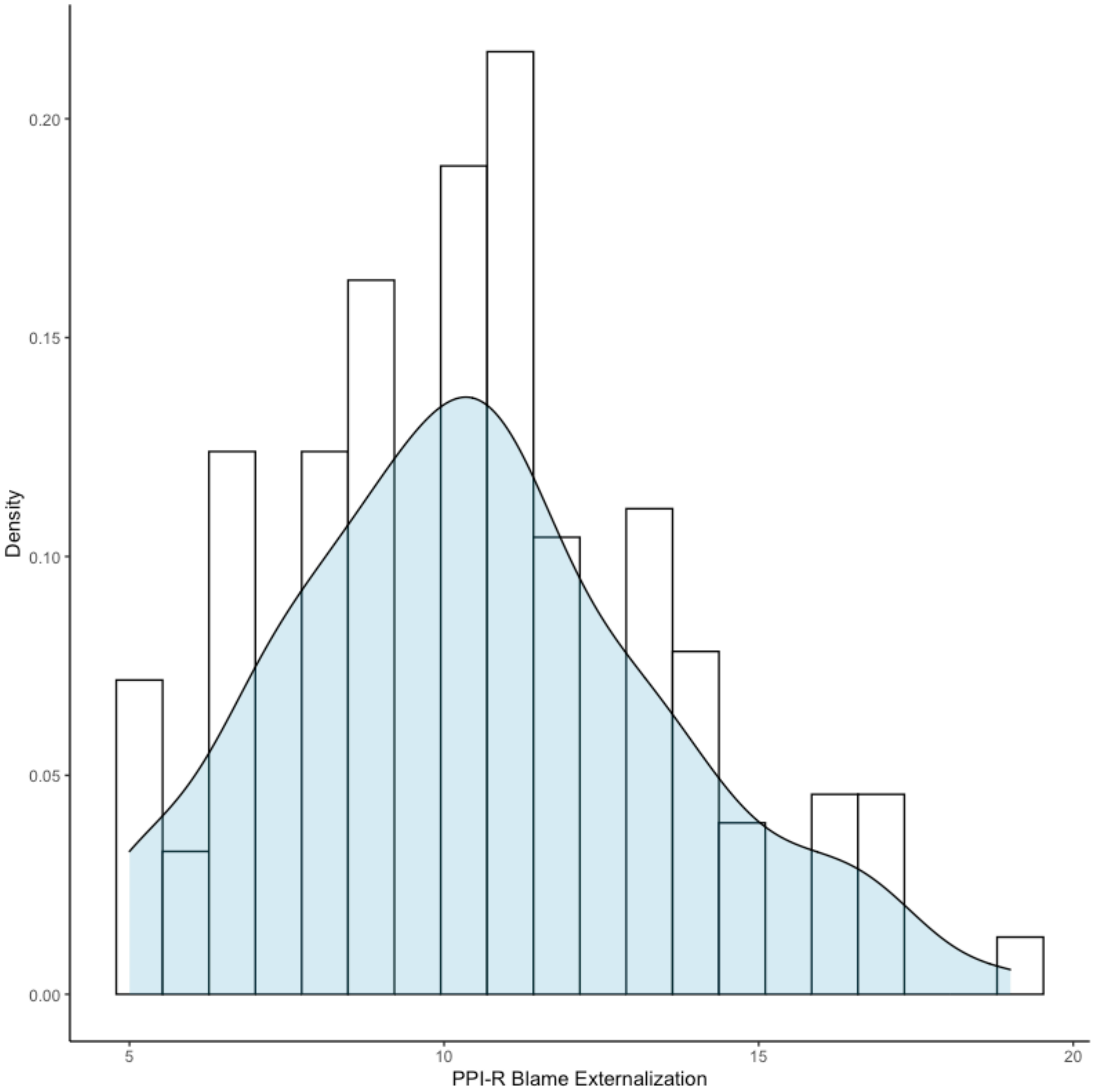
S20. *Distribution of the Psychopathic Personality Inventory – Revised Coldheartedness scores.*



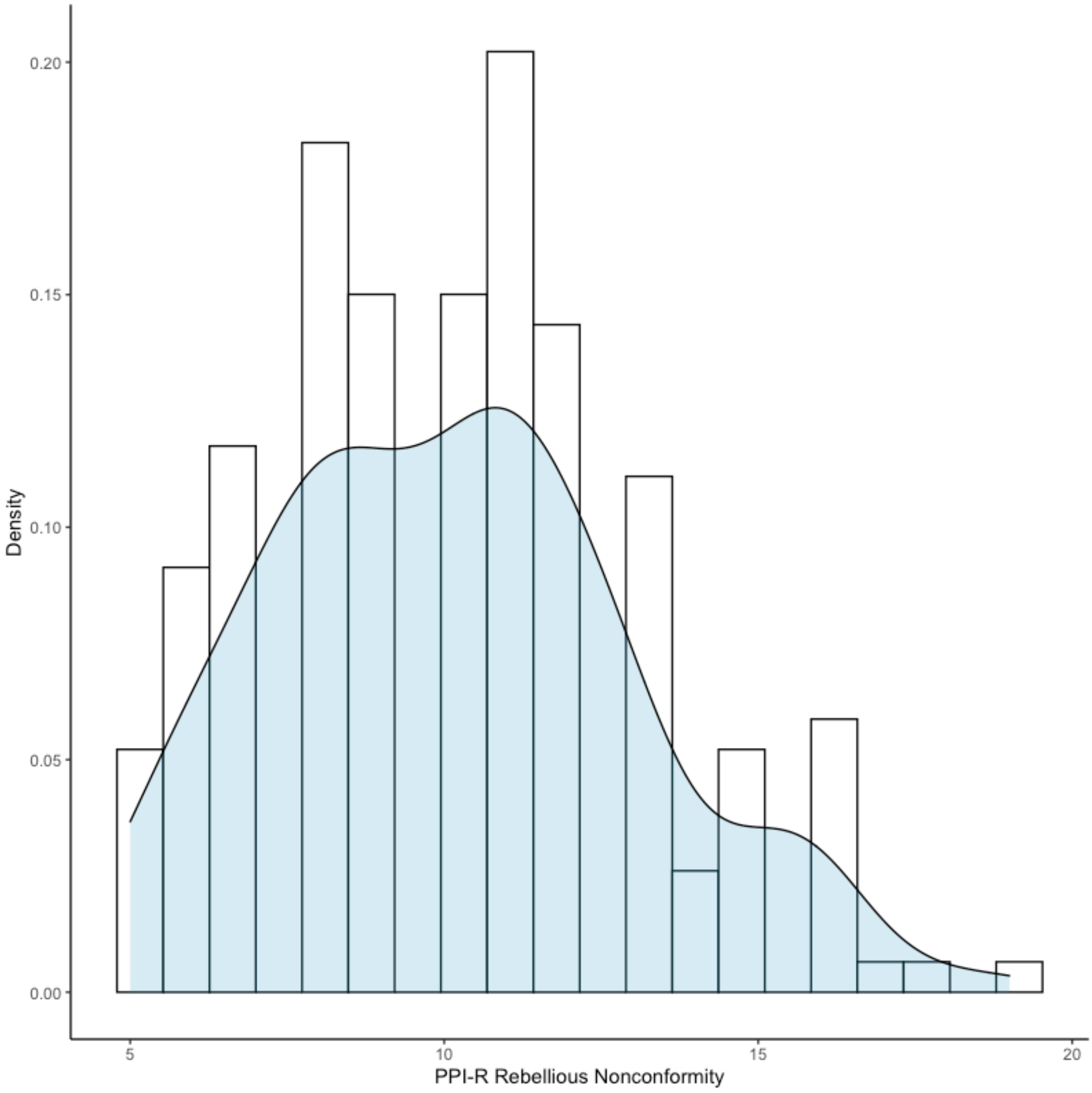
S21. *Distribution of the Psychopathic Personality Inventory – Revised Machiavellian Egocentricity scores.*



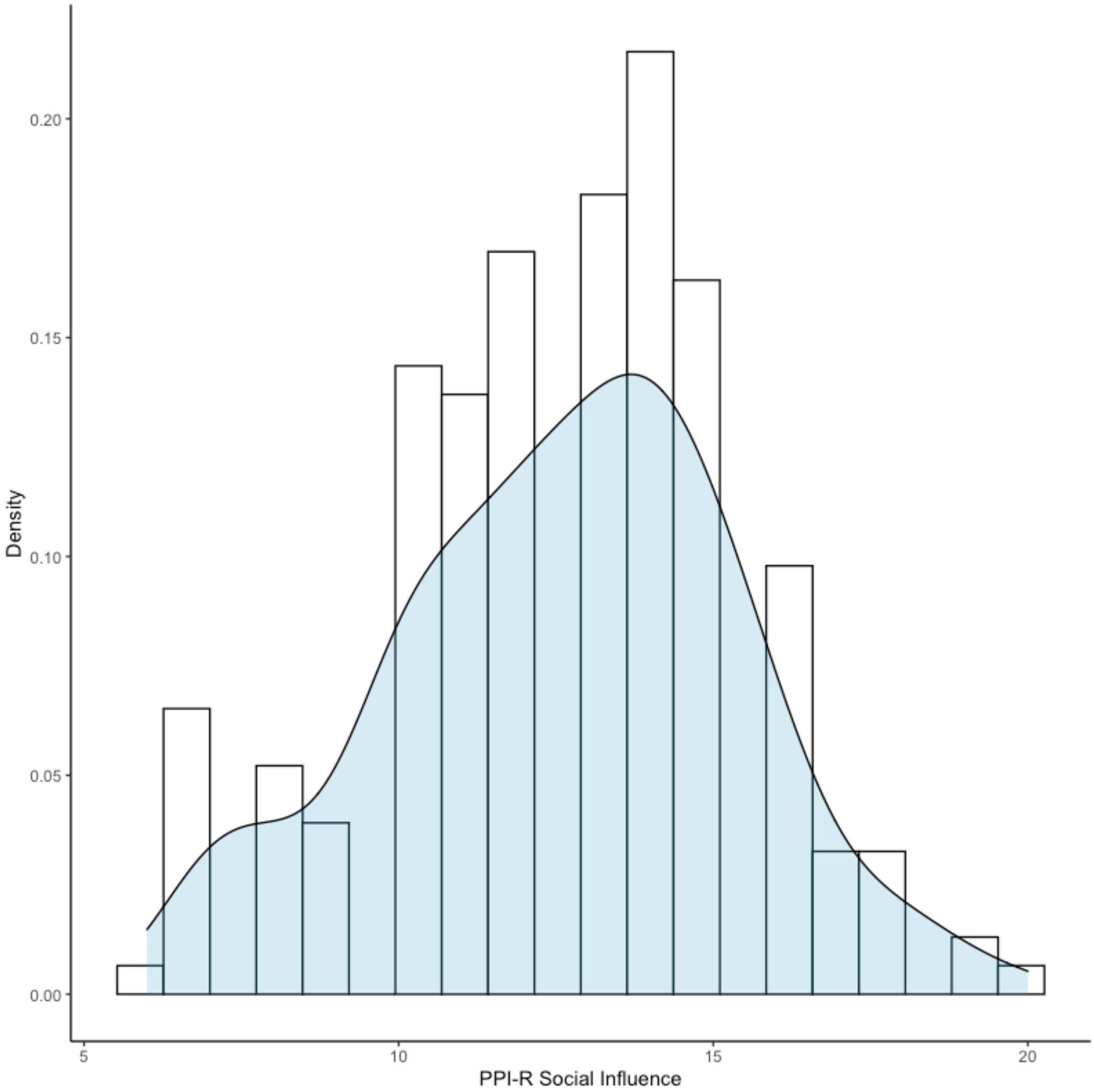
S22. *Distribution of the Psychopathic Personality Inventory – Revised Blame Externalization scores.*



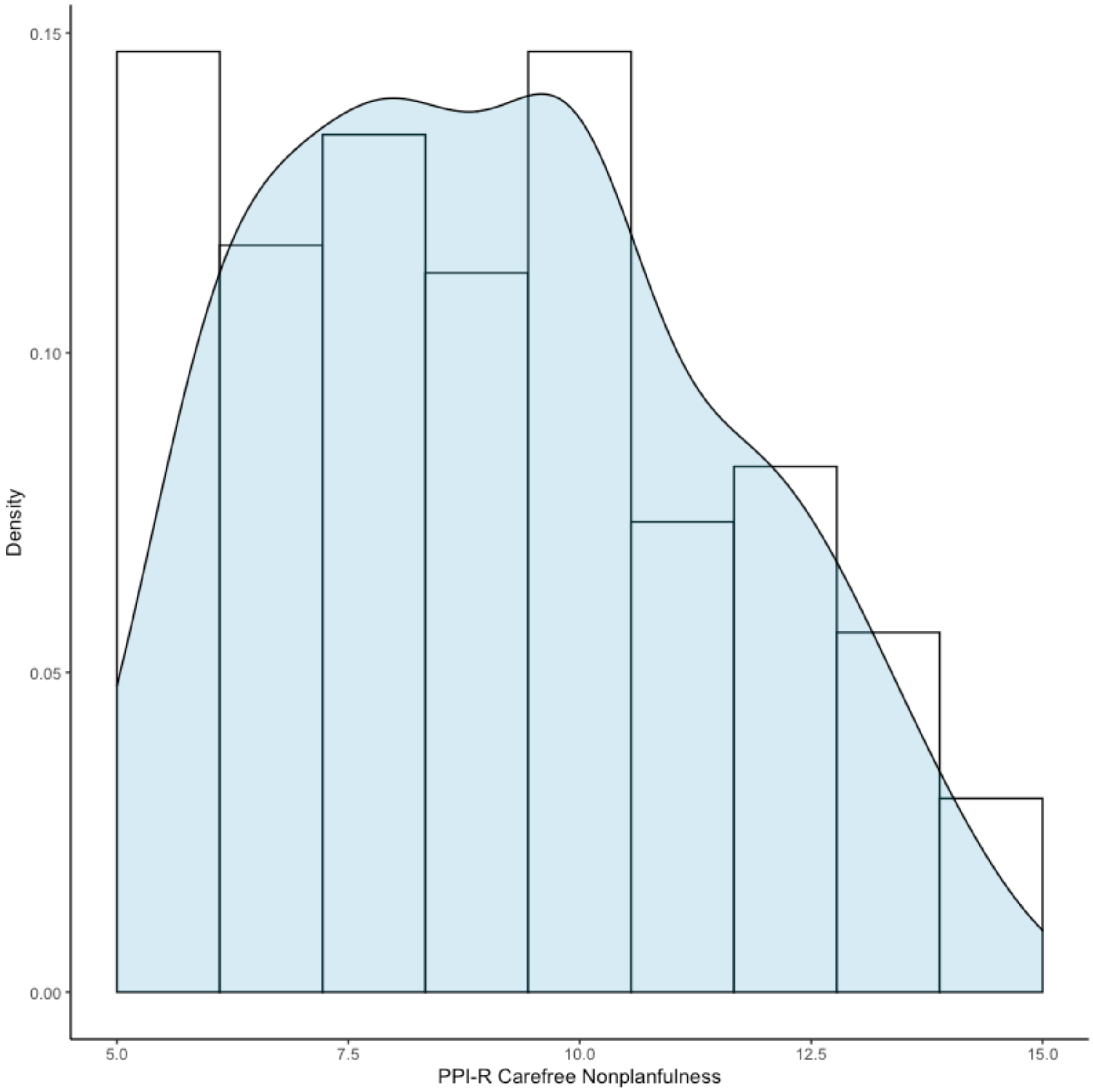
S23. *Distribution of the Psychopathic Personality Inventory – Revised Rebellious Nonconformity scores.*



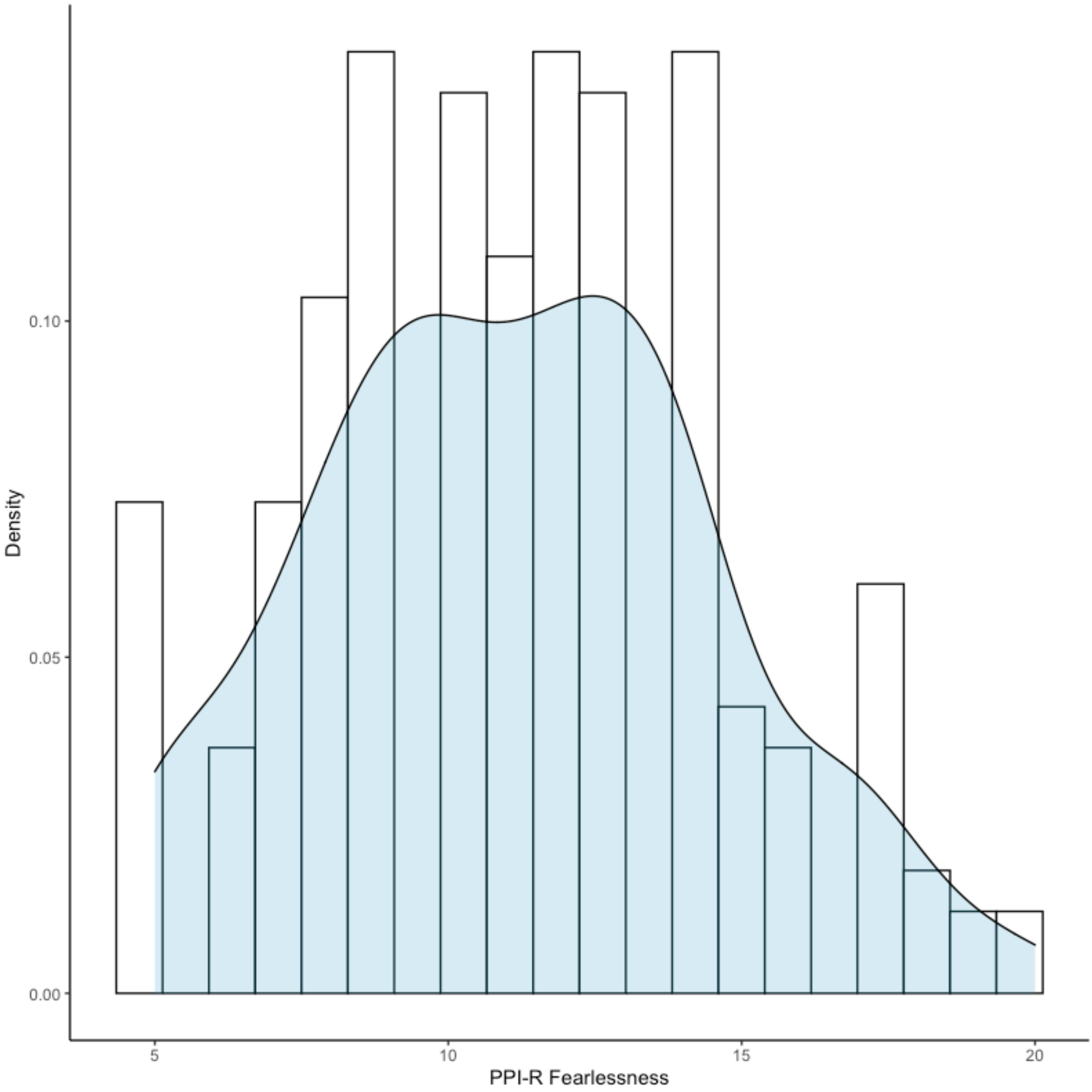
S24. *Distribution of the Psychopathic Personality Inventory – Revised Social Influence scores.*



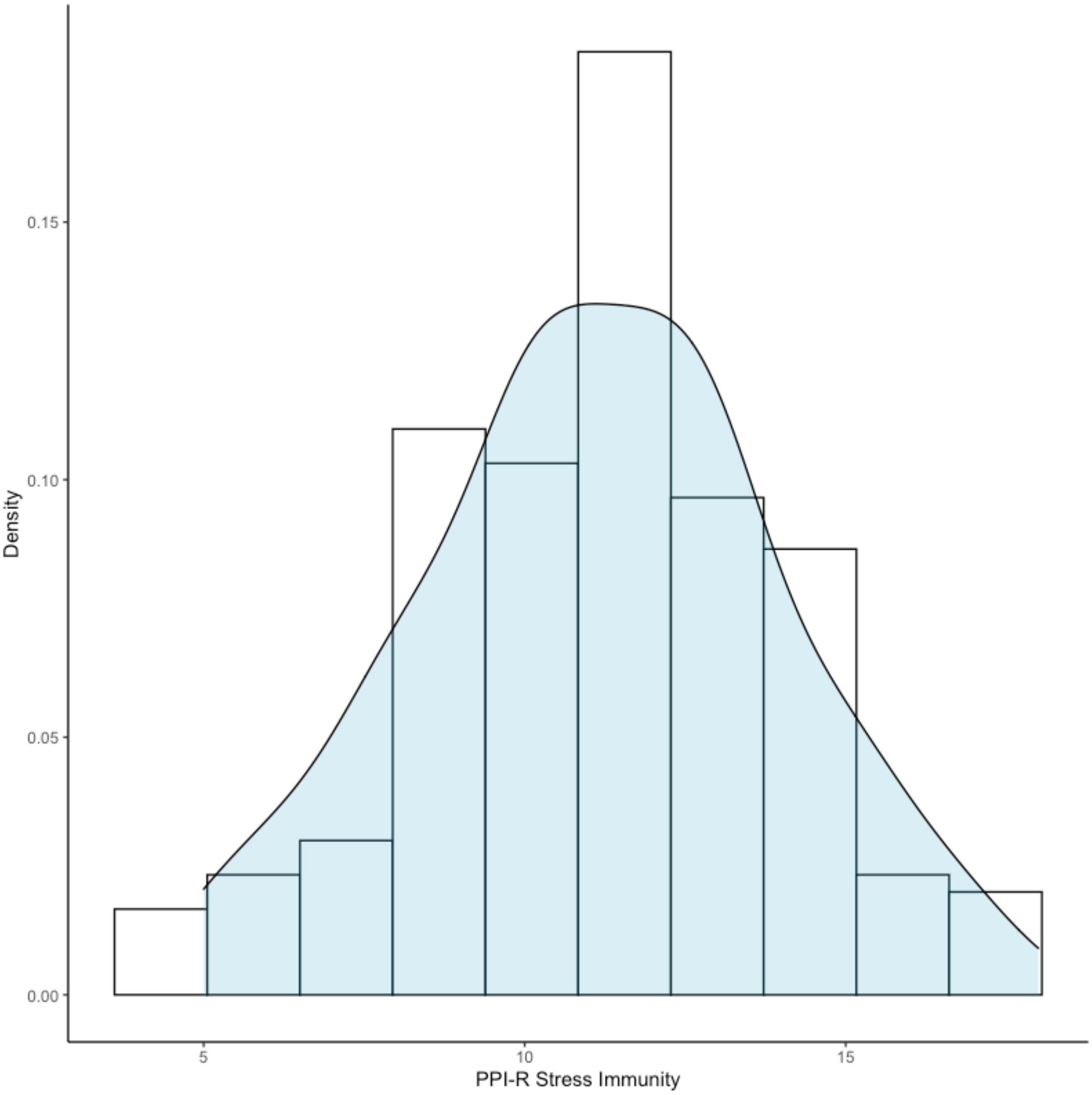
S25. *Distribution of the Psychopathic Personality Inventory – Revised Carefree Nonplanfulness scores.*



S26. *Distribution of the Psychopathic Personality Inventory – Revised Fearlessness scores.*



S27. *Distribution of the Psychopathic Personality Inventory – Revised Stress Immunity scores.*


Density
PPI-R Stress Immunity