

Law & Criminology Journal, Vol. 3(1): 82-122, 2026.

ISSN: 3041-2854

Content is licensed under CC BY 4.0 

© Ann De Buck & Lieven J.R. Pauwels

DOI: <https://doi.org/10.21825/lcj.95481>

[lcj.ugent.be](http://lcj.ugent.be)

## Empathy, Altruism, and the Moral Commitment Problem. Findings from a Prisoner's Dilemma Study

*Ann De Buck\* & Lieven J.R. Pauwels\*\**

**ABSTRACT** *Cooperation is the cornerstone of human social life, essential for survival in complex groups. Yet individuals often face the temptation to pursue short-term self-interest over long-term cooperation—a challenge known as the moral commitment problem. In many situations, external enforcement is absent or insufficient. This study examines evolved social preferences—empathy and altruism—as internal mechanisms facilitating cooperation. Using a one-shot, anonymous Prisoner's Dilemma, we found that most participants choose conditional cooperation based on trust, even in a one-time setting. Empathy and altruism significantly influence decisions, with effects varying by cost-sensitivity and sex. Individuals with strong empathic concern and medium-cost altruism are more likely to choose the “sucker” option, cooperating despite exploitation risk. While empathy and altruism promote cooperation, they can increase vulnerability to suboptimal outcomes. Overall, social preferences may act as commitment devices supporting sustained human cooperation, though the design captures only a limited aspect of complex social interactions.*

**KEYWORDS** commitment problem, prisoner's dilemma, social preferences, empathy, altruism

---

\* Ghent University, Belgium. ORCID: [0000-0002-0579-1011](https://orcid.org/0000-0002-0579-1011).

\*\* Ghent University, Belgium. ORCID: [0000-0002-4491-6374](https://orcid.org/0000-0002-4491-6374).

## 1 INTRODUCTION

Cooperation is a cornerstone of human social life, essential for survival and flourishing within complex social groups (Cliquet & Avramov, 2018; Sapolsky, 2017). Yet cooperation often involves a tension between individual self-interest and collective welfare, leading to what is known as the *moral commitment problem* (Frank, 1988; Hirshleifer, 1987; Schelling, 1960). This problem is also referred to in the literature as a social dilemma (Dawes, 1980; Ostrom, 1990) or cooperation failure or breakdown (Axelrod & Hamilton, 1981; Fehr & Fischbacher, 2003). The moral commitment problem arises when individuals are incentivized to prioritize immediate personal benefits at the expense of the group's long-term interests. It manifests in various contexts: disputes over property rights, breaches of trust in reciprocal exchanges, and failures to uphold social contracts often lead to antisocial behaviours such as theft, deception, and retaliation, undermining social cohesion (Durrant & Ward, 2015).

The Prisoner's Dilemma (PD) is a classic game-theoretical model that formalizes this tension.<sup>1</sup> It models how cooperation yields the best collective outcome, yet each individual is tempted to cheat—to make the selfish choice—thereby undermining group welfare (Axelrod, 1984; Rand & Nowak, 2013). Although the PD clearly illustrates this strategic conflict, it does not explain why some individuals resist the temptation to cheat and opt for cooperation instead.

This study focuses specifically on the role of evolved social preferences—particularly empathy and altruism—as potential psychological mechanisms that enable individuals to overcome self-interest and cooperate in situations involving commitment problems. Using the PD as a semi-experimental study design, we examine whether these evolved social preferences reduce the likelihood of cheating, thereby fostering cooperative choices.

## 2 THE MORAL COMMITMENT PROBLEM

Cooperation is undeniably a defining characteristic of human societies. While not unique to humans, the scale and complexity of human cooperation far exceed those observed in other mammals (Apicella & Silk, 2019; Henrich & Muthukrishna, 2021). Across cultures, people help one another by sharing, teaching, caregiving, and enforcing social norms, making humans an ultra-social and super-cooperative species (Nowak & Highfield, 2012; Tomasello, 2016, 2019, 2024). Yet individuals continuously face the challenge of balancing short-term self-interest against the long-term benefits of sustaining cooperation—a challenge captured by the moral commitment problem.

---

<sup>1</sup> Prisoner's dilemma was first devised in 1950 by Merrill Flood and Melvin Dresher, who worked at RAND in Santa Monica, California. The original think tank, the RAND (Research and Development) Corporation was founded as Project RAND in December 1945 by the U.S. Army Air Force and by defense contractors (Nowak & Highfield, 2011).

## 2.1 The moral commitment problem

The moral commitment problem arises because individuals, shaped by unique interests and competing for finite resources (Alexander, 1979, 1987), are often tempted to pursue immediate personal gain even when it conflicts with the group's welfare. When such temptations prevail, cooperation can break down: rules are violated, trust is betrayed, and behaviours like cheating or aggression destabilize the social order (Durrant & Ward, 2015). Although these situations are sometimes described as “cooperation failures”, they ultimately reflect the same core issue: the difficulty of sustaining commitment to shared norms when self-interest is at stake (Frank, 1988).

Frank (1988, 2001, 2011) illustrates the commitment problem<sup>2</sup> through a series of real-world examples that are particularly relevant in legal and institutional contexts. In business partnerships, for instance, individuals with complementary skills may benefit from cooperation, yet each has opportunities to cheat without detection, such as skimming profits or accepting kickbacks. In liability contexts, a party may cause foreseeable harm (e.g., by failing to fence in livestock), but the cost of legal action may exceed the damages, making threats of litigation non-credible. In bargaining situations, the party under less pressure often wields greater leverage—unless the weaker party can credibly threaten to walk away from unfair offers. Even in marriage, individuals may hesitate to invest fully without assurances of long-term commitment—yet legal mechanisms like prenuptial agreements rarely eliminate this uncertainty.

These examples highlight two key insights (Frank, 1988, 1993). First, the commitment problem is not a theoretical abstraction; it is a pervasive challenge in law, economics, and everyday life. Second, while binding commitments often improve outcomes, the mechanisms to enforce them are frequently vague, impractical, or legally unenforceable. This underlines the potential importance of internal commitment mechanisms—such as evolved social preferences—that may help individuals uphold cooperation even in the absence of external enforcement (Pauwels & De Buck, 2026).

## 2.2 The Prisoner's Dilemma

To analyse these real-world challenges more systematically, we turn to a formal model that captures their underlying strategic structure. The Prisoner's Dilemma is perhaps the best-known and most widely studied framework for representing this tension. It distills the core conflict between individual and collective interests

---

<sup>2</sup> Building on Frank's (1988) analysis of the commitment problem, we introduce the term *moral commitment problem* to describe situations in which individuals uphold cooperative behaviour not because of external enforcement, but due to evolved and internalized social preferences such as empathy, altruism, or guilt. While Frank does not use this exact phrase, his work suggests that emotions serve as honest signals of commitment, allowing individuals to maintain cooperation in the face of short-term temptations to cheat. The *moral* dimension emphasises the role of evolved psychological dispositions that make cooperation possible even when cheating would be materially advantageous.

into a simple yet powerful structure: although cooperation maximizes group welfare, each faces a strong temptation to cheat. This logic mirrors the dynamics observed in Frank’s examples, where cooperation would be beneficial, but individual incentives often pull in the opposite direction.

The structure of the Prisoner’s Dilemma can be formalized using a payoff matrix that illustrates the outcomes of mutual cooperation, mutual cheating, and unilateral cheating. As shown in Table 1, mutual cooperation yields a reward for both players, while unilateral cooperation results in the *sucker’s payoff*—a disadvantage for the cooperator when the other cheats. Conversely, unilateral cheating offers the *temptation payoff*, benefiting the cheater at the cooperator’s expense. If both players cheat, they receive the *punishment payoff*, which is worse than mutual cooperation but better than being exploited. This simple matrix illustrates real-world commitment problems: although mutual cooperation is collectively optimal, individual incentives often favour cheating (Rand & Nowak, 2013; Rapoport & Chammah, 1965).

Despite variations in formulation, such as the stakes involved, the severity of punishment, or the context in which cooperation is tested, the central logic remains the same. It can be represented by a payoff matrix. This matrix not only summarizes the four possible outcomes of the game but also reflects the basic tensions of life itself: between self-interest and mutual benefit, between short-term gain and long-term trust (Nowak & Highfield, 2011).

**Table 1**

*Payoff matrix of the Prisoner’s Dilemma (Adapted from Rand & Nowak, 2013)*

	<b>1 Cooperates</b>	<b>1 Cheats</b>
<b>2 Cooperates</b>	Reward for cooperation	Individual 2: Sucker’s payoff Individual 1: Temptation payoff
<b>2 Cheats</b>	Individual 2: Temptation payoff Individual 1: Sucker’s payoff	Punishment for mutual cheating

*Note.* This matrix represents the four possible outcomes in a one-shot Prisoner’s Dilemma. Columns represent the choices of Individual 1, and rows represent the choices of Individual 2. Each cell specifies the outcomes for both players. When both cooperate, each receives the mutual reward. When one cheats and the other cooperates, the cheating player receives the temptation payoff, while the cooperating player receives the sucker’s payoff. When both cheat, each receives the punishment payoff for mutual cheating, which is worse than cooperation but better than being exploited. This clarified presentation makes explicit which payoff applies to which individual in each outcome. Adapted from Rand & Nowak (2013).

Despite the strategic logic of the Prisoner’s Dilemma and the real-world examples of commitment problems, many individuals routinely refrain from cheating, even when the risk of detection is negligible. As Frank (1988) observes, there are countless everyday opportunities to cheat without consequence: keeping found money, padding expense reports, littering in isolated places, or disabling environmental controls in one’s car. While such acts are difficult to detect and often materially advantageous, not everyone engages in them. If they did, the consequences—dirtier beaches, more polluted air, widespread tax evasion—

would be far more visible. This behavioural paradox raises a fundamental question: why do some people consistently choose not to cheat, even when doing so would serve their material self-interest? Cultural norms and legal deterrents may play a role, but they are often insufficient to explain consistent moral restraint, especially in situations where enforcement is weak or absent. This paradox challenges purely materialist models of human behaviour, which struggle to account for the persistence of unselfish conduct. One possible explanation is that individuals who refrain from cheating may be observably different from others, and that these differences are linked to internal psychological mechanisms that support cooperation. Importantly, the Prisoner's Dilemma does not explain why some individuals systematically resist selfish impulses. While the strategic structure highlights the tension between self-interest and collective benefit, the model remains silent on the determinants that shape actual decision-making. Why do some individuals consistently resist the temptation to cheat, even when doing so would serve their material interests? And under what conditions do people uphold commitments, even in the absence of external enforcement?

To address these questions, the next section explores a set of psychological factors that have been proposed to influence choices in commitment situations. Building on Robert Frank's (1988, 2011) evolutionary perspective on the commitment problem, this section explores psychological factors that may underpin individuals' ability to make cooperative choices despite temptations to defect. These factors include personality traits such as impulsivity, risk-taking, and self-regulation, as well as evolved social preferences like empathy (both affective and cognitive) (Tomasello, 2016) and altruism (Batson, 2016).

### **3 PERSONALITY TRAITS AND EVOLVED SOCIAL PREFERENCES**

This section examines the individual-level determinants that influence decision-making in situations entailing a commitment problem. Building on the commitment problem outlined earlier, we explore the role of several dispositional determinants—personality traits and evolved social preferences—that may shape the likelihood of cooperative versus selfish choices.

#### **3.1 Personality traits**

Personality traits are relatively stable internal characteristics that serve as the building blocks of personality. They influence how individuals typically think, feel, and behave across various situations and over time (Larsen & Buss, 2021). These traits function as shorthand descriptors for behavioural tendencies and systematically differ between individuals (American Psychological Association, n.d.). From a dispositional perspective, traits are defined by their stability over time and consistency across contexts. Early trait theories, such as those proposed by Allport and Odbert (1936), Cattell (1973), and Eysenck (1991), offer descriptive taxonomies of recurring behavioural patterns. However, these approaches often stop short of explaining the underlying mechanisms. More recent perspectives emphasise

that traits influence decision-making processes by shaping how individuals evaluate options, anticipate outcomes, and regulate impulses (Mitchell, 2018). Some theorists go further, conceptualizing traits as conditional dispositions—patterns of behaviour that emerge only in particular contexts (Mischel & Shoda, 1995; Fleeson, 2013). From this view, personality is not merely a set of relatively stable tendencies but a dynamic system of responses that interact with situational cues.

This section focuses on three traits that are especially relevant to decisions involving cooperation and rule-following: impulsivity, risk-taking, and self-control (here defined as self-regulation; see below for further discussion). Although these traits are interrelated, each reflects a distinct underlying process with different implications for behaviour.

### ***Impulsivity***

Impulsivity refers to acting on urges without sufficient forethought or consideration of consequences (APA, n.d.). It involves a preference for immediate rewards, diminished inhibitory control, and heightened approach motivation to attractive stimuli (Gray, 1975). Often regarded as a failure of self-regulation, impulsivity reflects impaired ability to inhibit maladaptive behaviours (Zuckerman, 1994). It is not a unitary construct but a multidimensional disposition, encompassing tendencies such as acting without deliberation, discounting future outcomes, and failing to inhibit automatic responses. These facets are measurable through self-report and behavioural tasks assessing delay of gratification, response inhibition, and decision-making under uncertainty, which reveal stable individual differences (Mitchell, 2018). Importantly, genetic and neurobiological research has identified specific brain regions—such as the prefrontal cortex, orbitofrontal cortex, and basal ganglia—as well as neural circuits involving brainstem neuromodulatory systems, particularly the dopaminergic and serotonergic systems, as key contributors to impulsivity (Mitchell, 2018). Mitchell’s hierarchical model further conceptualizes impulsivity as a mid-level trait influenced by lower-level decision-making parameters like novelty salience, delay discounting, and punishment sensitivity, which feed into higher-level constructs including Extraversion, Conscientiousness, and Neuroticism.<sup>3</sup>

### ***Risk-taking***

Risk-taking, by contrast, involves a willingness to engage in behaviours with uncertain or potentially negative outcomes, typically in pursuit of high rewards. This tendency may reflect either insensitivity to risk or an active preference for uncertainty (APA, n.d.). Closely related is sensation seeking, defined as the pursuit of novel, complex, and intense experiences—even at the cost of physical, social, or legal risks (Larsen & Buss, 2021). While impulsivity and risk-taking can lead to similar behaviours, such as cheating, aggression, or rule-breaking, they differ

---

<sup>3</sup> Extraversion, Conscientiousness, and Neuroticism are three of the five traits of the Big Five (Five-Factor Model) of personality, as operationalized in the NEO Personality Inventory (Costa & McCrae, 1992).

conceptually: impulsivity reflects insufficient forethought, whereas risk-taking involves calculated engagement with potential loss. Zuckerman (1994) proposed that sensation seeking and impulsivity share overlapping features and may jointly form a broader behavioural style, which he termed impulsive sensation seeking—a super trait marked by strong approach motivation, low inhibition, and heightened sensitivity to immediate rewards. Steinberg et al. (2008) further emphasise the distinction between impulsivity and risk-taking by showing that these traits follow different developmental trajectories: risk-taking peaks in adolescence due to heightened reward sensitivity, whereas impulsivity tends to decline more steadily with the maturation of cognitive control systems.

### *Self-regulation*

Self-regulation—or self-control—is the capacity to monitor and regulate behaviour, emotions, and impulses, especially when immediate temptations conflict with long-term goals (APA, n.d.). Milyavskaya, Berkman, and De Ridder (2018) highlight that “self-control” is often inconsistently used and thus define it precisely as overcoming a temptation or dominant response in favour of a competing goal. Similarly, Mischel, Cantor, and Feldman (1996) describe it as a dynamic regulatory process aligning behaviour with intentions over time. Unlike impulsivity and risk-taking, which often represent failures of inhibition or heightened reward sensitivity, self-regulation supports deliberation, consistency, and planning (Nigg, 2017). It involves overriding a dominant, stimulus-driven response by activating internally generated representations linked to long-term goals. As such, this process is important for sustaining cooperation, ethical behaviour, and goal-directed persistence.

These variables, impulsivity, risk-taking, and self-regulation, are central to criminological theories explaining criminal behaviour. Gottfredson and Hirschi (1990) defined low self-control as a stable trait expressed through impulsivity, egocentrism, risk-taking, and preference for immediate gratification, positing it as the primary cause of deviance. Meta-analyses support low self-control as a strong predictor of criminal and antisocial behaviour (Pratt & Cullen, 2000; Vazsonyi et al., 2017). However, more recent research emphasises impulsivity and risk-taking as especially predictive of delinquency (Jones, Miller & Lynam, 2001; Zuckerman & Kuhlman, 2000). Moreover, current views advocate a nuanced understanding of self-control as a dynamic, developmentally sensitive set of self-regulatory processes influenced by underlying dispositions like impulsivity and risk preference (Burt, 2020). This aligns with models viewing self-control not as a fixed trait but as composite processes modulated by situational factors (Miller & Cohen, 2001).

From a neurobiological perspective, Sapolsky (2017) frames self-control as a maturational achievement of the prefrontal cortex—one that enables individuals to “do the harder thing when it's the right thing to do” (p. 25). This capacity relies on the delayed development of the prefrontal cortex and its increasing top-down regulation over subcortical, affectively driven regions. In

Sapolsky's view, self-control is not merely a dispositional trait but a biologically grounded regulatory function that develops through adolescence and early adulthood, reflecting increased cognitive control and foresight. This framing underscores the developmental, effortful, and context-sensitive nature of self-regulation and highlights its distinction from more or less stable traits like impulsivity and risk-taking.

In sum, many key individual differences and personality traits manifest most strongly in interpersonal contexts (Larsen & Buss, 2021), where they shape cooperation, trust, and conflict. This broader framework not only reaffirms the theoretical relevance of personality traits like impulsivity, risk-taking, and self-regulation but also underscores their importance in explaining real-world failures of commitment and adherence to rules. Consequently, we adopt the term *self-regulation* to refer specifically to these dynamic, process-oriented mechanisms of behaviour regulation, thereby distinguishing it from relatively stable traits like impulsivity and risk-taking, which are often precursors to regulatory capacity. The term *self-control* is retained when citing frameworks that explicitly use this terminology.

### 3.2 Evolved social preferences

To define evolved social preferences, it is useful to distinguish between evolutionary (ultimate) and proximate explanations—a distinction originally introduced by Mayr (1961) and Tinbergen (1963). Evolutionary explanations address *why* certain social behaviours or preferences evolved by highlighting their adaptive value, whereas proximate explanations focus on *how* these preferences are expressed through psychological mechanisms such as cognition, emotion, and motivation that drive behaviour in specific contexts. In the case of social preferences, evolutionary theories suggest that these tendencies emerged because they solved recurring coordination problems in small-scale social groups (Tomasello, 2016). Mechanisms such as kin selection (Hamilton, 1964) and reciprocal altruism (Trivers, 1971, 2001) provided a selective advantage for individuals who were motivated to help kin or trusted partners, thus increasing inclusive fitness or long-term reciprocity. These selective pressures likely gave rise to psychological dispositions—such as empathy and altruistic tendencies—that predispose individuals to care preferentially for others under specific conditions, particularly within kin networks and close social groups (Krebs, 2011). From a proximate perspective, such preferences manifest relatively stable psychological tendencies that are activated in social interactions and shape how individuals evaluate the interests of others alongside their own (van Lange et al., 2014; Camerer & Fehr, 2004).

#### *Empathy*

Empathy and sympathy are widely regarded as foundational to moral development, a view already expressed by 18th-century philosophers such as Adam Smith and David Hume. In *The Theory of Moral Sentiments* (1759), Smith

distinguished between spontaneous emotional resonance and the more deliberate capacity to imagine another's emotional state. The term *empathy*—derived from the German *Einfühlung*, meaning “to feel into”—was introduced much later by Titchener (1909), yet it continues to be defined in various ways. This conceptual ambiguity is also reflected in the literature. In a review of 43 definitions, Cuff et al. (2016) highlight the multidimensional nature of empathy and the importance of distinguishing it from related concepts such as sympathy or compassion. Hein and Singer (2008) define empathy as “feeling as” another, in contrast to sympathy's “feeling for”—a distinction supported by neuroimaging studies showing partially distinct neural circuits (Decety & Michalska, 2010).

Many scholars conceptualize empathy as a layered construct involving emotional, motivational, and cognitive components (Goetz & Simon-Thomas, 2024). Empathic concern likely has evolutionary roots in parental care, with parallels in primates (de Waal, 2010; Preston & de Waal, 2002). De Waal's “Russian doll” model (2010) captures this structure: from basic affective contagion to more advanced perspective-taking and targeted helping, supported by prefrontal systems (Decety, 2011). A key feature of empathy is its contextual flexibility. Empathic responses are typically amplified in cooperative contexts and suppressed in antagonistic ones (Krebs, 2022). This flexibility allows empathy to extend beyond kin to strangers and even non-human others, guided by cognitive appraisal. As such, it underpins empathy's role as a foundational mechanism for cooperation, although the extension of altruism to modern, large-scale societies likely depends on shared norms and cultural institutions. Tomasello (2016, 2019) argues that while evolved socio-affective capacities like empathy provide the motivational foundation for prosociality, their large-scale application requires culturally shaped moral frameworks that support norm-based cooperation among unrelated individuals.

Scholars typically distinguish between affective empathy, or empathic concern—an emotional response motivated by another's need—and cognitive empathy, or perspective-taking—the ability to infer another's thoughts and feelings (Batson, 2011; Davis, 2006; Blair, 2005; Pizarro, 2000). Empathic concern is associated with moral emotions such as guilt and compassion (Hoffman, 2001; Tangney & Dearing, 2002) and plays a key role in prosocial motivation (Pinker, 2011; Ward & Durrant, 2015). However, empathy does not always lead to helping. Its effects are shaped by group membership, bias, and social norms (Bloom, 2017; Ward & Durrant, 2013). Given these complexities, researchers increasingly argue against treating empathy as a monolithic construct. In line with this view, the present study measures empathic concern and perspective-taking separately, allowing for a more precise investigation of how distinct empathic processes influence cooperative decision-making.

### ***Altruism***

Altruism is typically defined as the motivation to enhance another's welfare, even when doing so involves a personal cost (Batson, 2011). Evolutionary explanations

focus on the fitness consequences of altruistic behaviours such as increased survival of kin or reciprocation over time (Sober & Wilson, 1998) and emphasise that cooperation and altruism are fundamental adaptations<sup>4</sup> that evolved, at least in part, to support social living (Sussman & Cloninger, 2011). Ricard (2015) further argues that altruism should be seen primarily as a benevolent motivation: a stable disposition to value others' well-being for its own sake. Psychological models emphasise the intention behind the act, regardless of its outcomes. Importantly, altruistic behaviour does not require suffering or self-denial; what matters is that the motivation is genuinely other-regarding, not instrumental. Within psychology, the Empathy-Altruism Hypothesis (Batson, 2017) proposes that empathic concern can generate a genuinely altruistic motivation to help others, even in the absence of kinship or expectations of reciprocity. This framework highlights how affective processes, such as empathy, contribute to prosocial motivation and complement evolutionary accounts of social preferences. Nevertheless, the concept of altruism has been variously defined across disciplines, with ongoing debate about whether it denotes genuinely selfless concern or merely disguised self-interest (Scott & Seglow, 2007).

A key insight from evolutionary theory is that altruistic tendencies are not uniformly applied across social contexts. Individuals are more likely to help kin, close friends, or familiar others, reflecting mechanisms of inclusive fitness and reciprocal trust (Hamilton, 1963, 1964; Trivers, 1971, 2006). Burnstein, Crandall, and Kitayama (1994) demonstrated that the likelihood of helping increases with genetic relatedness—e.g., helping a sibling over a cousin. Moreover, altruism is not only determined by the degree of closeness to the target, but also by the costs associated with the altruistic act for the actor. Altruism inherently involves a personal cost to the actor, distinguishing it from purely self-interested behaviour. This costliness is a central concept in evolutionary and economic theories explaining altruistic behaviour (Fehr & Fischbacher, 2003). Altruism ranges from low-cost, easily performed actions to high-cost behaviours involving significant personal sacrifice, with individuals' willingness to help, depending on how acceptable these costs are within the given context. Moreover, altruism can sometimes lead to negative outcomes for the actor, especially when personal costs are very high or when helping becomes compulsive (Oakley et al., 2012).

Taken together, empathy and altruism have both been proposed as *commitment devices* that help individuals overcome selfish impulses and promote cooperative behaviour in social interactions (e.g., Bowles & Gintis, 2011; Fehr & Fischbacher, 2003; Frank, 1988; Tomasello, 2019). As commitment devices, these social preferences create internal motivations that reduce the likelihood of selfish

---

<sup>4</sup> A key concept in evolutionary theory is that of adaptations (Williams, 1966). Adaptations refer to evolutionary processes that consistently shape an organism's traits and characteristics, enabling it to effectively respond to recurring challenges related to survival and reproduction, whether directly or indirectly (Hamilton, 1964; Tooby & Cosmides, 1990). The main purpose of adaptations is to solve recurrent problems in ways that improve an organism's reproductive success (Buss, Haselton, Shackelford, Bleske & Wakefield, 1998).

choices, even when immediate personal costs are involved. In this context, empathy—especially affective empathy or empathic concern—represents an emotional resonance with another’s experience, while cognitive empathy involves the more deliberate capacity to understand another’s perspective without necessarily sharing their feelings. Altruism, by contrast, is characterized by a motivation to benefit others that includes a willingness to incur personal costs, distinguishing it from empathy, which does not inherently require costly action (Batson, 2011). The present study explores how both evolved social preferences, differentiated by their affective and cognitive components as well as the costliness of helping, influence decision-making in moral commitment problems, particularly the extent to which they inhibit selfish choices in a Prisoner’s Dilemma.

### 3.3 Individual and sex differences

Empathy and altruism are widely regarded as near-universal human capacities. Yet substantial individual differences exist, with sex emerging as one notable source of variation (Mitchell, 2018). Research generally finds that women, on average, tend to score slightly higher on empathic concern and altruistic motivation, while men often exhibit higher levels of sensation-seeking and risk-taking (Löffler & Greitemeyer, 2021). Moreover, men display greater variability in social preference traits, meaning they are more frequently found at both extremes—either highly cooperative or strongly self-interested—whereas women’s scores tend to cluster more closely around the group average (Boehm, 2012; Jolliffe & Farrington, 2021). Importantly, these sex differences are typically small in magnitude, with substantial overlap between men and women.

Evolutionary theory provides valuable insights into these patterns. Betzig (2012) presents comparative evidence indicating that males generally have higher variance in reproductive success than females, reflecting different selective pressures. Contemporary research suggests that sex differences in mating strategies and social behaviour reflect adaptive responses to differing reproductive and caregiving roles (Hrdy, 2009; Sommer, 2020). For instance, men on average tend to score somewhat higher on traits such as risk-taking and status-seeking, whereas women tend to score higher on empathy and caregiving (Daly & Wilson, 2001; Durrant, 2019).

This evolutionary tension is well captured by Sapolsky’s (2017) distinction between monogamous pair-bonded and promiscuous tournament species. In pair-bonded species, males invest in parenting and long-term partnerships, and females select mates based on caregiving traits. In tournament species, males compete intensely for mates with minimal parental involvement, while females favour dominance and genetic fitness. Humans appear to combine features of both systems, reflecting a dual evolutionary heritage. This duality is echoed in behavioural sex differences: some men emphasise nurturance and cooperation, while others prioritize competition and status, whereas women tend to show more

consistent empathy and prosociality, likely shaped by evolutionary pressures favouring caregiving and social cohesion (Campbell, 2002).

Such evolutionary and psychological trade-offs predict small sex differences in traits like impulsivity, status-seeking, and altruism, which may influence social decision-making in cooperative or competitive contexts. A compelling example of male variability comes from longitudinal research on antisocial behaviour. Moffitt et al. (2004), using the Dunedin Longitudinal Study, documented robust sex differences in prevalence, onset, and persistence of antisocial conduct. Males were more likely to follow a "life-course persistent" trajectory characterized by early and enduring aggression and rule-breaking, whereas females more often exhibited "adolescence-limited" antisocial behaviours, less severe and more situationally driven.

Together, these data support the conclusion that men show greater variability in social preferences and behaviours, from highly cooperative to intensely competitive or antisocial. This male variance in traits such as impulsivity and empathy may underlie their overrepresentation at both extremes of the moral spectrum, encompassing heroic altruists as well as chronic offenders. Humans exhibit traits associated with both monogamous and promiscuous mating systems, leading to intermediate sexual dimorphism and a wide range of behavioural expressions. Rather than strict sexual dimorphism, human traits appear flexible and shaped by a complex interplay of biological predispositions and sociocultural influences (Cliquet & Avramov, 2018).

#### **4 THE PRESENT STUDY**

The present study draws on evolutionary and psychological accounts of empathy and altruism as evolved social preferences that can function as commitment devices, promoting cooperative choices even when it comes at a personal cost. These preferences are assumed to vary across individuals and may help explain individual differences in cooperative choices even in anonymous or one-shot contexts where selfish choices are tempting. The study investigates how such social preferences—along with personality variables like impulsivity, risk-taking, and self-regulation—relate to behavioural choices in an anonymous Prisoner's Dilemma. Our study is structured around three sets of research questions.

##### **4.1 Descriptive RQ1**

RQ1a: What is the distribution of participants' choices in an anonymous Prisoner's Dilemma?;

RQ1b: To what extent do sex differences exist in participants' choices and the central constructs (i.e., personality traits and evolved social preferences)?

In line with descriptive RQ1, we first explore how participants' choices are distributed and whether sex differences emerge across choices, underlying traits, and social preferences. Building on prior research suggesting that females tend to score slightly higher on empathy and altruism, and that males show greater

variability (e.g., Betzig, 2012; Campbell, 2002; Hrdy, 2009; Moffitt et al., 2004), we expect that women will, on average, report higher levels of prosocial traits and make fewer selfish choices, while male responses may show greater dispersion.

## 4.2 Exploratory RQ2

RQ2a: How do demographic variables (sex and age) and control variables influence participants' choices? (Model 1);

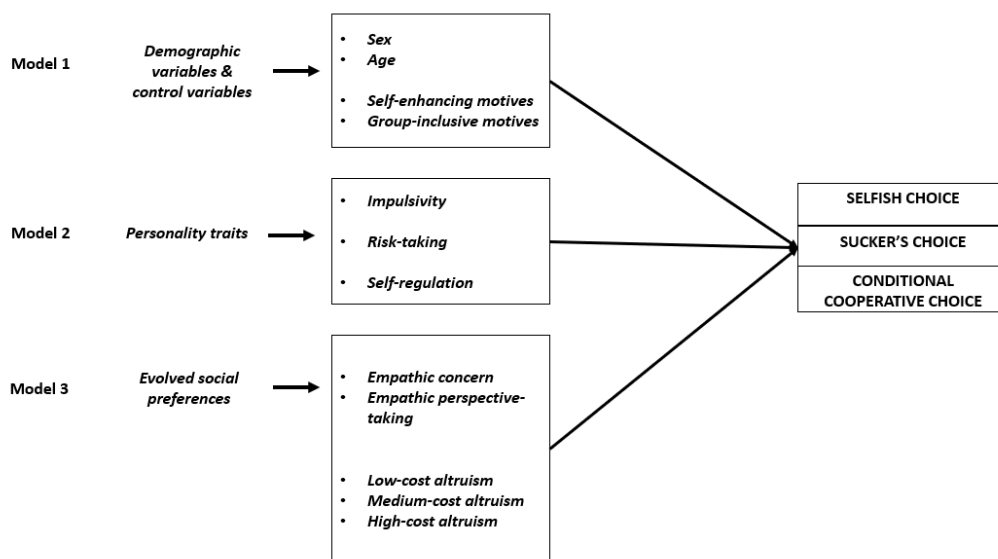
RQ2b: How do personality traits (impulsivity, risk-taking, and self-regulation) influence participants' choices, controlling for demographic variables? (Model 1 + Model 2);

RQ2c: How do evolved social preferences (empathic concern, empathic perspective-taking, altruism by costliness) influence participants' choices, controlling for demographics and personality traits? (Full Model)

These questions are guided by a hierarchical conceptual model (Figure 1) that sequentially incorporates demographics, personality traits, and evolved social preferences as predictors of PD choices.

**Figure 1**

*Hierarchical conceptual model in a one-shot Prisoner's Dilemma: The role of personality traits, and social preferences*



*Note.* Figure 1 illustrates the successive models of determinants used to examine participants' choices in the one-shot Prisoner's Dilemma. Demographic variables are entered first, followed by personality traits, and finally, evolved social preferences. Self-enhancing and group-inclusive motives (Declerck & Boone, 2016) are included as control variables to account for alternative decision rules influencing cooperative choices.

Following exploratory RQ2, we expect that individual differences in empathic concern, perspective-taking, and altruism will be negatively associated with selfish choices in the Prisoner's Dilemma, even when controlling for demographic and personality traits. Such a pattern would support the idea that

evolved social preferences sustain cooperative choices by constraining short-term self-interest in social decision-making.

### **4.3 Additional explanatory research question (RQ3)**

As an extension of the regression analyses, an additional explanatory research question (RQ3) is addressed through a structural equation model (SEM). This model examines the direct associations between affective (empathic concern) and cognitive (perspective-taking) components of empathy, as well as low-, medium- and high-cost forms of altruism, and participants' choices in the Prisoner's Dilemma. The SEM includes only the significant predictors identified in prior models to clarify the most relevant direct pathways from evolved social preferences to decision-making. By restricting the SEM to these significant predictors, we increase model parsimony and focus on the most relevant pathways influencing cooperative decision-making. Furthermore, the model tests for potential sex differences in these relationships by estimating separate paths for male and female participants.

RQ3: How do different forms of altruism (low-, medium-, and high-cost) and the dimensions of empathy (empathic concern and perspective-taking) relate directly to participants' choices in the Prisoner's Dilemma, and do these relationships differ by sex?

We do not have strong *a priori* hypotheses about sex differences in these direct effects; therefore, this analysis serves both as an explanatory test of the underlying mechanisms and an exploratory assessment of possible sex-specific patterns.

### ***Contributions of the present study***

This study makes several important contributions to the literature on commitment problems and evolutionary criminology (e.g., Durrant & Ward, 2015; Posick, Rocque & Barnes, 2021). First, it explicitly distinguishes between two dimensions of empathy: empathic concern (affective empathy) and empathic perspective-taking (cognitive empathy). These components are theorized to influence prosocial behaviour in distinct ways and are therefore modelled separately.

Second, this study advances the conceptualization of altruism beyond a single, uniform trait by framing it as a context-dependent tendency that varies with the costliness of helping behaviour. To capture this nuance, altruism is operationalized across three distinct levels—low-cost, medium-cost, and high-cost—each reflecting progressively greater degrees of personal sacrifice and investment. This approach aligns with theoretical perspectives, emphasising that altruistic actions are shaped not only by relatively stable individual dispositions (such as empathic concern) but also by situational demands and perceived costs, thereby offering a more fine-grained understanding of how evolved social preferences translate into cooperative decision-making.

## 5 METHODOLOGY

### 5.1 Data collection and participants

In the present study, the Prisoner's Dilemma (PD) was included in an internet-based survey. It was presented as a written vignette and embedded in a larger study on antisocial decision-making. After reading the vignette, participants were asked what they would do in the situation described (see Section 5.2). Importantly, the game was not played in a real experimental setting. Instead, it was framed as a hypothetical, one-shot, anonymous interaction. In such games, participants make decisions in a single, non-repeated encounter without knowing the identity of their counterpart. This setup eliminates reputational and strategic considerations and allows for the measurement of intrinsic social preferences (Camerer, 2004).

Data were collected from a large convenience sample of undergraduate students at Ghent University, one of the five largest universities in the Flemish region of Belgium. Between September and October 2019, one of the authors attended lectures to promote an online survey, distributing leaflets containing information about the study and a link to the survey. Additionally, the survey was promoted via a dedicated Facebook page shared with various student associations across faculties.

Participants provided informed consent before starting the survey, and full anonymity was ensured. The survey took approximately 20–25 minutes to complete, and participants could enter a raffle to win a €25 coupon upon completion. Demographic information, including sex and age, was collected at the end of the survey. The sample comprised 1,496 participants (70.9% female, 29.1% male) with a mean age of 19.9 years ( $SD = 3.20$ ). There was no missing data for these variables.

### 5.2 Dependent variable

The dependent variable in this study is participants' self-reported choice in a one-shot Prisoner's Dilemma (PD) scenario adapted from Nelissen (2013). The vignette describes a hypothetical situation in which two students face a joint liability problem after a theft, with payoffs structured to reflect the classic PD logic: the temptation to defect (pay €0) dominates mutual cooperation (pay €8), which in turn is preferred over mutual defection (pay €18), while the sucker's payoff (pay €30) is least desirable. In this adapted design, mutual admission (€8 each) represents the cooperative outcome and yields the best collective payoff, whereas mutual denial (€18 each) reflects the worst collective outcome.

The scenario, originally developed by Mark Nelissen (2013: p. 130), professor emeritus of behavioural biology, was presented as follows. Imagine the following situation:

Albert, a shopkeeper, employs two young students, Bernard and Siska, as weekend assistants. After closing time, he calls them into his office because he has discovered that an expensive bottle of vodka has gone missing from the shelf. The

bottle was neither scanned nor paid for. Albert did not take it himself, so the only remaining possibility is that one of the two youngsters stole the alcohol—or more likely both of them, since he saw them coming into the shop that morning with bleary eyes and clearly suffering from a hangover. They probably had quite a party. The two students are now standing in front of Albert's desk. Albert tells them that a €30 bottle of vodka has gone missing and that he cannot simply let it slide. He offers Bernard and Siska the following choice:

Look, here's my proposal: if you partied together, you pay together. So, you each reimburse me €8, and I'll cover the rest myself—after all, it's the first time something like this has happened. It is, of course, possible that only one of you stole the bottle. In that case, the culprit pays the full €30, and the other pays nothing. That seems fair to me. But it is also possible that you both claim to be innocent and point at each other as the culprit. In that case, you share the responsibility, and each pays €18. The extra €3 can be considered a penalty for mutual accusations.

Participants were asked to imagine themselves in Bernard's or Siska's position and choose one of three options: (1) blame the other student and deny involvement, stating that they did not steal the bottle. This option, referred to as the Selfish Choice, operationalizes cheating within the Prisoner's Dilemma framework: it prioritizes self-interest by denying personal responsibility and placing perceived responsibility on the other student to avoid personal cost, reflecting the defection strategy inherent in the PD. (2) Take full responsibility, accepting the consequences themselves. This Sucker's Choice reflects unconditional cooperation, in which the participant accepts the worst individual payoff. (3) Share responsibility only if the other person does the same. This Conditional Cooperative Choice represents reciprocity-based cooperation, contingent on mutual trust, and conceptually aligns with the reward for mutual cooperation.

The PD scenario was designed to assess participants' behavioural intentions rather than direct actions. While the Selfish Choice corresponds to cheating, the Sucker's Choice reflects unconditional cooperation, and the Conditional Cooperative Choice represents contingent cooperation assuming reciprocity from the other player. Although this third option does not map directly onto a single cell in the payoff matrix, it conceptually aligns with the reward for mutual cooperation.

This adaptation does not represent a canonical Prisoner's Dilemma. Instead, it functions as a moral dilemma with PD-like features, capturing normative and emotional motivations rather than purely strategic reasoning. As Frank (1988) argues, emotions such as empathy serve as commitment devices that sustain cooperation even when short-term incentives favour defection. Including a conditional option allows participants to express reciprocity-based strategies rather than unconditional self-interest. This approach is consistent with findings from neuroeconomics, where trust-based cooperation is linked to group-inclusive

motives (Declerck & Boone, 2016). Conditional cooperation is widely discussed in behavioural economics and evolutionary theory as a mechanism for sustaining cooperation (e.g., Axelrod, 1984; Rand & Nowak, 2013). The inclusion of this conditional element adds a coordination-like aspect, departing from the strict payoff logic of a classic PD.

Participants were not financially compensated; the scenario was entirely hypothetical. Nonetheless, the three response options reflect distinct motivational orientations: self-interest (Selfish Choice), altruistic sacrifice (Sucker's Choice), and trust-based reciprocity (Conditional Cooperative Choice). Table 2 presents the 2×2 payoff matrix for the one-shot PD involving two players, Bernard and Siska, showing the hypothetical monetary outcomes, with the Selfish Choice corresponding to the temptation to cheat, the Sucker's Choice corresponding to the sucker's payoff, and the Conditional Cooperative Choice corresponding to the reward for cooperation.

The adapted scenario operates on two levels: firstly, classic PD tension: Denying minimizes personal cost, whereas mutual admission minimizes joint cost; secondly, normative/moral tension: participants must weigh honesty and trust against self-interest. Sucker's Choice exemplifies moral risk: taking full responsibility exposes the individual to exploitation without assurance of reciprocity or factual guilt. This reflects the double-edged nature of evolved social preferences such as empathy—they can promote helping behaviour while simultaneously creating vulnerability to exploitation.

**Table 2**

*Defining the one-shot Prisoner's dilemma (PD): Payoff matrix (adapted from Nelissen, 2013)*

	Siska admits (cooperates)	Siska denies (cheats)
Bernard admits (cooperates)	<b>Reward for cooperation</b> Both admit to the theft <i>Each pays €8</i>	<b>Sucker's payoff</b> Bernard pays €30 Siska pays €0
Bernard denies (cheats)	<b>Temptation to cheat</b> Bernard pays €0 Siska pays €30	<b>Punishment for mutual cheating</b> Both deny & point to each other as the culprit Each pays €18

*Note.* The payoff matrix reflects the scenario described in the vignette with: Reward for mutual cooperation: Both admit to the theft → each pays €8. Temptation to cheat: Deny while the other admits → pay €0. Sucker's payoff: Admit while the other denies → pay €30. Punishment for mutual cheating: Both plead innocent and point to each other as the culprit → each pays €18 (including a €3 penalty for mutual cheating). This structure preserves the classic PD logic while introducing a normative dimension related to honesty and trust.

This measure captures participants' behavioural intentions in a social dilemma that reflects the commitment problem: while mutual cooperation yields the best collective outcome, individual incentives favour cheating. In the context of this study, cheating is operationalized as the Selfish Choice, accusing the other to avoid personal cost. The hypothetical, one-shot scenario isolates intrinsic social

preferences by minimizing external influences such as accountability or long-term consequences, enhancing internal validity while limiting ecological realism (Camerer, 2004).

### 5.3 Independent variables

#### *Personality traits*

Impulsivity and risk-taking—both linked to rule-breaking behaviours such as crime and delinquency (e.g., Armstrong et al., 2020; Burt et al., 2014; Forrest et al., 2019; Vazsonyi & Ksinan, 2017)—were included as personality traits. Impulsivity was assessed with three items on acting without foresight (e.g., “I do things without thinking about them first”), and risk-taking with four items from the UPPS-P scale (Whiteside & Lynam, 2001), capturing novelty and thrill seeking (e.g., “I quite enjoy taking risks”). Higher mean scores indicate greater impulsivity and sensation seeking. Both scales showed relatively good reliability (impulsivity: Cronbach’s  $\alpha = .79$  and risk-taking: Cronbach’s  $\alpha = .70$ ).

Self-regulation was assessed using a set of items adapted from the Capacity for Self-Control scale developed by Hoyle and Davison (2016). This scale itself builds on the Low Self-Control Scale derived from the General Theory of Crime (Grasmick et al., 1993) and the Self-Control Scale by Tangney et al. (2004). For this study, only items targeting the *inhibition* aspect of self-regulation were selected. This facet reflects the ability to manage impulses, urges, or situational pressures that clash with one’s current goals—for instance, the goal to uphold moral principles, such as honesty, when tempted to act otherwise. The chosen items specifically assess the tendency to suppress temptations, cravings, impulses, wants, and harmful habits. Participants indicated their agreement with eight statements on a five-point Likert scale (1 = *strongly agree* to 5 = *strongly disagree*). Example items include: “I am able to resist temptations” and “I have trouble resisting my cravings.” All items were reverse-coded so that higher scores reflect stronger self-regulation. The internal consistency of this measure in the present study was relatively good ( $\alpha = .74$ ).

#### *Evolved social preferences*

Empathic concern and perspective taking were assessed using the seven-item Empathic Concern (EC) and Perspective Taking (PT) subscales of the Interpersonal Reactivity Index (IRI; Davis, 1980, 1983, 1994). The PT subscale assesses the tendency to adopt others’ viewpoints in everyday situations, while the EC subscale measures other-oriented feelings of compassion and concern for people in distress. Participants rated their agreement on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Example items for the EC scale include “I often have tender, concerned feelings for people less fortunate than me” and “When I see someone being treated unfairly, I sometimes don’t feel very much pity for them” (reverse-scored). Example items for the PT scale include “I try to look at everybody’s side of a disagreement before I make a decision” and

“I sometimes find it difficult to see things from the other person’s point of view” (reverse-scored). Both constructs were modelled as latent factors using confirmatory factor analysis (CFA). The EC scale showed good internal consistency (Cronbach’s  $\alpha = .81$ ), and the PT scale demonstrated acceptable reliability (Cronbach’s  $\alpha = .71$ ; DeVellis, 2003; Nunnally & Bernstein, 1994; Tavakol & Dennick, 2011).

Altruism was measured using an adapted, scenario-based version of the Self-Report Altruism Scale (SRA; Rushton, Chrisjohn, & Fekken, 1981). Instead of reporting past behaviour frequency, participants indicated the likelihood of engaging in 20 specified altruistic actions across hypothetical situations, using a 5-point Likert scale ranging from 1 (*very unlikely*) to 5 (*very likely*). This adaptation maintains the original scale’s validity while allowing prospective assessment of altruistic tendencies. Exploratory and confirmatory factor analyses identified a three-factor structure corresponding to altruistic behaviours of varying costliness: low-cost, medium-cost, and high-cost altruistic acts. Items with low factor loadings were excluded to ensure scale reliability and validity. The three factors showed significant positive correlations: low-cost altruism (6 items, Cronbach’s  $\alpha = .69$ ) correlated moderately to strongly with medium-cost altruism (3 items;  $r = .566, p < .001$ ) and strongly with high-cost altruism (4 items;  $r = .630, p < .001$ ). Medium-cost altruism also correlated moderately with high-cost altruism ( $r = .397, p < .001$ ). These findings indicate related but distinct dimensions of altruistic preference. Cronbach’s alphas are respectively .69, .51, and .71. Although the Cronbach’s alpha for the medium-cost altruism scale is relatively low ( $\alpha = .51$ ), it is important to note that Cronbach’s alpha is a conservative estimate of reliability (Sijtsma, 2009). Moreover, the factor loadings of all items were sufficiently high, indicating that the scale items adequately represent the underlying construct (For a detailed discussion, see Pauwels & De Buck, 2026). This suggests that despite the low alpha, the scale demonstrates acceptable construct validity. These three subscales were employed in further analyses.<sup>5</sup>

### ***Control variables and demographic variables***

Self-enhancing and group-inclusive motives, adapted from Declerck and Boone (2016), represent two decision rules underlying cooperative behaviour; These

---

<sup>5</sup> The decision to retain three cost-based factors (low-, medium-, and high-cost altruism) was guided by theoretical and empirical considerations. The distinction between cost levels is well established in the literature on costly helping (e.g., Fehr & Fischbacher, 2003; Rushton et al., 1981). In a separate validation study (Pauwels & De Buck, 2026) exploratory and confirmatory factor analyses supported a three-factor model with acceptable fit indices (CFI = .932; TLI = .914; RMSEA = .066). All retained items showed factor loadings  $> .40$ , indicating adequate construct representation. While the medium-cost factor’s reliability is modest ( $\alpha = .51$ ), this is partly due to its brevity (3 items), and Cronbach’s alpha is a conservative estimator that tends to underestimate reliability in short scales (Sijtsma, 2009). Inter-factor correlations were moderate to strong ( $r = .40$  to  $.63$ ), but not high enough to indicate redundancy. Collapsing the subscales would therefore obscure theoretically meaningful distinctions and differential predictive patterns observed in our analyses.

concepts originate from the field of neuroeconomics, which integrates neuroscience, economics, and psychology to study how people make decisions. Self-enhancing motives reflect incentive-based cooperation driven by self-interest and evolutionary fitness benefits, promoting self-serving behaviour in the absence of external incentives (e.g., punishment). This construct was measured by agreement with the statement: “I prioritize my interests unless I have specific reasons to consider those of others as well” rated on a 5-point Likert scale. Group-inclusive motives capture trust-based cooperation rooted in group-mindedness and an intrinsic willingness to cooperate with one’s group, conditional on trust signals to avoid exploitation by selfish individuals. This was measured by agreement with: “I prioritize the interests of others unless I believe they are unreliable” using the same scale. These motives were included as control variables to account for alternative decision rules influencing cooperative behaviour beyond the core evolved social preferences.

Biological sex (0 = female, 1 = male) and age (in years) were included as demographic controls.

#### **5.4 Analysis plan**

A combination of analyses was conducted to address the three research questions: descriptive patterns (RQ1), predictive relationships (RQ2), and explanatory mechanisms (RQ3).

##### ***Descriptive patterns (RQ1)***

To address the first research question (RQ1), descriptive statistics were computed to examine the distribution of participants’ responses in the one-shot Prisoner’s Dilemma (PD). In addition, one-way ANOVAs were conducted to explore sex differences in PD choices and the central psychological constructs. Where appropriate, effect sizes ( $\eta^2$ ) were reported to quantify the magnitude of group differences.

##### ***Exploratory relationships - Hierarchical multinomial logistic regression (RQ2)***

To investigate the predictors of participants’ choices in the PD (RQ2), a hierarchical multinomial logistic regression analysis was conducted using R (version 4.5.2) with the nnet package. This analytic approach is appropriate for a nominal dependent variable with three unordered categories. Conditional Cooperative Choice, the most frequently selected response ( $n = 1,259$ ; 84.2%), was set as the reference category in all comparisons. The regression models were specified hierarchically: Model 1 included demographic variables : sex (coded 0 = female, 1 = male) and age (continuous), as well as decision heuristics, specifically self-enhancing and group-inclusive motives. Model 2 added personality traits, namely impulsivity, risk taking, and self-regulation. Model 3 further incorporated social preference variables, including empathic concern, empathic perspective-taking, and low-, medium-, and high-cost altruism.

Multinomial logistic regression estimates the probability of membership in each outcome category relative to a reference group. For each comparison, unstandardized coefficients (B) and odds ratios (Exp(B)) are reported. Odds ratios greater than 1 indicate increased odds of selecting the comparison category, whereas values below 1 indicate decreased odds (Field, 2024; Hosmer, Lemeshow, & Sturdivant, 2013).

To enhance substantive interpretability, we additionally report Average Marginal Effects (AMEs) for the significant predictors in Model 3. AMEs represent the average change in the probability of selecting a specific choice (Selfish, Sucker, or Conditional Cooperation) associated with a one-unit increase in a predictor, or a shift to the next category for categorical variables, in percentage points (Greene, 2018; Williams, 2012). For z-standardized predictors, a one-unit increase corresponds to a one standard deviation change. Positive AME values indicate an increased likelihood of the outcome, whereas negative values indicate a decreased likelihood. This approach provides a more intuitive measure of effect size in terms of changes in probability.

### ***Explanatory mechanisms - Structural Equation Modeling (RQ3)***

To address the explanatory research question (RQ3), a supplementary Structural Equation Model (SEM) was conducted using Mplus (Version 8.11). This SEM model included only those variables that emerged as statistically significant predictors in the preceding regression analyses. The model tested whether altruism (at low, medium, and high-cost levels) mediates the relationship between empathic concern, empathic perspective-taking and participants' choices in the PD. Given the nominal nature of the outcome variable (PD choice), a multinomial logistic regression framework was applied within the SEM. Robust maximum likelihood estimation (MLR) with Monte Carlo integration was used to account for potential deviations from multivariate normality (Kline, 2023; Muthén & Muthén, 2017). Model fit was evaluated using standard indices (e.g., CFI, RMSEA, SRMR). Standardized path coefficients were reported to facilitate interpretation and assess the strength and direction of direct effects.

## **6 RESULTS**

### **6.1 Descriptive results (RQ1)**

A total of 1,496 participants chose a one-shot Prisoner's Dilemma scenario. The majority ( $n = 1259$ ; 84.2%) selected the conditional cooperative option, indicating trust-based cooperation, that is, a willingness to cooperate based on the expectation that the other party would do the same. A smaller proportion ( $n = 179$ ; 12.0%) chose the self-interested option, opting to cheat regardless of the other party's behaviour. A minority ( $n = 58$ ; 3.9%) selected the Sucker's Choice, indicating a willingness to cooperate unconditionally, even if the other party might cheat, thereby accepting full responsibility for the outcome.

**Table 3**

*Frequencies and percentages of participants' choices in the one-shot Prisoner's Dilemma (n = 1496)*

PD Choice	Females (n = 1061)	Males (n = 435)	Total (n = 1496)
Selfish Choice	120 (11.3%)	59 (13.6%)	179 (12.0%)
Sucker's Choice	32 (2.9%)	27 (6.2%)	58 (3.9%)
Conditional Cooperative Choice	910 (85.8%)	349 (80.2%)	1259 (84.2%)
<b>Total</b>	1062 (100%)	435 (100%)	1496 (100%)

A chi-square test of independence showed a statistically significant association between sex and participants' choices in the one-shot Prisoner's Dilemma,  $\chi^2(2, n = 1,496) = 11.02, p = .004$ . Men were slightly more likely than women to choose both the Selfish Choice and the Sucker's Choice, while women more frequently opted for the conditional cooperative strategy. The effect size was small, Cramér's  $V = .086, p = .004$ .

One-way ANOVAs were conducted to examine sex differences in the central psychological constructs. Results indicated statistically significant differences in several variables (Table 4). Women scored higher than men on empathic concern, empathic perspective-taking, low-cost and high-cost altruism, and group-oriented thinking. In contrast, men scored higher on self-enhancing motives and risk-taking. No meaningful sex differences were found for impulsivity, medium-cost altruism, or self-regulation. To assess the practical relevance of these findings, effect sizes were calculated using eta squared ( $\eta^2$ ).

Based on Cohen's (1988) guidelines—small ( $\eta^2 \approx .01$ ), medium ( $\eta^2 \approx .06$ ), and large ( $\eta^2 \geq .14$ )—the following effects were observed: large effects for empathic concern ( $\eta^2 = .14$ ) and high-cost altruism ( $\eta^2 = .13$ ); medium effects for selfish tendencies ( $\eta^2 = .05$ ) and sensation seeking ( $\eta^2 = .03$ ); and small but significant effects for empathic perspective-taking ( $\eta^2 = .02$ ) and group-oriented thinking ( $\eta^2 = .02$ ).

Additionally, score variances were consistently higher among men than women, indicating a broader distribution of responses. This suggests that men were more frequently represented at both extremes of the scale—for example, displaying either very high or very low levels of empathy—whereas women's scores were more tightly clustered around the mean. This pattern aligns with our expectations.

## **6.2 Exploratory results: Hierarchical multinomial logistic regression (RQ2)**

### *Check for multicollinearity*

Before conducting the multinomial logistic regression to address RQ2, multicollinearity among the independent variables was assessed by inspecting the zero-order Pearson correlations (Table 5). No correlation exceeded the threshold of  $|.80|$ . To further assess multicollinearity, linear regression was run with all predictors entered simultaneously. Variance Inflation Factor (VIF) values ranged

**Table 4***Summary descriptive statistics, and internal consistency of the core constructs*

Exogenous variables	Cronbach $\alpha$	Min-Max	Full sample ( $n = 1496$ )		Males ( $n = 435$ )		Females ( $n = 1061$ )		ANOVA (Sex)	
			Mean	SD	Mean	SD	Mean	SD	$F(1, 1494)$	$\eta^2$ (x 100)
Empathic concern	.81	1-5	3.74	0.65	3.36	0.66	3.90	0.58	<b>243.19***</b>	14%
Empathic persp. taking	.71	1-5	3.60	0.52	3.48	0.53	3.64	0.51	<b>32.63***</b>	2%
Low-cost altruism	.69	1-5	4.28	0.49	4.17	0.52	4.33	0.46	<b>34.48***</b>	2%
Medium-cost altruism	.51	1-5	2.96	0.74	2.98	0.78	2.95	0.73	0.81 (ns)	0.1%
High-cost altruism	.71	1-5	3.71	0.72	3.30	0.79	3.87	0.62	<b>219.81***</b>	12.8%
Self-enhancing motives	--	1-5	2.59	1.11	2.98	1.14	2.43	1.05	<b>80.98***</b>	5.1%
Group-inclusive motives	--	1-5	3.49	0.90	3.31	0.90	3.57	0.88	<b>26.92***</b>	1.8%
Impulsivity	.79	1-5	2.76	0.85	2.77	0.93	2.75	0.85	0.23 (ns)	0%
Risk-taking	.70	1-5	3.19	0.85	3.40	0.84	3.10	0.83	<b>38.79***</b>	2.5%
Self-regulation	.74	1-5	3.29	0.59	3.26	0.61	3.30	0.59	1.27 (ns)	0.1%

\*\*\* $p < 0.001$  \*\* $p < 0.01$  ns= not significant**Table 5***Zero-order Pearson correlations between the central constructs*

Exogenous variables	1	2	3	4	5	6	7	8	9
1. Empathic concern	1								
2. Empathic persp. taking	<b>.368***</b>	1							
3. Low-cost altruism	<b>.352***</b>	<b>.276***</b>	1						
4. Medium-cost altruism	<b>.170***</b>	<b>.162***</b>	<b>.331***</b>	1					
5. High-cost altruism	<b>.461***</b>	<b>.246***</b>	<b>.446***</b>	<b>.227***</b>	1				
6. Self-enhancing motives	<b>-.395***</b>	<b>-.236***</b>	<b>-.231***</b>	<b>-.094***</b>	<b>-.298***</b>	1			
7. Group-inclusive motives	<b>.373***</b>	<b>.240***</b>	<b>.194***</b>	<b>.107***</b>	<b>.263***</b>	<b>-.365***</b>	1		
8. Impulsivity	-.014 (ns)	<b>-.161***</b>	-.018 (ns)	<b>.151***</b>	.011 (ns)	<b>.096***</b>	-.048 (ns)	1	
9. Risk-taking	<b>-.106***</b>	-.003 (ns)	<b>.063*</b>	<b>.160***</b>	-.019 (ns)	<b>.069**</b>	-.001 (ns)	<b>.279***</b>	1
10. Self-regulation	.038 (ns)	<b>.177***</b>	<b>.099***</b>	<b>-.110***</b>	.051 (ns)	<b>-.077**</b>	.050 (ns)	<b>-.449***</b>	<b>-.123***</b>

\*\*\* $p < 0.001$  \*\* $p < 0.01$  ns= not significant

from 1.135 to 1.615, and tolerance values ranged from .619 to .881. These values fall well within accepted thresholds ( $VIF < 5.0$ ; tolerance  $> .20$ ), indicating no evidence of problematic multicollinearity (Field, 2024). Therefore, all predictor variables were retained for inclusion in the multinomial logistic regression models.

### ***Model 1 - Demographics and control variables (RQ2a)***

Table 6 presents the results of Model 1, examining the effects of sex, age, and two heuristic motives on participants' choices. Self-enhancing motives significantly increased the likelihood of selecting the Selfish Choice relative to Conditional Cooperation ( $OR = 1.31, p = .002$ ) and significantly decreased the likelihood of selecting the Sucker's Choice ( $OR = 0.63, p = .005$ ). Biological sex was a significant predictor for the Sucker's Choice, with men more likely than women to choose this option ( $OR = 2.806, p < .001$ ). Age and group-inclusive motives were not significant predictors in either comparison.

### ***Model 2 – Demographics, control variables + personality traits (RQ2b)***

When including demographics, control variables, and personality traits in the multinomial logistic regression, self-enhancing motives remain a significant predictor of participants' choices. Specifically, higher self-enhancing motives increased the likelihood of making a Selfish Choice compared to the Conditional Cooperative Choice ( $B = 0.29, p = .001, \text{Exp}(B) = 1.33$ ). For the Sucker's Choice, both self-enhancing motives ( $B = -0.46, p = .005, \text{Exp}(B) = 0.63$ ) and biological sex ( $B = 1.097, SE = 0.28, p < .001, \text{Exp}(B) = 2.994$ ) were significant predictors, indicating that individuals with stronger self-enhancing motives and males were more likely to choose the sucker option relative to conditional cooperation.

### ***Model 3 – Demographics, control variables, personality traits + evolved social preferences (RQ2c)***

For the selfish choice, only self-enhancing motives ( $B = 0.192, p = .037$ ), self-regulation ( $B = 0.189, p = .045$ ), and medium-cost altruism ( $B = -0.249, p = .006$ ) significantly predicted participants' likelihood of choosing this option. Specifically, higher self-enhancing motives and stronger self-regulation increased the odds of a Selfish Choice, while higher medium-cost altruism decreased it.

Regarding the Sucker's Choice, significant predictors included sex (male;  $B = 1.044, p < .001$ ), self-enhancing motives ( $B = -0.442, p = .011$ ), empathic concern ( $B = 0.470, p = .014$ ), medium-cost altruism ( $B = 0.357, p = .018$ ), and high-cost altruism ( $B = -0.344, p = .039$ ). Compared to men, women were less likely to make the Sucker's Choice, while higher empathic concern and medium-cost altruism increased its likelihood. In contrast, higher self-enhancing motives and high-cost altruism decreased the odds of choosing the sucker's option.

**Table 6**

*Hierarchical multinomial logistic regression predicting participants' choices in a one-shot Prisoner's Dilemma (Reference category: Conditional Cooperative Choice)*

Variables z-standardized	B (Selfish Choice)	Exp(B) (Selfish Choice)	B (Sucker Choice)	Exp(B) (Sucker Choice)
Model 1				
Sex <sup>a</sup>	.070 (ns)	1.072	<b>.1.032***</b>	<b>2.806</b>
Age	.103 (ns)	1.108	-.311 (ns)	.733
Self-enhancing motives	<b>.272**</b>	<b>1.313</b>	<b>-.464**</b>	<b>.628</b>
Group-inclusive motives	-.084 (ns)	.920	.017 (ns)	1.017
Model 2				
Sex <sup>a</sup>	.092 (ns)	1.096	<b>1.097***</b>	<b>2.994***</b>
Age	.084 (ns)	1.087	-.353 (ns)	.703
Self-enhancing motives	<b>.286***</b>	<b>1.332***</b>	<b>-.464**</b>	<b>.629**</b>
Group-inclusive motives	-.082 (ns)	.921	.030 (ns)	1.030
Impulsivity	.028 (ns)	1.028	-.028 (ns)	.972
Risk-taking	-.051 (ns)	.951	-.206 (ns)	.814
Self-regulation	.178 (ns)	1.194	-.131 (ns)	.877
Model 3				
Sex <sup>a</sup>	-.113	.894	<b>1.044***</b>	<b>2.840***</b>
Age	.113	1.120	-.321	.725
Self-enhancing motives	<b>.192*</b>	<b>1.212*</b>	<b>-.442*</b>	<b>.643*</b>
Group-inclusive motives	.033	1.033	.007	1.007
Impulsivity	.050	1.051	-.087	.916
Risk-taking	-.013	.987	-.180	.835
Self-regulation	<b>.189*</b>	<b>1.208*</b>	-.068	.934
Empathic concern	-.161	.852	<b>.470*</b>	<b>1.601*</b>
Empathic perspective- taking	-.172	.842	-.123	.884
Low-cost altruism	.017	1.017	-.174	.841
Medium-cost altruism	<b>-.249**</b>	<b>.780**</b>	<b>.357*</b>	<b>1.429*</b>
High-cost altruism	-.111	.895	<b>-.344*</b>	<b>.709*</b>

\*\*\* $p < 0.001$  \*\* $p < 0.01$  \* $p < 0.05$  ns = not significant

<sup>a</sup> 0 = female, 1 = male

Table 7 presents the Average Marginal Effects (AMEs) for the significant predictors in Model 3 of the multinomial logistic regression (reference category: Conditional Cooperative Choice). AMEs express the average change in the probability of selecting a specific choice associated with a one-unit increase in the

predictor (or a shift to the next category for categorical variables), in percentage points.

For the Selfish Choice, higher self-enhancing motives and stronger self-regulation were associated with small increases in the likelihood of choosing this option (+2.10 p.p. and +1.94 p.p., respectively), whereas higher medium-cost altruism reduced this likelihood by -2.65 p.p.

For the Sucker’s Choice, being male increased the probability by +4.54 p.p., while higher empathic concern and medium-cost altruism also raised the likelihood (+1.74 p.p. and +1.37 p.p., respectively). In contrast, higher self-enhancing motives and high-cost altruism decreased the probability of Sucker’s Choice (-1.65 p.p. and -1.19 p.p., respectively). Changes in these predictors were accompanied by corresponding shifts in the probability of Conditional Cooperative Choice.

**Table 7**  
*Average marginal effects in percentage points (Model 3)*

Variables	Selfish Choice (Δ p.p.)	Sucker’s Choice (Δ p.p.)	Conditional Cooperative Choice (Δ p.p.)
Sex (0 to 1)	<b>-1.55</b>	<b>+4.54</b>	<b>-2.99</b>
Self-enhancing motives	<b>+2.10</b>	<b>-1.65</b>	<b>-0.45</b>
Self-regulation	<b>+1.94</b>	<b>-0.31</b>	<b>-1.63</b>
Empathic concern	<b>-1.80</b>	<b>+1.74</b>	<b>+0.05</b>
Medium-cost altruism	<b>-2.65</b>	<b>+1.37</b>	<b>+1.28</b>
High-cost altruism	<b>-1.00</b>	<b>-1.19</b>	<b>+2.19</b>

*Note.* AMEs are expressed in percentage points (p.p.). For z-standardized predictors, effects reflect a +1 SD increase; for Sex, the effect is the discrete change from 0 (female) to 1 (male). Positive values increase the probability of the outcome; negative values decrease it.

### 6.3 Explanatory results: Structural Equation Modeling (RQ3)

#### *Preliminary note on the specification of the model*

Only the significant predictors identified through the hierarchical multinomial logistic regression analyses were included in the subsequent structural equation modeling (SEM). This approach ensured that the SEM focused on the variables most strongly associated with the outcome, thereby enhancing model parsimony and interpretability. A confirmatory factor analysis (CFA) was conducted to evaluate the measurement model of the latent constructs in the study. The CFA model comprised empathic concern (seven indicators), medium-cost altruism (three indicators), high-cost altruism (four indicators), and self-regulation (seven indicators). This measurement model served as the foundation for the subsequent structural analyses. Model fit statistics are provided in Footnote 1.<sup>6</sup> Overall, the

<sup>6</sup> The results indicated that the model fit the data reasonably well. The chi-square test of model fit was significant ( $\chi^2 = 875.86$ ,  $df = 183$ ,  $p < .001$ ), which is common in large samples due to the test’s sensitivity to sample size. The Root Mean Square Error of Approximation (RMSEA) was

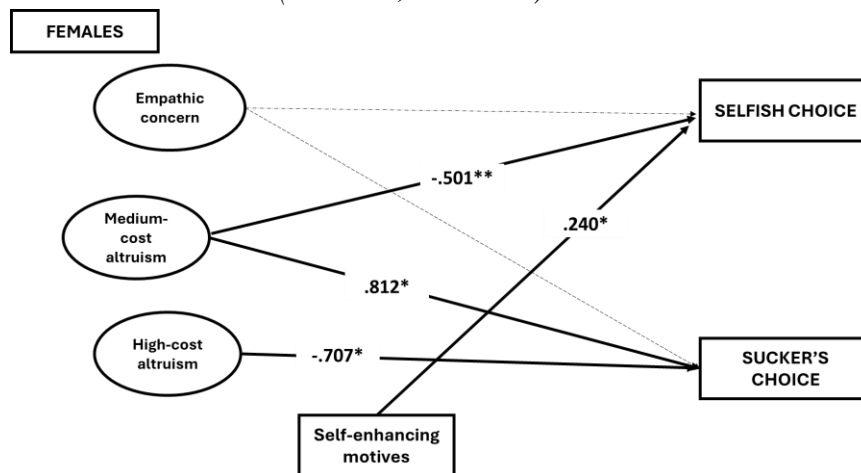
CFA confirmed the construct validity of the latent variables, indicating that the observed indicators effectively captured the underlying constructs. Both the CFA and SEM analyses were conducted using Mplus version 8.11 (Muthén & Muthén, 2017).

### *Main analysis*

In the baseline SEM, distinct dimensions of empathy and altruism emerged as significant albeit ambivalent predictors of participants' choices in the one-shot Prisoner's Dilemma. Specifically, medium-cost altruism reduced the likelihood of Selfish Choices but increased the likelihood of choosing the sucker option. Empathic concern (affective empathy) significantly predicted sucker choices but showed only a marginal effect on selfish choices. Interestingly, high-cost altruism was associated with a decreased likelihood of choosing the sucker option. Self-enhancing motives, included as a control variable, remained a robust and significant predictor of both non-cooperative choices, positively associated with Selfish Choices, and negatively with sucker choices. Sex differences were evident only for the Sucker's Choice, with males more likely to select this option. These findings support the decision to conduct subsequent analyses separately by sex, as they suggest sex-specific patterns.

**Figure 2**

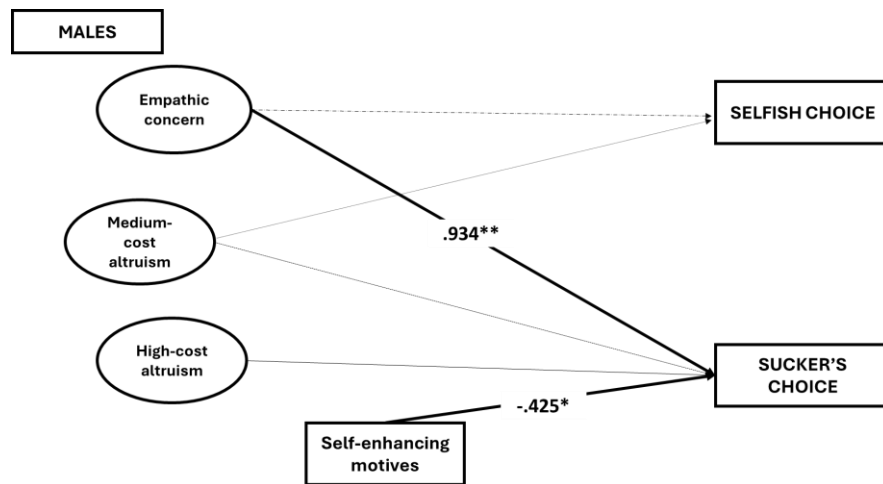
*Structural equation model predicting participants' choices in the one-shot Prisoner's Dilemma (Females; n = 1061)*



0.050, with a 90% confidence interval ranging from 0.047 to 0.054, and the probability that RMSEA was less than or equal to 0.05 was 0.433, suggesting acceptable fit (Hu & Bentler, 1999). Additionally, the Standardized Root Mean Square Residual (SRMR) was low at 0.043, further supporting the model's adequacy (Hu & Bentler, 1999). However, the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) were 0.890 and 0.873, respectively, which fall slightly below the conventional threshold of 0.90 generally recommended for good fit (Kline, 2023). These slightly lower values suggest that while the model reasonably represents the data, there is some room for improvement in model specification. Given the complexity of the constructs and the sample size, the measurement model was considered sufficiently robust to proceed with subsequent structural analyses, though caution is warranted in interpreting the results related to the latent constructs (Marsh, Hau, & Wen, 2004).

**Figure 3**

*Structural equation model predicting participants' choices in the one-shot Prisoner's Dilemma (Males; n = 435)*



*Note.* The models in Figure 2 and Figure 3 illustrate the standardized effects of empathic concern, medium-cost altruism, high-cost altruism, and self-enhancing motives on participants' likelihood of making a selfish or sucker choice, relative to a Conditional Cooperative Choice (reference category) in the Prisoner's Dilemma. Empathic concern and altruism at medium and high-cost levels are modelled as a latent construct. Only statistically significant paths are displayed. Arrows in bold visualize the significant effects.

Given the observed sex difference in sucker choices in the full model, separate structural models were estimated for males and females. The results revealed notable differences in the predictors of Prisoner's Dilemma decisions across sexes.

**Table 8**

*Results from the structural equation model predicting participants' choices in a one-shot Prisoner's Dilemma: Odds Ratios, 95% CI, and standardized coefficients (Reference category: Conditional Cooperative Choice)*

Exogenous variables	Outcome variables	Sex	Odds Ratio	95% CI lower	95% CI upper	Standardized coefficients
Empathic Concern	Selfish choice	Female	0.681	0.405	1.146	-0.212
	Sucker's choice	Female	1.637	0.511	5.251	0.273
Medium-cost altruism	Selfish choice	Male	0.754	0.424	1.340	-0.196
	<b>Sucker's choice</b>	<b>Male</b>	<b>3.833</b>	<b>1.536</b>	<b>9.562</b>	<b>0.934</b>
High-cost altruism	<b>Selfish choice</b>	<b>Female</b>	<b>0.382</b>	<b>0.199</b>	<b>0.734</b>	<b>-0.501</b>
	<b>Sucker's choice</b>	<b>Female</b>	<b>4.751</b>	<b>1.420</b>	<b>15.892</b>	<b>0.812</b>
	Selfish choice	Male	0.805	0.175	3.708	-0.086
	Sucker's choice	Male	1.183	0.243	5.746	0.066

Exogenous variables	Outcome variables	Sex	Odds Ratio	95% CI lower	95% CI upper	Standardized coefficients
High-cost altruism	Selfish choice	Female	1.173	0.635	2.167	0.091
	<b>Sucker's choice</b>	<b>Female</b>	<b>0.290</b>	<b>0.113</b>	<b>0.749</b>	<b>-0.707</b>
	Selfish choice	Male	0.665	0.346	1.280	-0.334
Self-enhancing motives	Sucker's choice	Male	0.505	0.217	1.174	-0.560
	<b>Selfish choice</b>	<b>Female</b>	<b>1.271</b>	<b>1.042</b>	<b>1.551</b>	<b>0.240</b>
	Sucker's choice	Female	0.743	0.511	1.081	-0.297
	Selfish choice	Male	1.021	0.781	1.336	0.021
	<b>Sucker's choice</b>	<b>Male</b>	<b>0.654</b>	<b>0.447</b>	<b>0.956</b>	<b>-0.425</b>

For females, empathic concern did not emerge as a significant predictor of either decision outcome for this group. Medium-cost altruism was a strong and consistent predictor: it significantly reduced the likelihood of Selfish Choices and increased the likelihood of choosing the sucker option. Interestingly, high-cost altruism was associated with a lower likelihood of sucker choices, suggesting that more extreme altruistic tendencies may not translate into self-sacrificial decisions when reciprocity is uncertain. Self-enhancing motives significantly increased the odds of selfish choices, though it was not significantly related to sucker choices in this group.

For males, a different pattern emerged. Empathic concern was the strongest predictor of sucker choices, with a significant positive association indicating that affective empathy plays a central role in motivating unreciprocated cooperation among men. In contrast to the female model, medium-cost altruism did not significantly predict either selfish or sucker choices in males, suggesting that this form of altruism may not drive their behaviour in the same way. High-cost altruism was negatively, but not significantly, associated with sucker choices, mirroring the direction observed in females. Self-enhancing motives predicted a reduced likelihood of selecting the sucker option among males, but—unlike in females—it did not significantly predict Selfish Choices.

Although confidence intervals for key predictors overlapped across sexes—statistically suggesting the absence of strong sex moderation—the divergence in which predictors reached significance within each model points to potentially meaningful differences in the psychological pathways underlying choices in a one-shot PD.

## 7 DISCUSSION

Commitment problems are ubiquitous in everyday life. From intimate relationships and friendships to workplace relationships and civic obligations, individuals frequently encounter situations where their interests may conflict with those of others. In such contexts, navigating social life often requires people to act with others' interests in mind—even in the absence of guarantees of reciprocity—

or at least to treat others' interests as equally important as their own (Tomasello, 2016). The present study's central question was to what extent evolved social preferences, particularly empathy and altruism, help individuals overcome these commitment problems by reducing selfish choices and, as such, fostering cooperation.

## **7.1 Summary of the results**

In this study, we modelled the commitment problem using a one-shot anonymous Prisoner's Dilemma (PD), a scenario in which individuals must decide whether to cooperate or cheat without knowing what the other will do. In this setting, the stakes are high: cheating offers immediate gain, but cooperation opens the possibility of mutual benefit (Camerer, 2004; Rand & Nowak, 2013). Consistent with theoretical expectations, most participants in our study selected the conditional cooperative option, meaning they were willing to cooperate only if they believed the other person would do the same. This approach reflects a trust-based strategy, where, even in an anonymous one-shot interaction, cooperative choices depend on the expectation of reciprocation, balancing the potential benefits against the risk of being exploited. Accordingly, the descriptive findings point to a general inclination toward conditional cooperation, a pattern that has been documented across experimental, anthropological, and cultural-evolutionary accounts of human sociality (e.g., Boehm, 2012; Henrich, 2017; Sperber & Baumard, 2012; Tomasello, 2016, 2019).

Building on this, our second research question explored which psychological traits and evolved social preferences (empathy and altruism) can explain individual variation in these decisions. Regarding personality traits, the results provide only limited support for their explanatory power in predicting choices in a one-shot PD study. Among the three traits examined—impulsivity, risk-taking, and self-regulation—only self-regulation showed a statistically significant association, and this was restricted to the Sucker's Choice. Specifically, individuals with higher self-regulation were more likely to choose the Selfish Choice. However, the effect was modest and did not replicate in the structural equation modeling framework, suggesting that self-regulation may exert only an indirect influence or share variance with other constructs. The lack of significant findings for impulsivity and sensation seeking also indicates that these traits may play a less central role in single-shot cooperative dilemmas when compared to more socially embedded traits such as empathy and altruism.

Our study adopted a differentiated approach by separating affective and cognitive components of empathy, and by parsing altruism into low-, medium-, and high-cost forms, to examine their distinct contributions to decision-making in a commitment dilemma. Affective empathy (empathic concern) was positively associated with the Sucker's Choice, but only in the male subsample in the structural equation models, suggesting that emotional responsiveness plays a unique role in motivating unreciprocated cooperation among men in the study

sample. In contrast, cognitive empathy (perspective-taking) showed no significant effects, indicating that understanding others' mental states may be less influential in one-shot interactions characterized by uncertainty about what the other would do. Altruism, too, exhibited differentiated effects: medium-cost altruism emerged as an ambivalent predictor of increased cooperation, particularly by reducing Selfish Choices, but at the same time increasing the likelihood of Sucker's Choice.

High-cost altruism displayed a more straightforward role. Specifically, it was negatively associated with sucker choices in the exploratory models and showed a significant effect only among females in the SEM models, suggesting that high-cost self-sacrificial tendencies do not always translate into unconditional cooperation. These findings underscore the theoretical importance of disaggregating prosocial tendencies into cost-sensitive and sex-specific profiles to better understand cooperative choices in a commitment problem.

Interestingly, our finding that empathic concern and medium-cost altruism increase the likelihood of choosing the sucker option highlights a double-edged sword: while empathy and altruism can reduce selfish behaviour, they may simultaneously render individuals more vulnerable to exploitation. This puzzling pattern resonates with Bloom's (2016) critique of empathy, in which he cautions against the "empathy trap"—the impulse to act on emotionally charged concern for others, which may feel morally justified but can result in suboptimal or even harmful outcomes. Bloom further argues that genuine moral action requires not less care for others, but a more reflective, rational form of compassion—one that weighs consequences and recognizes the risk of being exploited. This insight is particularly relevant in commitment problems, where empathic individuals may prioritize trust-based cooperation even when it leads to personal loss. Although such behaviour may be considered admirable from a moral perspective, it can be strategically naïve. Our study empirically illustrates this tension in the sample: empathy promotes cooperative choices but also increases susceptibility to being taken advantage of—a finding that, while intuitive, is rarely quantified in experimental settings (Bloom, 2016). This pattern also resonates with the concept of pathological altruism—a form of well-intentioned helping behaviour that inadvertently leads to harm, either to the self or others (Oakley, Knafo, Madhavan, & Wilson, 2012). In this framework, empathy-driven responses can override strategic reasoning or self-protection, particularly in social dilemmas where cooperation is not reciprocated. Pathological altruism highlights the darker side of prosocial motivations: under certain conditions, they may not only fail to produce beneficial outcomes but may facilitate exploitation.

Among the control variables, self-enhancing motives emerged as a significant and robust predictor. Specifically, self-enhancement increased the likelihood of self-interested choices and decreased the likelihood of unconditional cooperation. This finding aligns with theoretical work by Declerck and Boone (2016), who conceptualize self-enhancing motives as decision rules rooted in incentive-based cooperation. These motives reflect a tendency to cooperate primarily when it serves one's interest, driven by the brain's reward system, which

evaluates personal benefits in social exchanges. In the absence of external incentives such as punishment or reputational gain, self-enhancing individuals are more likely to cheat, prioritizing short-term personal gain over long-term collective benefit. In our study, this orientation manifested in a greater likelihood of selecting the selfish option in the Prisoner's Dilemma. The persistence of this effect, even after controlling for personality traits and evolved social preferences, underscores the role of self-focused heuristics in shaping behaviour in commitment situations. By contrast, group-inclusive motives—which promote trust-based cooperation rooted in a sense of group belonging—did not emerge as significant predictors in this context, suggesting that in anonymous, one-shot interactions, trust signals may carry less weight in guiding decisions.

Moreover, the results revealed small but statistically significant sex differences in choices in the one-shot Prisoner's Dilemma. Women were more likely to select the conditional cooperative strategy, while men were slightly overrepresented in both the self-interested and unconditional (sucker's) cooperation choices. This pattern suggests that women tend to cluster around a prosocial norm, whereas men display greater variability, appearing at both ends of the behavioural spectrum.

These behavioural differences align with significant sex differences in key psychological constructs: women scored higher on empathic concern, perspective-taking, altruism, and group-oriented thinking, while men scored higher on selfish tendencies and sensation seeking. Effect sizes ranged from small to moderate, with particularly notable differences in empathic concern and high-cost altruism.

The greater variability in men's scores supports evolutionary and neurobiological theories emphasising sex-specific adaptations shaped by hormonal and social pressures related to competition and risk-taking (Geary, 2010; Winegard et al., 2014). Developmental psychopathology research further supports these findings, showing males' higher likelihood of early-onset antisocial behaviour and broader behavioural variability (Moffitt et al., 2001). Evolutionary perspectives suggest this reflects a dual heritage of cooperative and competitive tendencies with greater male variance in social traits (Betzig, 2012; Sapolsky, 2017). However, although sex emerged as a significant predictor in the exploratory analyses, justifying the decision to run separate structural equation models for males and females to better capture potential sex-specific pathways in cooperative decision-making, the smaller sample size for males may have limited the statistical power to detect significant effects within this group. Therefore, some observed differences between sexes should be interpreted cautiously, and replication with larger and more balanced samples is warranted to strengthen these findings. Nevertheless, these small but consistent sex differences align with the understanding that humans are a species characterized by relatively minor sexual dimorphism, where men and women differ primarily in averages rather than absolute terms, with substantial overlap between the sexes (Cliquet & Avramov, 2018).

## 7.2 Limitations of the study

A central limitation of the present study is the use of a one-shot, anonymous version of the Prisoner's Dilemma (PD). While this design is valuable for isolating intrinsic social preferences, by removing reputational concerns and strategic incentives, it also represents an artificial setting that may not fully capture the dynamics of real-world social interactions (Camerer, 2004). In repeated games, cooperation can emerge as a rational strategy driven by long-term self-interest. In contrast, one-shot games are designed to reveal intrinsic prosocial tendencies. However, the lack of repeated interaction and anonymity may limit the ecological validity of the findings.

A further limitation concerns the uncertainty regarding personal culpability: participants were not informed whether they themselves or the other person had stolen the bottle. This departs from classical narrow rational-choice models, which assume perfect information and predictable outcomes (Simon, 1979). Under such uncertainty, individuals often rely on adaptive heuristics (Gigerenzer, 2003; Gigerenzer & Gaissmaier, 2011), including fairness norms or guilt avoidance, which may foster cooperation even when self-interest models would predict defection. Empirical work similarly suggests that tolerance for ambiguity, rather than risk attitudes, predicts costly prosocial behaviours such as cooperation and trust (Vives & FeldmanHall, 2018). Moreover, emotions and social preferences, such as empathy and altruism, can function as commitment devices that counteract short-term temptations (Frank, 1988), potentially leading some individuals to cooperate despite uncertainty about their own culpability. Because our design did not manipulate or measure this uncertainty, we cannot determine its effect on cooperative choices. Future research should vary the degree of certainty about personal responsibility to assess its impact on cooperative outcomes.

This limitation reflects a broader theoretical critique. As Tomasello (2016) argues, traditional game-theoretic models often portray individuals as isolated agents—*asocial monads* (p.14)—in competition with others, neglecting the complex web of interdependencies that characterizes human social life. In contrast, the stakeholder model (Roberts, 2005; Tomasello, 2016) emphasises that individuals often have a vested interest in the well-being of others, whether as kin, partners, colleagues, or group members. Helping, in such contexts, is not a sacrifice but an investment in the stability of ongoing relationships. This interdependence fundamentally alters the payoff structure of social interactions: when individuals care about each other's welfare because their outcomes depend on it, the logic of the Prisoner's Dilemma no longer applies. From this perspective, altruism is not a deviation from self-interest, but a natural feature of social life.

The abstraction inherent in one-shot PD designs may therefore obscure the role of interdependence in motivating cooperative behaviour. In real life, individuals are embedded in complex social networks where the well-being of others directly affects their outcomes. Cooperative choices are highly context-

sensitive: people are more likely to cooperate when they expect others to do so, and more likely to cheat when they expect cheating (Frank, 1988, 2011). This dynamic helps explain why cooperation can flourish in some environments and collapse in others. The present findings must thus be interpreted in light of the broader insight that human social preferences are not static traits but have evolved to meet the demands of interdependent living. In cognitively and socially complex species—humans in particular—individuals are not isolated decision-makers but are enmeshed in webs of mutual dependence (e.g. Bowles & Gintis, 2011; Krebs, 2022; Lindenfors, 2017; Turner, 2014, 2020).

Frank (1988, 2011) developed the commitment model, a foundational account of moral behaviour in which he argued for evolved emotions that serve as commitment devices to motivate cooperative actions even when short-term incentives favour cheating. His model highlights how feelings such as guilt, shame, and moral outrage inhibit opportunistic behaviour by fostering aversion to exploiting others, even in anonymous or one-shot interactions. Crucially, Frank argues that the best way to be seen as a trustworthy, cooperative group member is to be one. Emotional predispositions thus align short-term behaviour with long-term reputational benefits, addressing the commitment problem through internalized moral motivation. While our study builds on this influential framework, it did not explicitly examine these moral emotions as commitment devices. Instead, we focused more narrowly on empathic concern as a foundational emotional capacity underlying cooperative decision-making. For a fuller treatment of moral emotions as commitment devices, see Pauwels & De Buck (2026), where this extension of Frank's model is developed in greater detail.

Although this study does not directly investigate moral emotions such as guilt or shame, it underscores the fundamental role of empathic concern and altruism in moral decision-making. Drawing on Tomasello's transactive framework (2019, 2024), empathy should be understood not as a fixed trait, but as a developing capacity shaped through social interaction and practice. In particular, the ability to care about others' well-being appears to be a vital prerequisite for the emergence of prosocial behaviour (de Waal, 2016). Without this capacity, individuals are less likely to internalize social expectations or regulate their behaviour in ways that consider others' interests. While empathy alone may not guarantee moral action, it plays an essential developmental and motivational role. Tomasello's developmental perspective further emphasises that moral agency arises alongside the child's growing abilities in shared intentionality, normative self-regulation, and recursive perspective-taking—capacities that depend fundamentally on empathy. From this viewpoint, empathy is not merely a moral sentiment but a core mechanism within the developmental architecture of moral cognition and behaviour.

In short, this evolutionary account frames empathy and altruism as core features of human moral psychology, shaped by the demands of interdependent living. It also suggests that the one-shot, anonymous PD used in the present study

may underestimate the full range of cooperative motivations that emerge in more ecologically valid, relationally embedded contexts (e.g. Gintis, 2017).

## 8 CONCLUSION

In sum, this study conceptualized the one-shot Prisoner's Dilemma (PD) as a moral commitment problem: situations in which individuals face tension between short-term self-interest and adherence to cooperative norms. Rather than focusing on payoff maximization, we examined how evolved social preferences, in particular empathy and altruism, function as intrinsic commitment devices, enabling individuals to cooperate even when external enforcement is absent (Frank, 1988; Krebs, 2022). It highlights that cooperation is often guided by expectations of reciprocity and trust, rather than by strict payoff logic. By framing norm-violating behaviour broadly as a moral commitment problem, rather than relying on narrow legalistic definitions, our framework transcends disciplinary boundaries. This approach also aligns with a growing scholarly interest in (anti)-social behaviour from an evolutionary and biosocial perspective (e.g., Durrant & Ward, 2015; Posick, Rocque, & Barnes, 2021; Svingen, 2023). Many criminological contexts, such as co-offending, joint liability or group-based norm enforcement, can be understood as interdependent moral commitment problems, in which evolved social preferences shape how individuals respond (Tomasello, 2016, 2019). These preferences can sustain cooperation while simultaneously exposing individuals to potential exploitation, illustrating the double-edged nature of moral emotions in social decision-making. Future research using ecologically embedded, relationally interdependent designs will be important for further elucidating how these commitment devices operate in real-world social environments, thereby advancing a theoretically grounded understanding of cooperation and moral behaviour in criminological contexts.

## BIBLIOGRAPHY

- Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycho-lexical study. *Psychological Monographs*, 47 (1, Whole No. 211).
- American Psychological Association. (n.d.). Personality trait. *APA Dictionary of Psychology*. Retrieved July 21, 2025, from <https://dictionary.apa.org/personality-trait>.
- Axelrod, R. (1984/2006). *The evolution of cooperation*. Basic Books.
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211(4489), 1390-1396.
- Batson, C. D. (2017). The empathy-altruism hypothesis: What and so what? In E. M. Seppälä, E. Simon-Thomas, S. L. Brown, M. C. Worline, C. D. Cameron, & J. R. Doty (Eds.), *The Oxford Handbook of Compassion Science* (pp. 27-40). Oxford University Press.
- Betzig, L. (2012). Means, variances, and ranges in reproductive success: comparative evidence. *Evolution and Human Behavior*, 33, 309-317.
- Bloom, P. (2016). *Against empathy. The case for rational compassion*. HarperCollins Publishers.
- Bluhm, R. (2017). Gender and empathy. In H. Maibom (Eds.), *The Routledge handbook of philosophy of empathy* (pp. 377-387). Routledge.
- Boehm, C. (2012). *Moral origins. The evolution of virtue, altruism, and shame*. Basic Books.
- Bowles, S., & Gintis, H. (2011). *A cooperative species. Human reciprocity and its evolution*. Princeton University Press.
- Burnstein, E., Crandall, C., and Kitayama, S. (1994). Some neo-Darwinian decision rules for altruism: Weighing cues for inclusive fitness as a function of the biological importance of the decision. *Journal of Personality and Social Psychology*, 67, 773-789.
- Burt, C. H. (2020). Self-control and crime: beyond Gottfredson & Hirschi's theory. *Annual Review of Criminology*, 3(1), 43-73.
- Buss, D. M. (2024). *Evolutionary psychology. The new science of the mind* (7th ed.). Routledge.
- Buss, D. M., Haselton, M. G., Shackelford, T. K., Bleske, A., & Wakefield, J. C. (1998). Adaptations, exaptations, and spandrels. *American Psychologist*, 53(5), 533-548.
- Camerer, C. (2004). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

- Campbell, A. (2013). *A mind of her own: The evolutionary psychology of women*. Oxford University Press.
- Cattell, R. B. (1973). *Personality and mood by questionnaire*. Jossey-Bass.
- Christov-Moore, L., Simpson, E. A., Coude, G., Grigaityte, K., Iacoboni, M., Ferrari, P. F. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience and Biobehavioral Reviews* 46: 604-627.
- Cliquet, R., & Avramov, D. (2018). *Evolution, Science and Ethics in the Third Millennium. Challenges and choices for humankind*. Springer.
- Cohen, J. (1988/2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory and Five-Factor Inventory professional manual*. Psychological Assessment Resources
- Daly, M., & Wilson, M. (2001). Risk-taking, intrasexual competition, and homicide. In J. A. French, A. C. Kamil, & D. W. Leger (Eds.), *The Nebraska Symposium on Motivation: Vol. 47. Evolutionary psychology and motivation* (pp. 1-36). University of Nebraska Press.
- Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, 31, 169-193.
- Decety, J. (2011). The neuroevolution of empathy. *Annals of the New York Academy of Sciences*, 1231(1), 35-45.
- Declerck, C., & Boone, C. (2016). *Neuroeconomics of prosocial behavior. The compassionate egoist*. Academic Press.
- De Waal, F. B. M. (2010). *The age of empathy. Nature's lessons for a kinder society*. Random House Inc.
- Durrant, R. (2019). Evolutionary approaches to understanding crime: Explaining the gender gap in offending. *Psychology, Crime & Law*, 25(6), 589-608.
- Durrant, R., & Ward, T. (2015). *Evolutionary criminology: Towards a comprehensive explanation of crime*. Academic Press.
- Eysenck, H. J. (1991). Dimensions of personality: 16, 5, or 3? Criteria for a taxonomic paradigm. *Personality and Individual Differences*, 12, 773-790.
- Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach*. Plenum.
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960), 785-791.
- Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. Sage Publications Ltd.

Fleeson, W. (2012). Perspectives on the person. Rapid growth and opportunities for integration. In K. Deaux & M. Snyder (Eds.), *The Oxford handbook of personality and social psychology* (pp. 33-63). Oxford University Press.

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. WW Norton & Co.

Frank, R. H. (2001). Cooperation through emotional commitment. In R. M. Nesse (Ed.), *Evolution and the capacity for commitment* (pp. 57-76). Russell Sage Foundation.

Frank, R. H. (2011). The strategic role of the emotions. *Emotion Review*, 3(3), 252-254.

Geary, D. C. (2021). *Male, female. The evolution of human sex differences* (3rd ed.). American Psychological Association.

Gigerenzer, G. (2023). *The intelligence of intuition*. Cambridge University Press.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451-482.

Gintis, H. (2017). *Individuality and entanglement: the moral and material bases of social life*. Princeton University Press.

Greene, W. H. (2018). *Econometric analysis* (8th ed.). Pearson.

Gottfredson, M. R., & Hirschi, T. (1990). A general theory of crime. In *A General Theory of Crime*. Stanford University Press.

Gray, J.A. (1975). *Elements of a two-process theory of learning*. Academic Press.

Henrich, J. (2017). *The secret of our success. How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.

Hoyle, R. H., & Davisson, E. K. (2016). Varieties of self-control and their personality correlates. In K. D. Vohs and R. F. Baumeister (Eds.), *Handbook of self-regulation: Research, theory, and applications* (3rd ed., 396-413). Guilford Press.

Hrdy, S. B. (2009/2011). *Mothers and others: The evolutionary origins of mutual understanding*. Harvard University Press.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.

Jones, S. E., Miller, J. D., & Lynam, D. R. (2011). Personality, antisocial behavior, and aggression: A meta-analytic review. *Journal of Criminal Justice*, 39(4), 329-337. <https://doi.org/10.1016/j.jcrimjus.2011.03.004>

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Press.

Krebs, D. L. (2011). *The origins of morality. An evolutionary account*. Oxford University Press.

Krebs, D. L. (2022). *Survival of the virtuous. The evolution of moral psychology*. Oxford University Press.

Larsen, R. J., & Buss, D. M. (2021). *Personality psychology. Domains of knowledge about human nature* (7th ed.). McGraw Hill.

Lindenfors, P. (2017). *For whose benefit? The biological and cultural evolution of human cooperation*. Springer.

Löffler, C. S., & Greitemeyer, T. (2023). Are women the more empathetic gender? The effects of gender role expectations. *Current Psychology*, 42(1), 220-231.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11(3), 320-341.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24(1), 167-202.

Milyavskaya, M., Berkman, E. T., & De Ridder, D. T. (2019). The many faces of self-control: Tacit assumptions and recommendations to deal with them. *Motivation Science*, 5(1), 79.

Mischel, W., Cantor, N., & Feldman, S (1996). Principles of self-regulation: The nature of willpower and self-control. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 329-360). Guilford.

Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P. A. (2004). *Sex differences in antisocial behaviour. Conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study*. Cambridge University Press.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, 58(4), 361-383.

Nowak, M. A., & Highfield, R. (2011). *SuperCooperators. Altruism, evolution, and why we need each other to succeed*. Free Press.

Oakley, B., Knafo, A., Madhavan, G., & Wilson, D. S. (Eds.) (2012). *Pathological altruism*. Oxford University Press.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Pauwels, L. J. R., & De Buck, A. (2026). *Breaking shared norms, the moral commitment problem, and moral emotions. An evolutionary analysis*. Cambridge Scholars Publishing.

Posick, C., Rocque, M., & Barnes, J. C. (2021). *Fitting the facts of crime: An invitation to biopsychosocial criminology*. Temple University Press.

Pratt, T. C., & Cullen, F. T. (2000). The empirical status of Gottfredson and Hirschi's general theory of crime: A meta-analysis. *Criminology*, 38(3), 931-964.

Preston, S. D., & De Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25(1), 1-20.

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in cognitive sciences*, 17(8), 413-425.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's Dilemma: a study in conflict and cooperation*. University of Michigan Press.

Sapolsky, R. (2017). *Behave: the biology of humans at our best and worst*. Penguin Press.

Schelling, T. C. (1960). *The strategy of conflict*. Oxford University Press.

Scott, N., & Seglow, J. (2007). *Altruism*. Open University Press.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120.

Simon, H. A. (1979). Rational decision making in business organizations. *The American Economic Review*, 69(4), 493-513

Sober, E., & Wilson, D. S. (1999). *Unto others. The evolution and psychology of unselfish behavior*. Harvard University Press.

Sommer, I. (2020). *Het vrouwenbrein*. Atlas Contact.

Sperber, D., & Baumard, N. (2012). Moral reputation: An evolutionary and cognitive perspective. *Mind & Language*, 27(5), 495-518.

Steinberg, L., Albert, D., Cauffman, E., Banich, M., Graham, S., & Woolard, J. (2008). Age differences in sensation seeking and impulsivity as indexed by behavior and self-report: evidence for a dual systems model. *Developmental Psychology*, 44(6), 1764-1778.

Sussman, R. W. & Cloninger, C. R. (Eds.) (2011). *Origins of altruism and cooperation*. Springer.

Svingen, E. (2023). *Evolutionary criminology and cooperation: retribution, reciprocity, and crime*. Springer.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72(2), 272-322.

Tomasello, M. (2016). *A natural history of human morality*. Harvard University Press.

Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Harvard University Press.

Tomasello, M. (2024). *Agency and cognitive development*. Oxford University Press.

Tooby, J., & Cosmides, L. (1990). On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality*, 58, 17-68.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46(1), 35-57.

Trivers R. L. (2006). Reciprocal altruism: 30 years later. In P. M. Kappeler & C. P. van Schaik (Eds.), *Cooperation in Primates and Humans* (pp. 67-83). Springer.

Turner, J. (2014). The evolution of affect, sociality, altruism, and conscience in humans. In V. Jeffries (Ed.), *The Palgrave handbook of altruism, morality, and social solidarity* (pp. 275-301). Palgrave Macmillan.

Turner, J. H. (2020). *On human nature: The biology and sociology of what made us human*. Routledge.

Vazsonyi, A. T., Mikuška, J., & Kelley, E. L. (2017). It's time: A meta-analysis on the self-control-deviance link. *Journal of Criminal Justice*, 48, 48-63.

Vives, M.-L., & FeldmanHall, O. (2018). Tolerance to ambiguous uncertainty predicts prosocial behavior. *Nature Communications*, 9(1), Article 2156. <https://doi.org/10.1038/s41467-018-04631-9>

Williams, G. C. (1966). *Adaptation and natural selection*. Princeton University Press.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata Journal*, 12(2), 308-331.

Winegard, B. M., Winegard, B., & Geary, D. C. (2014). Eastwood's brawn and Einstein's brain: An evolutionary account of dominance, prestige, and precarious manhood. *Review of General Psychology*, 18(1), 34-48.

Zuckerman, M. (1994). Impulsive unsocialized sensation seeking: The biological foundations of a basic dimension of personality. In J. E. Bates & T. D. Wachs (Eds.), *Temperament: Individual differences at the interface of biology and behavior*. American Psychological Association.

Zuckerman, M., & Kuhlman, D. M. (2000). Personality and risk-taking: Common biosocial factors. *Journal of Personality*, 68(6), 999-1029.