Judgment and Decision Making (2025), **vol. 20**:e16 1–18
doi:10.1017/jdm.2025.1

EMPIRICAL ARTICLE

Toward a (more) parsimonious account of the link between 'dark' personality and social decision-making in economic games

Benjamin E. Hilbig[1], and Isabel Thielmann[2]

[1]RPTU University Kaiserslautern-Landau, Germany and [2]Max Planck Institute for the Study of Crime, Security and Law, Freiburg, Germany

Corresponding author: Benjamin E. Hilbig; Email: b.hilbig@rptu.de

Received: 28 June 2024; **Revised:** 6 December 2024; **Accepted:** 22 January 2025

Keywords: Economic games; situational affordances; dark traits; D factor

Abstract

There are large individual differences in prosocial vs. antisocial behavior as studied via economic games. Prominent among the personality traits that have been considered as potential correlates are 'dark' traits (especially Machiavellianism, Narcissism, Psychopathy, and Sadism). However, although such traits should account for choices in games, the corresponding associations are weak and inconsistent, leading to a state of knowledge that lacks specificity and parsimony. We argue and demonstrate across 10 studies and 8 economic games (total N = 10,474) that a clearer picture emerges once (a) including games that (also) afford the expression of forgiveness (vs. retaliation) and/or (dis)trust and (b) considering the shared vs. unique aspects of dark traits. Specifically, we find that (i) the common core of all dark traits—the dark factor of personality, D—consistently predicts antisocial behavior in all games, (ii) dark traits also (though less strongly) predict antisocial behavior, and (iii) dark traits do so almost entirely due to D. We discuss the theoretical and methodological implications of these findings for the study of individual differences in pro- vs. antisocial behavior in economic games.

1. Introduction

Prosocial decisions—i.e., 'actions that benefit others, usually at personal costs, in situations of interdependence' (Thielmann et al., 2021, p.1)—are crucially important to societal functioning (Axelrod & Hamilton, 1981; Colman, 2003; Dawes & Messick, 2000; Nowak, 2006), from everyday interactions in dyads to some of the most pressing challenges faced by humans as a species (Van Lange & Rand, 2022). To better understand the determinants of (pro)social decision-making in interdependent situations, a range of economic games have been developed and studied (Camerer, 2011; Thielmann et al., 2021; van Dijk & De Dreu, 2021), some across hundreds of experiments and thousands of participants (for authoritative meta-anlyses, see, e.g., Balliet et al., 2011; Balliet et al., 2009; Balliet & Van Lange, 2012; Engel, 2011; Johnson & Mislin, 2011; Sally, 1995; Spadaro et al., 2022; Zelmer, 2003). Among the most fundamental insights from these many studies is the finding that there are substantial and consistent individual differences in corresponding choice behavior.

© The Author(s), 2025. Published by Cambridge University Press on behalf of Society for Judgment and Decision Making and European Association for Decision Making. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/jdm.2025.1 Published online by Cambridge University Press

Correspondingly, over the past two decades in particular, research has set out to identify the stable dispositions and, more specifically, personality traits that explain such individual differences in choice behavior. Recently, a notable number of studies (e.g., Baggio & Benning, 2022; Böckler et al., 2017; Fatfouta et al., 2018; Malesza, 2018, 2020; Mussel & Hewig, 2016; Nehrlich et al., 2019; Pfattheicher et al., 2017; Szijjártó et al., 2018) has focused on socially/ethically aversive (aka 'dark') traits, most prominently the 'dark tetrad' traits of Machiavellianism, Narcissism, Psychopathy, and Sadism (Paulhus et al., 2020). Given that such traits are commonly understood to be conducive to callous and exploitative behavior (Bonfá-Araujo et al., 2022), they are arguably prime candidates to explain individual differences in pro- vs. antisocial[1] choices in economic games. More specifically, (non)exploitativeness, forgiveness (vs. retaliation), and/or (dis)trust are key defining features of dark traits. The expression of these very features may, in turn, be afforded in economic games depending on whether the particular game allows for the expression of unconditional concern for others' welfare, conditional concern for others' welfare (reciprocity), and/or beliefs about others' prosociality (Thielmann et al., 2020).[2] Thus, it is reasonable to expect that dark traits will—depending on the affordances present in a game and the tendencies a trait subsumes (by its definition)—account for pro- vs antisocial selfish behavior in corresponding games.

Providing some support for this conjecture, an authoritative meta-analysis (Thielmann et al., 2020) confirmed that the dark tetrad traits are negatively linked to prosociality across several games affording the expression of said tendencies with small-to-moderate effect sizes ($-.11 < r < -.18$). However, on closer inspection, the picture is far from satisfactory. Even setting aside that effect sizes for the four dark tetrad traits exhibited the largest extent of publication bias among the more than 50 traits studied by Thielmann et al. (2020) (and that some of the largest theory-consistent effect sizes by far stem from studies with effective sample sizes as small as $N < 50$; e.g., Bereczkei et al., 2015; Mokros et al., 2008), the most noteworthy caveat was that effects on the level of specific games were strikingly inconsistent: As can be gleaned from Table 1, there were multiple theory-inconsistent null effects, especially (but not exclusively) for those games that do not primarily afford (non-)exploitation.[3] Specifically, none of the traits accounted for responder behavior in the Ultimatum Game that primarily taps into conditional concern for others' welfare (reciprocity) and thus affords the expression of forgiveness (vs. retaliation). Similarly, in games tapping into beliefs about others prosociality and thus affording the expression of (dis)trust—the Ultimatum Game as proposer, the Trust Game as trustor, and Social Dilemma Games—the majority of meta-analytical effects were indistinguishable from zero.

Moreover, none of the dark tetrad traits actually yielded a robust effect across all games, nor was there any indication that any dark tetrad trait (or subset of traits) is particularly, let alone exclusively, relevant for any one game (or subset of games). This is problematic insofar as all games included in the meta-analysis afford key aspects of the traits in question, namely, exploitation, retaliation (negative reciprocity), and/or distrust (Thielmann et al., 2021; Thielmann et al., 2020). Indeed, even across games sharing certain affordances (e.g., the Ultimatum Game as proposer, the Trust Game as trustor, and social dilemmas all afford the expression of distrust), there was no consistent association with any of the dark tetrad traits.

Overall then, the state of knowledge reduces to the near truism that some aversive traits sometimes relate to behavior in some games. Clearly, such a vague and unspecific conclusion cannot be satisfactory. Indeed, so long as the number of potential aversive traits is not bounded by any theory

[1] We use the term antisocial to denote any behavior that does not increase (or indeed decreases) others' welfare.

[2] The theoretical framework by Thielmann et al. (2020) further proposes self-regulation as a fourth dispositional tendency that may account for individual differences in pro-social behavior. However, evidence for the role of self-regulation is weak at best (see also Popov & Thielmann, in press).

[3] Of note, research published after (and thus not included in) the meta-analysis by Thielmann et al. (2020) does not indicate a more theory-consistent picture, reporting null effects—e.g., of Psychopathy on social dilemma cooperation (Baggio & Benning, 2022)—or even positive effects—e.g., of Narcissism on social dilemma cooperation (Malesza, 2020) and on sharing in a paradigm akin to the Dictator Game (Malesza & Kalinowski, 2019).

https://doi.org/10.1017/jdm.2025.1 Published online by Cambridge University Press

Table 1. *Mean true-score correlations corrected for unreliability (number of studies in parentheses) between dark tetrad traits and games as per the meta-analysis by Thielmann et al. (2020).*

		Dark trait			
Game	Affordances	M	N	P	S
Dictator Game	EX	**−.20** (16)	**−.16** (12)	**−.19** (15)	*
Ultimatum Game (proposer)	(EX), DI[a]	−.04 (6)	**−.11** (5)	−.05 (8)	*
Ultimatum Game (responder)	(EX), RE	−.02 (8)	.00 (8)	−.04 (12)	*
Trust Game (trustor)	(EX), DI	**−.16** (10)	*	.00 (3)	*
Trust Game (trustee)	EX, (RE)	**−.20** (11)	*	−.12 (3)	*
Social dilemmas	EX, (RE), DI	**−.16** (20)	−.07 (24)	**−.13** (17)	−.04 (6)

Note: [a] Note that in the Ultimatum Game as proposer, distrust implies more prosociality (higher offers due to fear of rejection)—unlike in the Trust Game and Social Dilemmas in which distrust implies less prosociality; * if the number of studies was smaller than 3, no coefficients were calculated or reported by Thielmann et al. (2020); Coefficients in boldface are statistically significant at $p < .05$; EX = exploitation; RE = retaliation; DI = distrust; N = Narcissism, M = Machiavellianism, P = Psychopathy, S = Sadism.

(on the contrary, it is continually increasing), the state of knowledge has very little '*empirical content*' in Popperian terms (Popper, 1934/2005), that is, it lacks precision ('Bestimmtheit') and is therefore difficult if not impossible to falsify (Glöckner & Betsch, 2011). Clearly, then, 'the inflationary number of trait concepts and corresponding measures gives cause for concern' (Thielmann et al., 2022, p. 290), implying that a more precise and parsimonious theoretical account is needed and ought to be strictly preferred unless some specificity (unique or even exclusive links between certain traits and certain games) can be established (for similar arguments, see Leising et al., 2022).

We see two primary (not mutually exclusive) reasons why this has not been established so far. The first is mainly empirical in nature and essentially concerns the games (not) commonly studied and thus the situational affordances (not) covered in the literature. The second is more theoretical in nature and concerns the shared vs. unique aspects of aversive traits, that is, aversive traits vis-à-vis each other and their common core. In what follows, we discuss both in turn.

On the empirical side, the lack of specificity may be due to the limited coverage of games and corresponding types of interdependent situations. Specifically, there is an almost exclusive reliance on games that primarily provide a possibility for (non-)exploitation. This major affordance is well-covered across commonly studied games and indeed largely isolated from other affordances in the Dictator Game: The prosocial response in the Dictator Game (giving) can only stem from nonexploitation but not from forgiveness or trust (Thielmann et al., 2021). By contrast, the commonly studied games rarely tap into, let alone properly isolate, the other two major affordances that ought to allow for the expression of dark traits: conditional concern for others' welfare (i.e., reciprocity, allowing for the expression of forgiveness vs. retaliation), and/or beliefs about others' prosociality (allowing for the expression of (dis)trust). In fact, among the games commonly studied, reciprocity is only primarily afforded in a single one (the Ultimatum Game as responder) and even in this case reciprocity is not properly isolated: Because retaliation is costly to the responder, the game confounds conditional concern for others (i.e., forgiveness) with greed and thus a lack of unconditional concern for others. In other words, the prosocial response (accepting even an unfair offer) can stem from either forgiveness or the desire to maximize one's own payoff (or both). Indeed, the very fact that retaliation and greed imply contrary responses in the Ultimatum Game (as responder) may actually explain the consistent null effects across dark traits (see Table 1). To remedy this problem, a game that allows for the expression of retaliation (vs. forgiveness) independent of greed—such as the uncostly retaliation game (Hilbig et al., 2016)—would need to be studied.

Similarly, whereas some games do afford beliefs about others' prosociality, namely the Trust Game (as trustor) and social dilemmas like the Prisoner's Dilemma or Public Goods Game, once more several dispositional tendencies are confounded within the same overt choices: In all of these games, trust is confounded with social welfare concerns (Thielmann et al., 2021) and thus the prosocial response (trust in the Trust Game and cooperation in a social dilemma) can stem either from trust or aspects of nonexploitation or both (Hilbig et al., 2018). To remedy this problem, games that separate trust from (non-)exploitation—such as the Faith Game (Kiyonari & Yamagishi, 1999) or a coordination game like the Stag Hunt Game (Skyrms, 2001)—would need to be studied. Indeed, the almost complete absence of studies linking personality to behavior in coordination games (such as the Stag Hunt) essentially means that prosociality in situations of compatible interests is hugely understudied—despite evidence that people perceive most of their everyday social interactions to be far more akin to coordination games (like the Stag Hunt) than to social dilemmas (like the Prisoner's Dilemma) (Balliet et al., 2022; Columbus et al., 2021). In summary, the overall picture linking aversive traits to behavior in games may well lack specificity simply because only a limited range of games has been studied, which, in turn, fails to isolate the affordances that would allow for the expression of forgiveness (vs. retaliation) and (dis)trust which are key aspects of specific dark traits.

Second, on the more theoretical side, the inconsistent empirical picture and lack of specificity may also be due to the conceptual nature of aversive traits, that is, their shared vs. unique aspects. There is now broad evidence and consensus that all aversive traits share a single common core (Muris et al., 2017; Schreiber & Marcus, 2020; Vize et al., 2018). This core has been termed the dark factor of personality, or simply D (Moshagen et al., 2018). Conceptually, D constitutes the shared 'aversive essence' of all aversive traits and thus accounts for their association with any behavior that can be understood as an expression of pro- vs. antisocial tendencies (Hilbig et al., 2023; Hilbig et al., 2024; Moshagen et al., 2018; Scholz et al., 2023). As per the very definition of D as 'the tendency to maximize one's individual utility—disregarding, accepting, or malevolently provoking disutility for others—accompanied by beliefs that serve as justifications' (Moshagen et al., 2018, p. 657), this common core of all aversive traits is immediately relevant for exploitation (individual utility maximization at others' cost), retaliation (provoking disutility for others), and distrust (beliefs that serve as justifications; Hilbig et al., 2022). Indeed, this also sets D conceptually apart from other relatively broad traits such as HEXACO Honesty–Humility (Ashton et al., 2014) or trait-like concepts such as Social Value Orientations (Murphy & Ackermann, 2014; Van Lange, 1999)—both of which capture individual differences in unconditional concern for others' welfare (i.e., (non)exploitation) rather than in conditional concern for others (i.e., retaliation vs. forgiveness) or beliefs about others' prosociality (i.e., distrust; Thielmann et al., 2020).

In any case, since all aversive traits share D to some extent, they may be associated with pro- vs. antisocial behavior in *any* economic game affording the expression of one or more of these tendencies (i.e., exploitation, retaliation, distrust). However, although all aversive traits can be understood as manifestations of D (Bader et al., 2023; Scholz et al., 2024), they differ in several important ways which, in turn, may explain the inconsistent and often weak evidence linking these traits to behavior in economic games. First, aversive traits differ in how broadly they reflect defining aspects of D. For example, whereas Sadism primarily reflects the aspect of inflicting disutility on others (in a very specific way), Machiavellianism reflects both the aspect of individual utility maximization (generally) and justifying beliefs (such as cynical distrust). At least in part, such differences might explain heterogeneity across games: Because different games allow for the expression of different tendencies, narrow traits that reflect only some aspects of D may only be associated with behavior in corresponding games (e.g., Machiavellianism especially with games that allow for the expression of distrust). Second, aversive traits generally differ notably in their D-saturation, that is, how much of the variance within such a trait is due to D and thus, ultimately, how aversive they are (Moshagen et al., 2018; Moshagen et al., 2020). In turn, single aversive traits typically entail other, per se nonaversive aspects beyond D—such as admiration in Narcissism (Back et al., 2013) or boldness in Psychopathy (Bader et al., 2023). Therefore, measures of these traits actually assess D (to varying degrees) plus other aspects that are not immediately and generally relevant to pro- vs. antisocial behavior. In other words, they 'dilute' the measurement of D (Hilbig et al., 2024)—their shared aversive essence that, in theory, links them to behavior in economic games—with other features. This would explain, at least in part, heterogeneity

https://doi.org/10.1017/jdm.2025.1 Published online by Cambridge University Press

between traits and recurring null findings: The stronger any trait dilutes D with aspects irrelevant to prosociality, the less likely it is to have any consistent effect on behavior in economic games.

Taken together, the following predictions can be derived from the conceptual framework of D. First, D should be negatively associated with prosocial behavior in all games that afford exploitation, retaliation, and/or distrust—which ought to hold in a broad array of games that, taken together, provide better coverage/isolation of these affordances than has been achieved in prior studies. Second, because all specific aversive traits are manifestations of D, they will typically also be negatively linked to prosocial behavior in games (although effect sizes may vary due to their differences in D-saturation and the defining aspects of D they reflect); however, because dark traits typically 'dilute' D with other aspects, their effect sizes should—on average across games—be smaller than the effect of D itself. Finally, and most importantly, because the D framework mandates that the effect of any aversive trait on game behavior is *due to* its aversive essence, D, controlling for D should lead to null (or even positive) effects of single aversive traits on prosociality in economic games. In other words, aversive traits should no longer be (negatively) linked to prosocial choices once controlling for D—a finding that has already been demonstrated for behavior in the Dictator Game (Moshagen et al., 2018) and a measure of social value orientation (Hilbig et al., 2023) but not for any game(s) affording forgiveness (vs. retaliation) or (dis)trust. Across a total of 10 studies (and eight economic games[4]), we tested these predictions.

2. Methods

The studies were not preregistered, but all entailed the exact same design and methodology that was fixed a priori, varying only in the economic game presented to participants. The studies are therefore reported on jointly. All studies were run in German, and all participants had at least good skills in the study language (≥95% were native speakers in each study). Each study involved a direct measure of D. The German version (Bader et al., 2022) of the D70 (Moshagen et al., 2020), which includes a total of 70 items, 50% of which are reverse-coded, that were presented in random order (per participant). Example items include 'My own pleasure is all that matters' or 'I tend to forgive the wrongs I have suffered' (reverse-keyed). In each study, the measure of D was followed by the German version (Blötner et al., 2022) of the short dark tetrad questionnaire (SD4, Paulhus et al., 2020) involving a total of 28 items presented in random order (per participant) to measure Narcissism, Machiavellianism, Psychopathy, and Sadism (with 7 items each). Example items include 'I know that I am special because people keep telling me so' (Narcissism subscale) and 'I know how to hurt someone with words alone' (Sadism subscale). Finally, in each study, one of the games described below was presented. Table 2 provides an overview of studies, games, the observed proportion of prosocial choices (which ranged from .55 to .91), as well as sample characteristics. As can be seen, all samples were heterogeneous in terms of sex and age.

All games were hypothetical and asked participants to imagine interacting with another individual randomly drawn from the population whom they do not know and will not knowingly meet. The verbatim instructions for all games can be found on the OSF (https://osf.io/5vjws/). Because participants self-selected into the studies (see below), the order of studies was quasi-random, that is, after completion of a study, the choice of which game would be included in the subsequent study was determined at random.

Participants were recruited via https://darkfactor.org/, a website open to the general public offering free and anonymous self-assessments and feedback on D. All studies were compatible with ethical requirements, and there was no deception involved. Data collection via the website was approved by the local ethics committee (IRB) of the RPTU University Kaiserslautern-Landau, Department of Psychology (approval #LEK-154 and #LEK-567). Participants provided informed consent prior to completing any items and confirmed the seriousness of their responses as well as consenting (again) to

[4] As detailed below, two of the games were studied twice: once in their more common variant with a nonbinary response format and once in a binary response variant.

https://doi.org/10.1017/jdm.2025.1 Published online by Cambridge University Press

Table 2. *Overview of studies 1–10.*

Study #	Game	Sample size (*N*)	Number (proportion) of prosocial responses	Sample demographics
1	Dictator Game	1057	625 (.59)	755 female, 271 male, 10 other; age: *M* = 31.7 (*SD* = 9.5), range: 18–74
2	Dictator Game (binary)	1043	570 (.55)	585 female, 439 male, 17 other; age: *M* = 35.2 (*SD* = 12.1), range: 18–86
3	Ultimatum Game (Proposer)	1053	960 (.91)	567 female, 469 male, 17 other; age: *M* = 38.3 (*SD* = 13.3), range: 18–80
4	Ultimatum Game (Responder)	1026	719 (.70)	573 female, 440 male, 12 other; age: *M* = 43.4 (*SD* = 16.5), range: 18–88
5	Uncostly Retaliation Game (Responder)	1010	713 (.71)	652 female, 342 male, 15 other; age: *M* = 29.1(*SD* = 7.6), range: 18–71
6	Uncostly Retaliation Game (Responder) (binary)	1077	978 (.91)	633 female, 431 male, 11 other; age: *M* = 35.2 (*SD* = 12.5), range: 18–90
7	Faith Game	1050	613 (.58)	739 female, 296 male, 14 other; age: *M* = 29.5 (*SD* = 8.3), range: 18–84
8	Prisoner`s Dilemma	997	565 (.57)	650 female, 328 male, 18 other; age: *M* = 30.9 (*SD* = 9.1), range: 18–75
9	Chicken Game	1100	656 (.60)	696 female, 378 male, 23 other; age: *M* = 31.7 (*SD* = 10), range: 18–81
10	Stag Hunt Game	1061	743 (.70)	596 female, 446 male, 18 other; age: *M* = 32 (*SD* = 9.9), range: 18–73

use of their data for scientific purposes after completing the measure of D. All exclusion criteria were set prior to running the first study and held constant across all studies (for details, see https://osf.io/93tw6/). After completing the D70 (but before receiving feedback on their level of D), participants were asked whether they would be willing to complete a few additional, unrelated questions voluntarily and without compensation (all were assured that they would receive feedback on D independent of whether they chose to answer the additional questions). Those who agreed, completed the SD4 and one economic game as specified below.

To determine the required sample size, a simulation was run for the crucial hypothesis that single dark tetrad traits should no longer be linked to prosocial choices once controlling for D. Specifically, mirroring the planned analytical procedure, the simulation implemented a logistic regression with a binary outcome (game choice) and two standardized predictors (D and one dark tetrad trait). Assuming a prevalence of prosociality of .60, a strong effect of D (OR = .40), and a correlation between D and any single dark tetrad trait of $r = .70$ (Hilbig et al., 2024), the simulation determined for different sample sizes the expected 95% CI width of the estimated OR of the dark tetrad trait (hypothesized to have no effect, i.e. OR = 1). It revealed that with $N > 800$ the expected width of the 95% CIs is sufficiently small to indicate when there is no relevant effect (i.e., with $N = 800$ the average CI for the null effect was [0.81, 1.25] and thus just within the thresholds of OR = 0.80 and OR = 1.25 which approximately correspond to Cohen's $d = .20$ and represent a small effect (Heck et al., 2018)). Given that 800 is the required minimum and the simulation is based on assumptions that may not turn out entirely accurate[5],

[5] Post hoc, the assumptions turned out relatively accurate as the ultimately observed prevalence of prosocial choices across games was .68 (see Table 2), the average effect of D was $M_{\text{OR}} = 0.46$ (see results section), and the average correlation between D

https://doi.org/10.1017/jdm.2025.1 Published online by Cambridge University Press

we conservatively set the target sample size to $N = 1000$. The simulation code is available on the OSF (https://osf.io/93tw6/).

2.1. Study 1: Dictator Game

The Dictator Game (Forsythe et al., 1994; Kahneman et al., 1986) is a sequential, zero-sum game that allows for the expression of (non-)exploitation (Thielmann et al., 2021). Participants were asked to imagine that a monetary endowment of 40€ had been randomly distributed between themselves and person X, resulting in an uneven distribution of 30€:10€ to their own advantage. They were given the power to decide on the final distribution and asked to choose between taking 10€, taking 5€, taking/giving nothing, or giving any amount between 5€ and 30€ (in steps of 5€). The resulting choice distribution was clearly bimodal and indeed almost perfectly binary as a full 87% of participants chose one of only two options: 38% chose to give/take nothing and 49% chose to give 10€. Thus, we coded anyone giving any nonzero amount as choosing the prosocial choice option (coded 1, otherwise 0). To ensure results were not due to us transforming the continuous responses into a binary variable, Study 2 implemented a binary version of the same game.

2.2. Study 2: Dictator Game (binary)

Instructions and the initial distribution of the endowment were the same as in Study 1, but participants were given the choice to either leave the (unfair) distribution unaltered or give 10€ to the other (thereby producing a fair distribution) which is the prosocial choice option (coded 1, otherwise 0).

2.3. Study 3: Ultimatum Game (proposer)

The Ultimatum Game (Güth et al., 1982) is a sequential (bargaining) game. For the first mover, the proposer, it is a game of imperfect information which allows for the expression of (non-)exploitation and (dis)trust (Thielmann et al., 2021). Participants were asked to imagine they had received a monetary endowment of 40€ and asked to propose how to distribute the endowment between themselves and the other who was then going to have the choice of either accepting the proposed distribution (in which case it would be (hypothetically) realized as proposed) or rejecting it (in which case neither player would receive anything, i.e., the endowment would be lost). Participants chose between two options, namely a distribution ratio of 80:20 (32€ vs. 8€) in their own favor or an even distribution (20€ each) which is the prosocial choice option (coded 1, otherwise 0).

2.4. Study 4: Ultimatum Game (responder)

The game was essentially the same as in Study 3, but participants were in the role of the responder reacting to an unfair proposal in which case the game allows for the expression of forgiveness vs. retaliation (Thielmann et al., 2021). Specifically, participants were asked to imagine the other—knowing the entire rules of the game—had proposed to split the endowment unevenly (80:20, i.e. 32€ vs. 8€) in the other's favor. Participants were given the choice to either reject the proposal (all money lost) or accept it (split realized as proposed) which is the prosocial choice option (coded 1, otherwise 0).

2.5. Study 5: Uncostly Retaliation Game (responder)

The Uncostly Retaliation Game (Hilbig et al., 2016) is similar to the Ultimatum Game but unconfounds forgiveness (vs. retaliation) from greed. Specifically, so long as the proposed offer is greater than zero,

and any dark tetrad trait was $r = .61$ (see supplementary results on the OSF; https://osf.io/93tw6/). Correspondingly, a simulation with these actual numbers and the mean achieved sample size (1047) showed that the expected width of the 95% CIs is .85 – 1.19 and thus well within the thresholds of a small effect size.

https://doi.org/10.1017/jdm.2025.1 Published online by Cambridge University Press

a responder in the Ultimatum Game can only retaliate at some cost (losing whatever was offered); vice versa: forgiveness (accepting an unfair proposal) is also individual payoff maximizing. In the Uncostly Retaliation Game, by contrast, the responder's choice does not affect their own payoff but only the proposer's. Specifically, the responder can decide on the proposer's final payoff, i.e. by how much to reduce the proposer's share (the responder does not receive what is taken; it is lost). Thereby, retaliation is not costly to the responder. Participants were asked to imagine that the proposer had split the endowment of 40€ unevenly (75:25, i.e. 30€ vs. 10€) in the other's favor. They were then given the choice to reduce the proposer's payoff by 30€ (to zero), reduce it by 20€ (to 10€), reduce it by 10€ (to 20€), or leave it intact (at 30€). The resulting choice distribution was bimodal and 87% of participants chose one of only two options: 71% chose to leave it intact and another 16% chose to take 20€. We coded leaving the proposer's payoff intact as choosing the prosocial option (coded 1, otherwise 0). Again, as with the Dictator Game, an additional study implemented a binary version of the same game to ensure results were not due to us transforming the continuous responses into a binary variable.

2.6. Study 6: Uncostly Retaliation Game (responder) (binary)

The game was the same as in Study 5, but participants were only given two choice options: Allowing the proposer to keep nothing (reducing their payoff from 30€ to zero) or allowing them to keep everything (30€) which is the prosocial choice option (coded 1, otherwise 0).

2.7. Study 7: Faith Game

The Faith Game (Kiyonari et al., 2006) is akin to the Dictator Game but focuses on the recipient side. More specifically, the (potential) recipient chooses whether to play the Dictator Game (as recipient) or opt-out and take a pre-specified payoff which is typically lower than the fair split in the Dictator Game. The game thereby allows for the expression of (dis)trust (Thielmann et al., 2021). Participants were asked to imagine an endowment of 60€ given to the other who would decide how to distribute it. Participants were asked whether they would prefer either to opt-out and receive 20€ (independent of the dictator's chosen distribution) or play the Dictator Game as the recipient (and receive whatever the dictator chose to give) which is the prosocial choice option (coded 1, otherwise 0).

2.8. Study 8: Prisoner`s Dilemma

The Prisoner`s Dilemma (Kollock, 1998; Rapoport & Chammah, 1965) is a simultaneous game with two players and a social dilemma structure. The players independently choose between cooperation (C) and defection (D) and the combination of choices determines their (four possible) payoffs: reward R for mutual cooperation (C, C), punishment P for mutual defection (D, D), temptation T for unilateral defection (D, C), and sucker S for unilateral cooperation (C, D). In the standard (symmetric) Prisoner's Dilemma, $T > R > P > S$, thus leading to the social dilemma structure: For each individual, defection strictly dominates cooperation, but mutual cooperation maximizes the sum of payoffs (social welfare). The game allows for the expression of (non-)exploitation and (dis)trust (Thielmann et al., 2021). The game and its four possible outcomes were explained to participants with T = 20€ > R = 10€ > P = 5€ > S = 0€, and they were asked to choose between defection and cooperation (both labeled neutrally as options A and B), the latter of which is the prosocial choice option (coded 1, otherwise 0).

2.9. Study 9: Chicken Game

The Chicken Game (de Heus et al., 2010; Rapoport & Chammah, 1969) is highly similar to the Prisoner's Dilemma; however, it is not a social dilemma, but a coordination game because the rank order of payoffs S and P is reversed (i.e. $T > R > S > P$) and thus unilateral cooperation is no longer the worst outcome (instead, mutual defection is). Thus, as compared to the Prisoner's Dilemma, the prosocial

https://doi.org/10.1017/jdm.2025.1 Published online by Cambridge University Press

choice (C) no longer affords trust but still affords nonexploitation (Hilbig et al., 2018). Participants received the same instructions and choice options as in the Prisoner's Dilemma, but with T = 20€ > R = 10€ > S = 5€ > P = 0€. Again, cooperation is the prosocial choice option (coded 1, otherwise 0).

2.10. Study 10: Stag Hunt Game

The Stag Hunt or Assurance Game (Skyrms, 2001) is another coordination game that essentially alters the payoff structure of the Prisoner's Dilemma. Specifically, the rank order of payoffs T and R is reversed (i.e. R > T ≥ P > S), and thus unilateral defection is no longer the best outcome (instead, mutual cooperation is). Thus, as compared to the Prisoner`s Dilemma, the prosocial choice (C) no longer affords nonexploitation but affords trust (Hilbig et al., 2018). In other words, whereas the Chicken Game isolates (non-)exploitation, the Stag Hunt isolates (dis)trust. Participants again received the same instructions and choice options as in the Prisoner's Dilemma but with R = 20€ > T = 10€ ≥ P = 10€ > S = 0€. Again, cooperation is the prosocial choice option (coded 1, otherwise 0).

3. Results

Descriptive statistics and internal consistencies of all trait scales (D, Narcissism, Machiavellianism, Psychopathy, and Sadism) as well as their intercorrelations per study can be found in the supplementary results on the OSF (https://osf.io/93tw6/). All studies were analyzed with the exact same procedure (and code, which is also available on the OSF; https://osf.io/93tw6/). All trait scales were standardized prior to the main analyses. First, to test the hypothesis that D would be negatively associated with prosocial choices in economic games, a logistic regression was run for each study with game choice (coded 1 for prosocial, otherwise 0) as the criterion and D as the predictor. The results are summarized in Figure 1, showing that D was negatively associated with prosocial choices in each game (0.32 ≤ OR ≤ 0.71; all 95% CI upper bounds < 1) with a large effect size on average (M_{OR} = 0.46).

Next, to test whether single dark tetrad traits are negatively associated with prosocial choices in economic games, a logistic regression was run for each study with game choice (coded 1 for prosocial, otherwise 0) as the criterion and one of the four dark tetrad traits (Narcissism, Machiavellianism, Psychopathy, or Sadism) as the predictor. As can be seen from the corresponding odds ratios (green circles) displayed in Figure 2, the expected negative association held for every trait in each game with only a single exception: Narcissism did not predict prosocial choices in the Stag Hunt Game (OR = 0.94, 95%CI: 0.83, 1.08]. For all other odds ratios (0.37 ≤ OR ≤ 0.87), the 95% CI upper bounds were below OR = 1. On average, single dark tetrad traits yielded a medium-sized effect (M_{OR} = 0.63). The exact numbers per trait and game can be found in the supplementary results on the OSF (https://osf.io/93tw6/).

Finally, to test whether single aversive traits are (no longer) negatively associated with prosocial choices in economic games when controlling for D, we ran multiple logistic regressions per study with game choice (coded 1 for prosocial, otherwise 0) as the criterion and D *and* one of the four dark tetrad traits (Narcissism, Machiavellianism, Psychopathy, or Sadism) as predictors. As can again be seen in Figure 2 (orange squares), in four (out of 40) cases, single dark tetrad traits were still negatively associated with prosocial choices once controlling for D: Machiavellianism had a small incremental effect over D in the Faith Game, the Prisoner's Dilemma, and the Chicken Game; also, Sadism had a small incremental effect over D in the uncostly retaliation game. In all other (36 out of 40) cases, single dark tetrad traits were no longer negatively associated with prosocial choices once controlling for D, and indeed, the average incremental effect of these traits over D was an almost perfect null effect (M_{OR} = 1.03). By contrast, D was negatively associated with prosocial choices beyond all single dark tetrad traits in every game (blue triangles in Figure 2; 0.28 ≤ OR ≤ 0.78; all 95% CI upper bounds < 1), and indeed, the average incremental effect of D was exactly as large as its average zero-order effect (M_{OR} = 0.46). In summary, the negative associations between single aversive traits and prosocial choices were

https://doi.org/10.1017/jdm.2025.1 Published online by Cambridge University Press

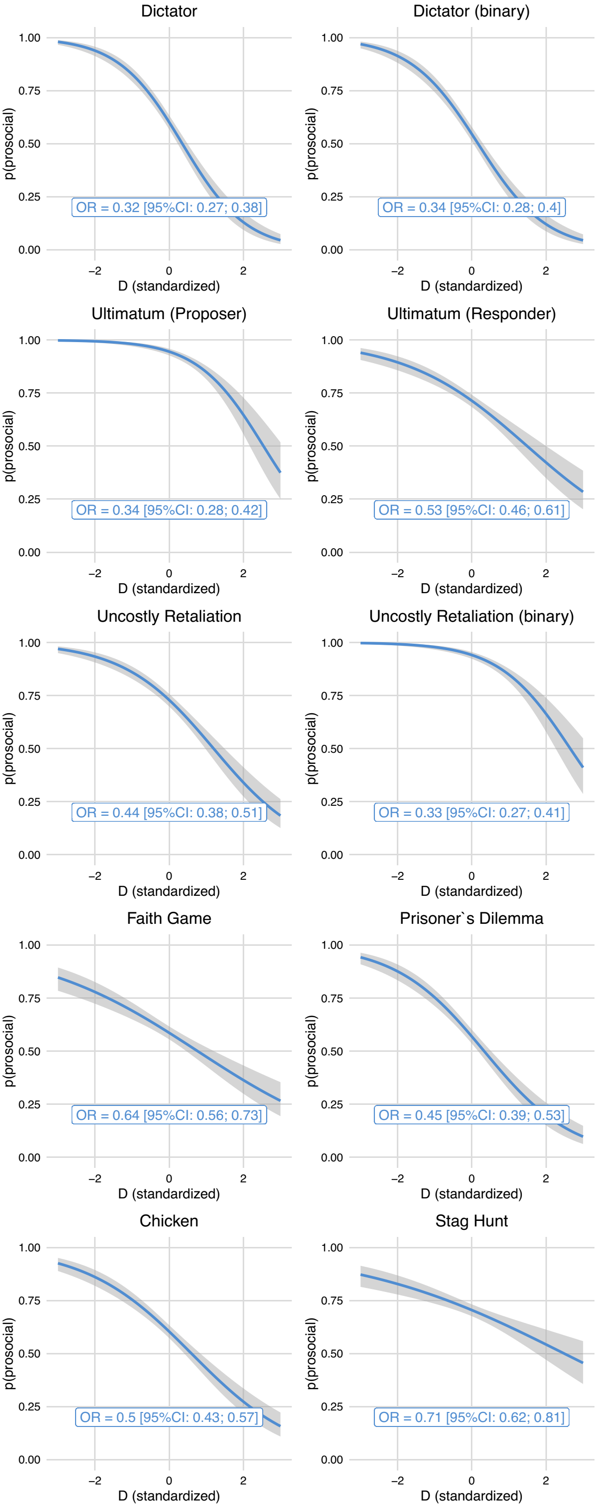

Figure 1. *Association (smoothed; GLM with a binomial link function) between D and the probability of prosocial choices per game. The shaded area represents the standard error. The reported odds ratios (and 95% CIs) are the results from the logistic regression (game choices on D) as described in the main text.*

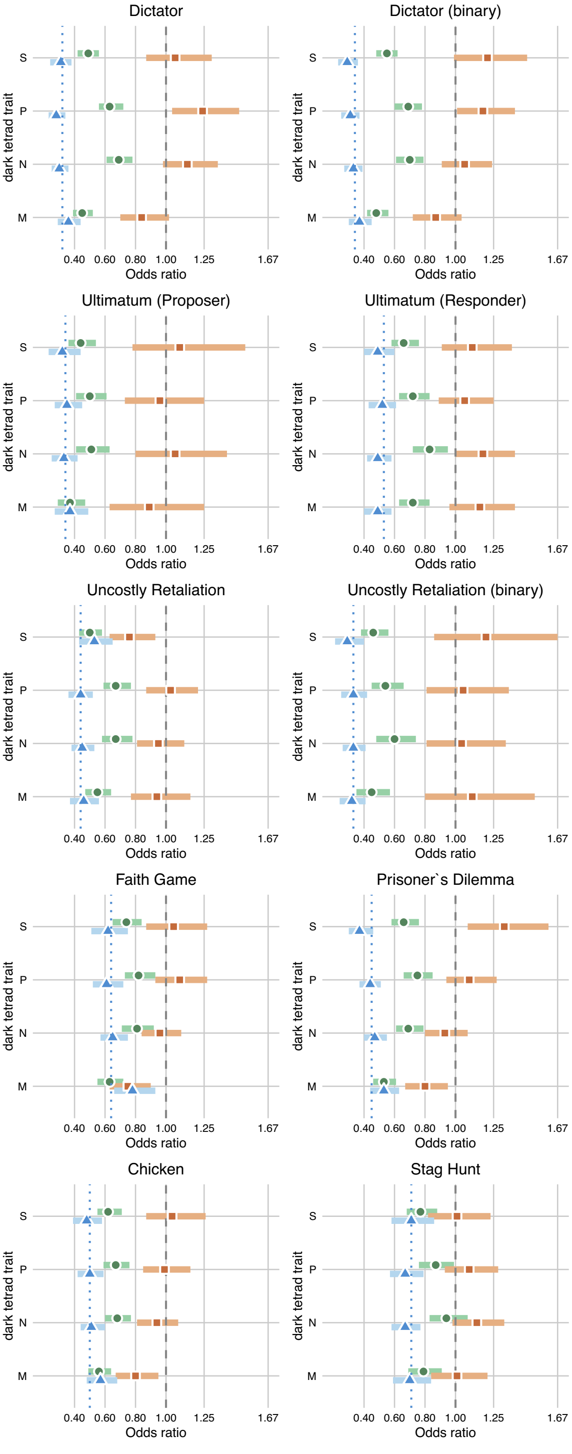

Figure 2. *Odds ratios (bars represent the 95% CIs) for single dark tetrad traits predicting prosocial choices alone (green circles), for single dark tetrad traits predicting prosocial choices when controlling for D (orange squares), and for D predicting prosocial choices controlling for each single dark tetrad trait (blue triangles). The dotted (blue) vertical line indicates the zero-order effect of D (see Figure 1). The dashed vertical line indicates a null effect (OR = 1). The remaining vertical lines indicate a small (OR = 0.80; OR = 1.25), medium-sized (OR = 0.60; OR = 1.67), or large (OR = .40) effect (Heck et al., 2018). N = Narcissism, M = Machiavellianism, P = Psychopathy, S = Sadism.*

https://doi.org/10.1017/jdm.2025.1 Published online by Cambridge University Press

almost entirely accounted for by D, whereas the reverse was never once the case. The exact numbers can be found in the supplementary results on the OSF (https://osf.io/93tw6/).

4. Discussion

In light of substantial individual differences in pro- vs. antisocial behavior as studied via economic games, several personality traits have been considered as potential explanations. Prominent among these are dark traits, especially the 'dark tetrad' traits of Machiavellianism, Narcissism, Psychopathy, and Sadism (Paulhus et al., 2020). Theoretically, such traits are obvious candidates to account for choices in games that allow for the expression of (non)exploitation, forgiveness (vs. retaliation), and/or (dis)trust; across studies, however, their effects are relatively small on average (and often enough indistinguishable from zero) and notably inconsistent (Thielmann et al., 2020). Most problematically, the empirical picture is largely devoid of specificity (linking certain dark traits to certain games or, by implication, affordances), let alone parsimony.

We herein argued that this unsatisfactory state of knowledge may arise from two limitations of prior work: first, the games commonly studied in prior research on individual differences typically afford the expression of the same dispositional tendency, namely (non-)exploitation, but rarely afford the expression of forgiveness (vs. retaliation) and/or (dis)trust alone, that is, unconfounded from aspects of (non-)exploitation. Second, prior work insufficiently accounted for the large overlap of dark traits, that is, their common core vis-à-vis their unique aspects. Specifically, whereas all dark traits share a single aversive core—the dark factor of personality (D; Moshagen et al., 2018)—they often entail other, nonaversive aspects and thus 'dilute' the measurement of their shared 'aversive essence', resulting in reduced and inconsistent effects on many instances of unethical and/or aversive behavior (Hilbig et al., 2024; Scholz et al., 2023). In turn, since dark traits can all be understood as manifestations of D, their associations with aversive behavior ought to be largely due to D (Hilbig et al., 2023).

In line with our predictions derived from the conceptualization and framework of D, we found—across 10 well-powered studies and eight economic games which, taken together, provide coverage/isolation of all theoretically relevant affordances—that (i) D is consistently negatively associated with prosocial behavior in all games yielding large effect sizes on average, (ii) dark traits are also negatively linked to prosocial behavior, yielding medium-sized effects and, crucially, (iii) controlling for D largely eliminates (i.e., accounts for) almost all of these negative effects of dark traits whereas the opposite (any dark trait accounting for the effect of D) was never once the case. The latter issue—i.e., whether dark traits predict incremental variance beyond D—can be considered the most critical test of the D framework. In turn, particularly strong counterevidence to this framework would be that single dark traits do explain relevant variance beyond D—optimally in a systematic/specific pattern. We welcome further critical tests of this nature, which do not need to be limited to behavior in economic games. Essentially, the D framework makes the same strong prediction for *any* outcome that primarily pits socially/ethically aversive behavior against prosocial/ethical behavior (e.g., (dis)honesty, Hilbig et al., in press).

In terms of the zero-order order effects, it is noteworthy that effect sizes were substantial even in games in which prosocial responses are predominant and thus there is a limited variation to explain. Such limited variance is common for coordination games like the Stag Hunt due to its nature of aligned interests (Balliet et al., 2022), the Ultimatum Game as proposer due to the strategic necessity of fairness to avoid retaliation (Güth & Tietz, 1990; Suleiman, 1996), and the Uncostly Retaliation Game as responder due to the competitive or even sadistic nature of retaliation that goes beyond what is mandated by a fairness/equality norm (Herrmann et al., 2008). Arguably, our relatively large sample sizes were particularly useful in this regard, since they ensure a sufficiently large number of observations per category even with highly skewed choice proportions.

The incremental effects, too, were in line with theoretical considerations and prior findings. In particular, although cases in which any dark tetrad trait accounted for incremental variance beyond

https://doi.org/10.1017/jdm.2025.1 Published online by Cambridge University Press

D were so few that one might attribute them to chance, they were not completely unsystematic. Three out of four of these cases were due to Machiavellianism and two out of these occurred in games that afford the expression of (dis)trust: the Faith Game and the Prisoner's Dilemma. This is well-aligned with prior arguments and findings that some variants of Machiavellianism may yield an excess of cynicism/distrust beyond D (Bader et al., 2023). Then again, the associated incremental effect sizes were relatively small (ORs ~ .80) and the absence of an incremental effect of Machiavellianism beyond D in the Stag Hunt Game, which also taps into (dis)trust (and indeed exclusively so), imply that this excess of cynicism/distrust beyond D is limited in scope. Moreover, it must be noted that D predicted incremental variance beyond Machiavellianism in every game affording the expression of (dis)trust (ORs ~ .60), thus suggesting that Machiavellianism does not cover all aspects of (dis)trust driving pro- vs. antisocial choices in such games.

More generally, the pattern of incremental effects adds to a growing body of literature demonstrating that most if not all of the variance that dark traits explain in ethically/socially aversive behavior is due to their single common core, D, by not vice versa (Hilbig et al., in press; Hilbig et al., 2023; Moshagen et al., 2018; Scholz et al., 2023). In other words, the aversive features of dark traits are subsumed by D and any remaining features these traits comprise beyond D (e.g. admiration in Narcisissim or boldness in Psychopathy; Bader et al., 2023) do not generally add to the explanation of aversive behavior; in fact, they may well hamper it because they 'dilute' D. Accumulating evidence thus highlights the inherent drawbacks of proposing an ever-growing number of allegedly distinct, narrow aversive traits and especially of studying these in small sets or even in isolation. When seeking to explain aversive behavior in general, maximizing theoretical parsimony clearly implies to start with D and to consider more specific aversive traits if and only if these clearly account for incremental variance (Hilbig et al., 2024). We would not claim this cannot occur—on the contrary, the D framework clearly mandates it should occur, namely whenever the unique, nonaversive aspects of a specific aversive trait (e.g., impulsivity in Psychopathy) are also relevant to the behavior in question (e.g., reckless driving; Bader et al., 2023). However, we maintain that the burden of evidence should always be on those seeking to add theoretical complexity (Leising et al., 2022; Popper, 1959). In the realm of ethically/socially aversive behavior, placing the bar at explaining variance beyond D is decidedly not 'to err on the side of parsimony' (Paulhus et al., 2020, p. 217), but to mitigate further construct inflation or, more specifically, dark trait sprawl.

Beyond further clarifying the relationship between narrow, specific aversive traits and their common core, D, the present findings also shed light on the distinctions between D and other broad traits relevant to pro- vs. antisocial behavior, such as HEXACO Honesty–Humility (Ashton et al., 2014) or trait-like concepts such as Social Value Orientation (Murphy & Ackermann, 2014; Van Lange, 1999). Unlike D, neither of these conceptually involve aspects of retaliation vs. forgiveness and their associations with behavior in games affording this tendency are indeed negligible (Hilbig et al., 2016; Thielmann et al., 2020). On the conceptual level, the same holds for (dis)trust, but the empirical picture is more mixed (Hilbig et al., 2018; Thielmann & Hilbig, 2014; Thielmann et al., 2020). In any case, our findings confirm that D—in line with its definition—is indeed linked to behavior in games that afford the expression of both retaliation vs. forgiveness and distrust. In addition, D also differs from Honesty–Humility in that it accounts not only for in-group cooperation vs exploitation but also for out-group harm (Columbus et al., 2024). As such, the findings add to a growing body of studies that provide evidence for the distinctness of D from other broad pro- vs. antisocial traits (Horsten et al., 2021; Scholz et al., 2022).

4.1. Limitations and future directions

As must be expected, the present research is not without limitations. First, the reported associations between personality traits and game behavior (and arguably the overall choice proportions) must be interpreted carefully due to the purely hypothetical design. Not only is it plausible that the lack of incentives shifts observed choices toward the socially desirable response options (Hertwig & Ortmann, 2001; Lanz et al., 2021; Thielmann et al., 2016), but there is actually meta-analytical evidence suggesting that the effects of dark traits on game behavior are larger in hypothetical as compared in incentivized, consequential games (Thielmann et al., 2020). This may also explain the discrepancy between our findings and prior evidence in that we found significant asssociations with prosocial behavior for all dark traits in (almost) all games. Moreover, larger effect sizes in hypothetical (as compared to incentivized) games have also been reported for D (Moshagen et al., 2018). Importantly, however, to the extent that the hypothetical setup inflated zero-order effects, it actually rendered the test of our crucial hypothesis—that dark traits do not account for variance beyond D—more conservative: The larger the zero-order effects, the larger the possible incremental effects (and vice versa: without a zero-order effect there cannot be an actual incremental effect); so, if anything, the hypothetical setup maximized the chances of dark traits to account for incremental variance[6]. Also, highly similar results (little to no incremental effect of dark traits over D) have already been reported for fully consequential behavior (Hilbig et al., in press; Hilbig et al., 2023; Moshagen et al., 2018) and are thus not limited to hypothetical situations. Nonetheless, given the lack of corresponding evidence especially for the less commonly studied games that we included herein (i.e., those affording the expression of forgiveness (vs. retaliation) and/or (dis)trust), future replications with incentives would undoubtedly add further confidence.

Moreover, even though we sought to study a broad array of games, still other games or related measures might complement the picture. For example, (dis)trust can also be measured by asking participants about their beliefs (what other players will do) in social dilemmas. Whereas D has also been shown to be associated with (dis)trust measured in such a way (Hilbig et al., 2022), this has not yet been tested vis-a-vis single dark traits. Moreover, game paradigms that allow involved parties to take away resources from other involved parties might provide additional insights. For example, the Moonlighting Game (Abbink et al., 2000) extends the classic Trust Game (Berg et al., 1995) by adding a taking option to both the trustor's and the trustee's choice options. Thereby, the Moonlighting Game taps into positive *and* negative reciprocity among trustees because the trustee can punish the trustor for distrustful behavior. One might speculate, for example, that individuals high in D—despite being distrustful themselves—expect/demand trust from others and punish those who do not trust them.

Finally, given that we recruited participants from a website open to the general public and visited by people interested in self-assessments on aversive personality, some selection effects are likely present in our samples. Whether they are more likely or indeed more prone to bias results than selection effects inherent in other typical approaches to recruiting participants[7] is difficult to judge and remains to be tested. Relatedly, our samples are not fully population representative; in particular, participants were younger than the population average. That said, we are confident that, if anything, our samples are more heterogeneous and more population-representative than many samples in the field, especially student samples. For example, the sample size weighted mean age in all unique studies in the Thielmann et al. meta-analysis on personality and economic games (n = 561 unique studies) is 29.4, which is younger and thus still less population representative than the mean age of the samples we report on herein. Nonetheless, still more randomly selected and more population-representative samples than ours would certainly constitute an additional improvement.

[6] To be sure, we ran a simulation varying the size of the zero-order effects of both predictors to be small (r = .10) vs. medium-sized (r = .30) on average (holding the correlation between predictors constant at r = .60). With small zero-order effects, the average incremental variance explained by either predictor was 0.3% (practically zero) whereas with medium-sized zero-order effects, it was 2.3% and thus almost an order of magnitude larger. The simulation code can be found on the OSF (https://osf.io/93tw6/).

[7] Clearly, individuals also self-select into the two most common participant sources in the field: crowdsource platforms (i.e., clickworkers) and higher education institutions (i.e., students). We are not aware of direct evidence whether selection by income (clickworkers), ability (students), or age (both)—as compared to interest in a free, anonymous self-assessment (our studies)—is any more or less likely to bias results.

https://doi.org/10.1017/jdm.2025.1 Published online by Cambridge University Press

5. Conclusions

Overall, the present research highlights two limitations characterizing prior studies on personality and pro- vs. antisocial behavior in economic games. Only a small subset of games is commonly studied, thus leading to a severe underrepresentation of certain situational affordances, and aversive traits are treated as distinct constructs while ignoring their overlap (which is what actually links them to pro- vs antisocial behavior). The latter problem is exacerbated by the sheer number of socially aversive traits (at present, we know of well over 20 traits that can be considered manifestations of D; Scholz et al., 2024) and the widespread practice of studying them in isolation, in triplets, or quadruplets at most. Whereas these two limitations are independent in principle, their effects are compounded when seeking to derive a general state of knowledge, e.g. in meta-analyses (Thielmann et al., 2020). Such efforts have allowed for little more than the near truism that some dark traits sometimes relate to behavior in some games—a statement almost entirely devoid of empirical content (Glöckner & Betsch, 2011). Instead, we offer and provide evidence for the conclusion that the single, common core of all aversive traits (D) predicts pro- vs. antisocial behavior in economic games that afford the expression of (non)exploitativeness, forgiveness (vs. retaliation), and/or (dis)trust.

Data availability statement. Materials, data, analysis code, and supplementary results are available on the Open Science Framework (https://osf.io/5vjws/)

Funding statement. This work was supported by a grant from the German Research Foundation to the second author (TH 2318/3-1).

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abbink, K., Irlenbusch, B., & Renner, E. (2000). The moonlighting game: An experimental study on reciprocity and retribution. *Journal of Economic Behavior & Organization*, *42*(2), 265–277. https://doi.org/10.1016/S0167-2681(00)00089-5

Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, *18*, 139–152. https://doi.org/10.1177/1088868314523838

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396. https://doi.org/10.1126/science.7466396

Back, M. D., Küfner, A. C. P., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. A. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology*, *105*(6), 1013–1037. https://doi.org/10.1037/a0034431

Bader, M., Hilbig, B. E., Zettler, I., & Moshagen, M. (2023). Rethinking aversive personality: Decomposing the Dark Triad traits into their common core and unique flavors. *Journal of Personality*, *91*, 1084–1109. https://doi.org/10.1111/jopy.12785

Bader, M., Horsten, L. K., Hilbig, B. E., Zettler, I., & Moshagen, M. (2022). Measuring the dark core of Personality in German: Psychometric properties, measurement invariance, and self-other agreement. *Journal of Personality Assessment*, *104*, 660–673. https://doi.org/10.1080/00223891.2021.1984931

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

Balliet, D., Molho, C., Columbus, S., & Dores Cruz, T. D. (2022). Prosocial and punishment behaviors in everyday life. *Curr Opin Psychol*, *43*, 278–283. https://doi.org/10.1016/j.copsyc.2021.08.015

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, *137*(4), 594–615. https://doi.org/10.1037/a0023489

Balliet, D., Parks, C., & Joireman, J. (2009). Social Value Orientation and cooperation in social dilemmas: A Meta-Analysis. *Group Processes Intergroup Relations*, *12*(4), 533–547. https://doi.org/10.1177/1368430209105040

Balliet, D., & Van Lange, P. A. M. (2012). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*. https://doi.org/10.1037/a0030939

Bereczkei, T., Papp, P., Kincses, P., Bodrogi, B., Perlaki, G., Orsi, G., & Deak, A. (2015). The neural basis of the Machiavellians' decision making in fair and unfair situations. *Brain and Cognition*, *98*, 53–64. https://doi.org/10.1016/j.bandc.2015.05.006

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142. https://doi.org/10.1006/game.1995.1027

Blötner, C., Ziegler, M., Wehner, C., Back, M. D., & Grosz, M. P. (2022). The nomological network of the Short Dark Tetrad Scale (SD4). *European Journal of Psychological Assessment*, *38*(3), 187–197. https://doi.org/10.1027/1015-5759/a000655

Böckler, A., Sharifi, M., Kanske, P., Dziobek, I., & Singer, T. (2017). Social decision making in narcissism: Reduced generosity and increased retaliation are driven by alterations in perspective-taking and anger. *Personality and Individual Differences*, *104*, 1–7. https://doi.org/10.1016/j.paid.2016.07.020

Bonfá-Araujo, B., Lima-Costa, A. R., Hauck-Filho, N., & Jonason, P. K. (2022). Considering sadism in the shadow of the Dark Triad traits: A meta-analytic review of the Dark Tetrad. *Personality and Individual Differences*, *197*. https://doi.org/10.1016/j.paid.2022.111767

Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral & Brain Sciences*, *26*, 139–198.

Columbus, S., Molho, C., Righetti, F., & Balliet, D. (2021). Interdependence and cooperation in daily life. *Journal of Personality and Social Psychology*, *120*, 626–650. https://doi.org/10.1037/pspi0000253

Columbus, S., Thielmann, I., Böhm, R., & Zettler, I. (2024). Personality Correlates of Out-Group Harm. *Social Psychological and Personality Science*. https://doi.org/10.1177/19485506241254157

Dawes, R. M., & Messick, D. M. (2000). Social dilemmas. *International Journal of Psychology*, *35*(2), 111–116. https://doi.org/10.1080/002075900399402

de Heus, P., Hoogervorst, N., & Dijk, E. v. (2010). Framing prisoners and chickens: Valence effects in the prisoner's dilemma and the chicken game. *Journal of Experimental Social Psychology*, *46*(5), 736–742. http://www.sciencedirect.com/science/article/B6WJB-500SK3F-1/2/3ae7682f2a47d50ad9817b898ff76c56

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*, 583–610. https://doi.org/10.1007/s10683-011-9283-7

Fatfouta, R., Rentzsch, K., & Schröder-Abé, M. (2018). Narcissus oeconomicus: Facets of narcissism and socio-economic decision-making. *Journal of Research in Personality*, *75*, 12–16. https://doi.org/10.1016/j.jrp.2018.05.002

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*(3), 347–369. https://doi.org/10.1006/game.1994.1021

Glöckner, A., & Betsch, T. (2011). The empirical content of theories in judgment and decision making: Shortcomings and remedies. *Judgment and Decision Making*, *6*(8), 711.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–388. https://doi.org/10.1016/0167-2681(82)90011-7

Güth, W., & Tietz, R. (1990). Ultimatum bargaining behavior: A survey and comparison of experimental results. *Journal of Economic Psychology*, *11*(3), 417–449. https://doi.org/10.1016/0167-4870%2890%2990021-Z

Heck, D. W., Thielmann, I., Moshagen, M., & Hilbig, B. E. (2018). Who lies? A large-scale reanalysis linking basic personality traits to unethical decision making. *Judgment and Decision Making*, *13*(4), 356–371s.

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*(5868), 1362–1367.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A challenge for psychologists? *Behavioral and Brain Sciences*, *24*(3), 383–403. https://doi.org/10.1037/e683322011-032

Hilbig, B. E., Kieslich, P. J., Henninger, F., Thielmann, I., & Zettler, I. (2018). Lead Us (Not) into Temptation: Testing the Motivational Mechanisms Linking Honesty–Humility to Cooperation. *European Journal of Personality*, *32*(2), 116–127. https://doi.org/10.1002/per.2149

Hilbig, B. E., Moshagen, M., Thielmann, I., & Zettler, I. (2022). Making rights from wrongs: The crucial role of beliefs and justifications for the expression of aversive personality. *Journal of Experimental Psychology: General*, *151*, 2730–2755. https://doi.org/10.1037/xge0001232

Hilbig, B. E., Thielmann, I., & Heck, D. W. (in press). Filling in the missing pieces: Personality traits (un)related to dishonest behavior. *European Journal of Personality*. https://doi.org/10.1177/08902070241293621

Hilbig, B. E., Thielmann, I., Klein, S. A., & Henninger, F. (2016). The two faces of cooperation: On the unique role of HEXACO Agreeableness for forgiveness versus retaliation. *Journal of Research in Personality*, *64*, 69–78. https://doi.org/10.1016/j.jrp.2016.08.004

Hilbig, B. E., Thielmann, I., Zettler, I., & Moshagen, M. (2023). The Dispositional Essence of Proactive Social Preferences: The Dark Core of Personality vis-a-vis 58 Traits. *Psychological Science*, *34*(2), 201–220. https://doi.org/10.1177/09567976221116893

Hilbig, B. E., Zettler, I., & Moshagen, M. (2024). A little parsimony goes a long way: Aversive ('dark') personality and pro-environmentalism. *Journal of Environmental Psychology*, *96*. https://doi.org/10.1016/j.jenvp.2024.102291

Horsten, L. K., Moshagen, M., Zettler, I., & Hilbig, B. E. (2021). Theoretical and empirical dissociations between the Dark Factor of Personality and low Honesty-Humility. *Journal of Research in Personality*, *95*, 104154. https://doi.org/10.1016/j.jrp.2021.104154

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889. https://doi.org/10.1016/j.joep.2011.05.007

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the Assumptions of Economics. *The Journal of Business*, *59*(4), S285–S300. https://doi.org/10.2307/2352761

https://doi.org/10.1017/jdm.2025.1 Published online by Cambridge University Press

Kiyonari, T., & Yamagishi, T. (1999). A comparative study of trust and trustworthiness using the game of enthronement. *Japanese Society of Social Psychology*.

Kiyonari, T., Yamagishi, T., Cook, K. S., & Cheshire, C. (2006). Does Trust Beget Trustworthiness? Trust and Trustworthiness in Two Games and Two Cultures: A Research Note. *Social Psychology Quarterly*, *69*(3), 270–283. https://doi.org/10.1177/019027250606900304

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, *24*, 183–214. https://doi.org/10.1146/annurev.soc.24.1.183

Lanz, L., Thielmann, I., & Gerpott, F. H. (2021). Are social desirability scales desirable? A meta-analytic test of the validity of social desirability scales in the context of prosocial behavior. *Journal of Personality*, *90*(2), 203–221. https://doi.org/10.1111/jopy.12662

Leising, D., Thielmann, I., Glöckner, A., Gärtner, A., & Schönbrodt, F. (2022). Ten steps toward a better personality science – how quality may be rewarded more in research evaluation. *Personality Science*, *3*. https://doi.org/10.5964/ps.6029

Malesza, M. (2018). The effects of the Dark Triad traits in prisoner's dilemma game. *Current Psychology*, *39*(3), 1055–1062. https://doi.org/10.1007/s12144-018-9823-9

Malesza, M. (2020). Grandiose narcissism and vulnerable narcissism in prisoner's dilemma game. *Personality and Individual Differences*, *158*. https://doi.org/10.1016/j.paid.2020.109841

Malesza, M., & Kalinowski, K. (2019). Willingness to share, impulsivity and the Dark Triad traits. *Current Psychology*, *40*(8), 3888–3896. https://doi.org/10.1007/s12144-019-00351-5

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

Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, *125*(5), 656. https://doi.org/10.1037/rev0000111

Moshagen, M., Zettler, I., & Hilbig, B. E. (2020). Measuring the dark core of personality. *Psychological Assessment*, *32*(2), 182–196. https://doi.org/10.1037/pas0000778

Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The Malevolent Side of Human Nature: A Meta-Analysis and Critical Review of the Literature on the Dark Triad (Narcissism, Machiavellianism, and Psychopathy). *Perspectives on Psychological Science*, *12*, 183–204. https://doi.org/10.1177/1745691616666070

Murphy, R. O., & Ackermann, K. A. (2014). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, *18*, 13–41. https://doi.org/10.1177/1088868313501745

Mussel, P., & Hewig, J. (2016). The life and times of individuals scoring high and low on dispositional greed. *Journal of Research in Personality*, *64*, 52–60. https://doi.org/10.1016/j.jrp.2016.07.002

Nehrlich, A. D., Gebauer, J. E., Sedikides, C., & Schoel, C. (2019). Agentic narcissism, communal narcissism, and prosociality. *Journal of Personality and Social Psychology*, *117*(1), 142–165. https://doi.org/10.1037/pspp0000190

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*, 1560–1563.

Paulhus, D. L., Buckels, E. E., Trapnell, P. D., & Jones, D. N. (2020). Screening for Dark Personalities: The Short Dark Tetrad (SD4). *European Journal of Psychological Assessment*, *37*(3), 208–222. https://doi.org/10.1027/1015-5759/a000602

Pfattheicher, S., Keller, J., & Knezevic, G. (2017). Sadism, the intuitive system, and antisocial punishment in the Public Goods Game. *Personality and Social Psychology Bulletin*, *43*(3), 337–346. https://doi.org/10.1177/0146167216684134

Popov, N., & Thielmann, I. (in press). The core tendencies underlying prosocial behavior: Testing a person-situation framework. *Journal of Personality*. https://doi.org/10.1111/jopy.12957

Popper, K. R. (1934/2005). *Logik der Forschung* (11th ed.). Mohr Siebeck.

Popper, K. R. (1959). *The logic of scientific discovery*. Basic Books.

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

Rapoport, A., & Chammah, A. (1969). The Game of Chicken. In *Game theory in the behavioral sciences* (pp. 151–175). University of Pittsburgh Press.

Sally, D. (1995). Conversation and cooperation in social dilemmas. A meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, *7*, 58–92. https://doi.org/10.1177/1043463195007001004

Scholz, D. D., Hilbig, B. E., Thielmann, I., Moshagen, M., & Zettler, I. (2022). Beyond (low) Agreeableness: Toward a more comprehensive understanding of antagonistic psychopathology. *Journal of Personality*, *90*(6), 956–970. https://doi.org/10.1111/jopy.12708

Scholz, D. D., Thielmann, I., & Hilbig, B. E. (2023). Down to the core: The role of the common core of dark traits for aversive relationship behaviors. *Personality and Individual Differences*, *213*. https://doi.org/10.1016/j.paid.2023.112263

Scholz, D. D., Zimmermann, J., Moshagen, M., Zettler, I., & Hilbig, B. E. (2024). Theoretical and empirical integration of 'dark' traits and socially aversive personality psychopathology. *Journal of Personality Disorders*, *38*, 241–267. https://doi.org/10.1521/pedi.2024.38.3.241s

Schreiber, A., & Marcus, B. (2020). The place of the "Dark Triad" in general models of personality: Some meta-analytic clarification. *Psychological Bulletin*, *146*(11), 1021–1041. https://doi.org/10.1037/bul0000299

Skyrms, B. (2001). The stag hunt. In *Proceedings and Addresses of the American Philosophical Association* (Vol. *75*, pp. 31–41). https://doi.org/10.2307/3218711

https://doi.org/10.1017/jdm.2025.1 Published online by Cambridge University Press

Spadaro, G., Graf, C., Jin, S., Arai, S., Inoue, Y., Lieberman, E., Rinderu, M. I., Yuan, M., Van Lissa, C. J., & Balliet, D. (2022). Cross-cultural variation in cooperation: A meta-analysis. *Journal of Personality and Social Psychology*, *123*(5), 1024–1088. https://doi.org/10.1037/pspi0000389

Suleiman, R. (1996). Expectations and fairness in a modified Ultimatum game. *Journal of Economic Psychology*, *17*(5), 531–554. https://doi.org/10.1016/S0167-4870%2896%2900029-3

Szijjártó, L., Kocsor, F., & Bereczkei, T. (2018). Machiavellian individuals' reciprocation tend to be smaller in a trust game. *Human Ethology Bulletin*, *33*, 39–48.

Thielmann, I., Böhm, R., Ott, M., & Hilbig, B. E. (2021). Economic games: An introduction and guide for research. *Collabra: Psychology*(7), 19004. https://doi.org/10.1525/collabra.19004

Thielmann, I., Heck, D. W., & Hilbig, B. E. (2016). Anonymity and incentives: An investigation of techniques to reduce socially desirable responding in the Trust Game. *Judgment and Decision Making*, *11*(5), 527.

Thielmann, I., & Hilbig, B. E. (2014). Trust in me, trust in you: A social projection account of the link between personality, cooperativeness, and trustworthiness expectations. *Journal of Research in Personality*, *50*, 61–65.

Thielmann, I., Hilbig, B. E., & Zettler, I. (2022). The dispositional basis of human prosociality. *Current Opinion in Psychology*, *43*, 289–294. https://doi.org/10.1016/j.copsyc.2021.08.009

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and Prosocial Behavior: A Theoretical Framework and Meta-Analysis. *Psychological Bulletin*, *146*(1), 30–90. https://doi.org/10.1037/bul0000217

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

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, *77*(2), 337–349. https://doi.org/10.1037/0022-3514.77.2.337

Van Lange, P. A. M., & Rand, D. G. (2022). Human cooperation and the crises of climate change, COVID-19, and misinformation. *Annual Review of Psychology*, *73*, 379–402. https://doi.org/10.1146/annurev-psych-020821-110044

Vize, C. E., Lynam, D. R., Collison, K. L., & Miller, J. D. (2018). Differences among dark triad components: A meta-analytic investigation. *Personality Disorders: Theory, Research, and Treatment*, *9*(2), 101. https://doi.org/10.1037/per0000222

Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics*, *6*(3), 299–310. https://doi.org/10.1023/A:1026277420119

Cite this article: Hilbig, B. E. and Thielmann, I. (2025). Toward a (more) parsimonious account of the link between 'dark' personality and social decision-making in economic games. *Judgment and Decision Making*, e16. https://doi.org/10.1017/jdm.2025.1