Judgment and Decision Making (2025), **vol. 20**:e38 1–17
doi:10.1017/jdm.2025.10013

The European Association for Decision Making
SOCIETY FOR JUDGMENT AND DECISION MAKING

EMPIRICAL ARTICLE

Strangers in the dark: assumed similarity in judgments of unknown others on aversive personality

Nicholas Poh-Jie Tan[1], Ben Hilbig[2], Morten Moshagen[3], Ingo Zettler[4], Sophia Payer[2] and Isabel Thielmann[5]

[1]The University of Zürich, Switzerland; [2]RPTU Kaiserslautern-Landau, Germany; [3]Ulm University, Germany; [4]University of Copenhagen, Denmark and [5]Max Planck Institute for the Study of Crime, Security and Law, Germany

Corresponding author: Nicholas P. Tan; Email: nicholas.tan@psychologie.uzh.ch

Received: 31 October 2024; **Revised:** 10 July 2025; **Accepted:** 2 September 2025

Keywords: dark factor; assumed similarity; person perception; personality

Abstract
The need to maintain cooperation in social dilemmas is a fundamental challenge. Responses to social dilemmas are affected by dispositions toward exploitativeness (i.e., the maximization of one's own utility) and distrust (i.e., the fear of being exploited by others). This is because the belief that others are untrustworthy justifies exploitative behaviors. The Dark Factor of Personality (D) is postulated to comprise the conjunction of these dispositions, implying that individuals will assume similarity on D. In this research, we sought to test this implication by examining whether individuals' self- and observer reports of unacquainted targets on D converge. Across five studies, we found that individuals assume similarity on D when unknown targets are described as 'typical' (Study 1) or when shown a photograph (Studies 2–5). These effects were not moderated by the congruency between rater and target sex (Studies 2 and 3); however, we found that higher attractiveness of targets led to greater assumed similarity on D (Studies 4 and 5). These findings are consistent with D reflecting the conjunction of exploitativeness and distrust while also suggesting that assumed similarity on D is moderated by the interpersonal attraction of those being rated.

1. Introduction

Among the most fundamental challenges to human interaction—from dyads to entire societies and nations—is the need to maintain cooperation in social dilemmas (Axelrod and Hamilton, 1981). The common structure of all social dilemmas is that cooperation is mutually beneficial (maximizing social welfare), whereas unilateral exploitation is individually utility maximizing but associated with less overall welfare (Kollock, 1998; Van Lange et al., 2013). Real-life examples of such social dilemmas range from doing the cleaning in a shared flat and paying for public transport to climate change mitigation and nuclear arms control.

Personal characteristics and perceptions of others are likely to influence the decision whether to defect or cooperate in social dilemmas. In general, two broad factors may influence the decision to defect in these situations (e.g., Bruins et al., 1989): First, unilateral defection is *tempting* because one may thereby maximize one's own utility, minimize the other's utility, or maximize the difference between one's own and the other's utility (Kelley et al., 2003; Kelley and Thibaut, 1978). That is, social dilemmas provide a possibility for exploitation (Thielmann et al., 2021), suggesting that person characteristics capturing the tendency to exploit should play a role for corresponding behavior. Second, one must *fear* unilateral cooperation and thus being exploited by the other. As such, social dilemmas also involve dependence on others under uncertainty: One's outcome will ultimately be influenced by interaction partners' behavior, which is unknown to the acting individual. Thereby one's beliefs about others' exploitativeness will influence one's own behavior (Balliet and Van Lange, 2013), suggesting the influence of person perception processes. Taken together, defection in social dilemmas can thus be seen as an expression of exploitativeness (i.e., low concern for others' welfare, also often referred to as *greed*) and/or distrust (i.e., pessimistic beliefs in others' prosociality, also often referred to as *fear*; Bruins et al., 1989; Hilbig et al., 2018; Thielmann et al., 2020a).

While corresponding dispositional tendencies or traits—exploitativeness and distrust—are often conceptualized separately from each other (i.e., within different broader traits), recent theoretical developments on the nature of socially/ethically aversive personality suggest that these tendencies are two sides of the same coin. Specifically, the conjunction of such tendencies is mandated by the substantive definition that has been put forward to describe the common core of all socially/ethically aversive personality traits: 'the general 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, emphases added). This tendency shared by all aversive traits—maybe most prominently, 'dark' traits such as narcissim, machiavellianism, and psychopathy (Paulhus and Williams, 2002)—has been termed the Dark Factor of Personality, or simply *D*. According to its definition, individuals who are more likely to exploit others should also be more likely to expect exploitation from others (e.g., distrust) because the latter serves as a justification for the former.

The explicit inclusion of justifying beliefs in the conceptualization of D sets it apart from other broad traits that also capture individual differences in aversive behavior (e.g., honesty–humility; Ashton et al., 2014). It also reflects the well-established notion that people are highly motivated to justify their immoral behavior in order to see themselves as moral (Paulhus and John, 1998; Prentice et al., 2019). Moreover, people are willing to modify their beliefs to maintain their positive self-image in the face of immoral behavior (Mazar et al., 2008). Correspondingly, D is not only associated with a range of distrust-related beliefs above and beyond traits like honesty–humility and agreeableness (Horsten et al., 2021; Moshagen et al., 2020b; Thielmann and Hilbig, 2023), but such beliefs also account for the link between D and actual exploitative behavior (Hilbig et al., 2022). In other words, those high in D expect others to be exploitative in the sense of a descriptive social norm (Schultz et al., 2007) to justify their own exploitativeness: As exploitation is expected to be widespread, it is acceptable or indeed necessary to do the same (i.e., 'everyone does it' and 'eat or be eaten'). In turn, the very definition of D implies that those high in D should judge others to be exploitative and thus also high in D. Indeed, items from the D inventory such as 'Most people are basically good and kind' [reversed scored] (Moshagen et al., 2020a) suggest that D directly involves judgments of others.

In research on person perception, the phenomenon of judging that others share one's own characteristics is commonly known as *assumed similarity*—the convergence between how individuals see themselves and how they see others (Cronbach, 1955).[1] By implication, testing for assumed similarity on D can provide a test of the notion that D represents the *conjunction* of exploitativeness and distrust, as per its substantive definition, and further advance research on person perception with regard to aversive personality traits. This work therefore sought (i) to test the hypothesis that individuals assume similarity when rating others on D and (ii) to identify potential boundary conditions of this relation to allow conclusions on its generalizability across contexts.

In addition to the theoretical rationale provided so far, there are also empirical indications that support the prediction of assumed similarity on D. First, previous research suggests that people assume

[1] Assumed similarity has been studied under different labels, including the false-consensus effect, social projection, and self-anchoring. We use assumed similarity in this research as it most accurately describes the observed phenomenon in question. For example, while the false-consensus effect describes individuals' belief that their characteristics are common, assumed similarity describes the belief that individuals share certain characteristics with a specific other person.

https://doi.org/10.1017/jdm.2025.10013 Published online by Cambridge University Press

similarity on specific aversive traits that can be considered manifestations of D (Hilbig et al., 2023; Moshagen et al., 2018). Specifically, Webster and Campbell (2023) assessed assumed similarity of known fictional television series characters by asking participants to not only rate themselves on the dark tetrad traits (i.e., machiavellianism, narcissism, psychopathy, and sadism) but also to choose one of 56 characters to rate on the dark tetrad. They found that participants assumed these characters were similar to themselves on the dark tetrad traits. Arguably, if people assume similarity across all the dark tetrad traits, then people may also assume similarity on what all these traits share, viz. D. Critically, however, because participants in this prior study rated targets they 'knew', it is possible that at least part of the assumed similarity effect observed can be explained by actual similarity between raters and targets.

Moreover, expecting assumed similarity on D also aligns with evidence and arguments on the trait specificity of assumed similarity. These arguments posit that traits more closely linked to personal values are particularly important to people's identity, thus triggering a motivational mechanism: Since people want others to share their values, they perceive similarity on corresponding traits in particular (Lee et al., 2009; Thielmann et al., 2020a). The origin of this theoretical idea lies in the observation that, among basic personality traits, the strongest assumed similarity effects ($r \geq 0.23$) have been found for those traits that are most strongly linked to values: Honesty–humility from the HEXACO (i.e., the tendency to be sincere, honest, and fair-minded; Ashton and Lee, 2007), agreeableness from the Big Five (i.e., the tendency to be considerate of the needs, feelings, and concerns of others; DeYoung et al., 2007), and openness to experience from both the HEXACO and the Big Five (i.e., the tendency to be imaginative, curious, and reflective; DeYoung, 2015). Moreover, assumed similarity has been shown to increase with increasing value relatedness of traits, even after accounting for other trait characteristics, such as their social desirability (Thielmann et al., 2023). D, in turn, is not only substantially related to value-related basic traits—most prominently HEXACO honesty–humility (Moshagen et al., 2018)—but it is itself also conceptually linked to self-enhancement versus self-transcendence values (García-Fernández et al., 2025). Specifically, the characteristics of those high on D, such as the motivation to increase reputational status and increase feelings of power (Moshagen et al., 2018), overlap with the content of self-enhancement values (Schwartz, 1992). Thus, the notion that value-related traits tend to produce assumed similarity would also suggest assumed similarity effects on D, which has yet to be tested directly.

2. *The present research*

The aim of this research was to test the theoretically derived prediction of assumed similarity on D—reflecting the defining assumption of a conjunction of exploitativeness and distrust within a single trait. Assumed similarity is usually inferred from the association between self-reported trait levels and judgments of a target's level on the same trait (i.e., observer reports of the target). In this research, we chose to focus on observer reports of unacquainted others (i.e., strangers) to offer a strong test of the proposition under scrutiny, controlling for the potential influence of known and actual similarity between rater and target person. Thus, across five studies, participants provided observer reports on D about unacquainted targets that were either imagined or depicted but unknown (self- and observer reports of D are summarized in Table 1). To test the robustness of the proposed assumed similarity effect, we further examined the potentially moderating influence of targets' sex (i.e., sex congruency) and attractiveness as two factors that have been argued to affect assumed similarity (Marks and Miller, 1982; Paunonen and Hong, 2013). Taken together, we hypothesized that individuals would assume unacquainted others have similar levels on D to themselves and sought to test whether this effect holds when rating sex-congruent and attractive targets.

All statistical analyses were performed using R (R Core Team, 2023) and assumptions for all models were assessed visually using the *sjPlot* package (i.e., homoscedasticity, normality of residuals, and outliers; Lüdecke, 2023). All research data and code are publicly available on the Open Science Framework (https://osf.io/t4s3u/). Study 3 was preregistered. Of note, the images used in this research

https://doi.org/10.1017/jdm.2025.10013 Published online by Cambridge University Press

Table 1. *Dark factor questionnaires used for self- and observer reports.*

Study	Self-report	Observer report
Study 1	D16	D16
Study 2	D16	D16
Study 3	D35	D35
Study 4	D70	D35
Study 5	D16	D16

Note. To see the items for these measures, see https://osf.io/t4s3u/. Histograms for these measures are shown in Supplementary Figures S2–S6. Although different versions of the D questionnaire were used across studies, and in some instances for self- and observer reports within the same study, it is unlikely that this affected our results as these measures all converge strongly (Moshagen et al., 2020a).

are copyrighted and thus not for public distribution.[2] Therefore, these images are not made publicly available, but they can be obtained from the first and last authors.

3. Study 1

3.1. Methods

3.1.1. Participants

We recruited participants among individuals who completed an online self-assessment at https://darkfactor.org (see Moshagen et al., 2020a, for details). This website provides general information about D and allows people to complete a self-report questionnaire measuring D to determine their D level relative to others. All procedures of the website are approved by the local ethics committee of the RPTU University Kaiserslautern-Landau, Department of Psychology (approval #LEK-154 and #LEK-567). Participants were given the option to complete a 16-, 35-, or 70-item measure of D (nowadays, individuals can only choose between the 16- and the 70-item measure). For the current investigation, we focused on individuals completing the English 16-item D questionnaire (D16) between September 2020 and November 2020 who agreed to work on an additional task as described next. For all studies run on https://darkfactor.org, exclusion criteria are set a priori; more details can be found at https://osf.io/93tw6/. The sample consisted of N = 702 participants (46% female) aged 18–79 (M = 42.6, SD = 16.7) years. See Supplementary Tables S1 and S2 for full demographic information. The sample size gave us 80% power to detect small effects of $r = 0.11$ (two-tailed, $\alpha = 0.05$).

3.1.2. Procedure and measures

This study was conducted in English and participants began by giving their informed consent before completing the D16 which asked participants to indicate their agreement to the statements (e.g., ‘When I get annoyed, tormenting people makes me feel better’) on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree; Moshagen et al., 2020). After completing the D16, participants were asked to report their demographic information and whether they consented to the use of their data for scientific purposes. Next, participants indicated whether they would like to volunteer for an additional task and were reassured that they would receive feedback on their level of D independent of whether they volunteered for the additional task or not. Those who agreed were then presented with the observer report form of the D16 and asked to rate the statements about a ‘typical other’. Specifically, participants were given the instructions ‘If we were to select some stranger randomly from the population, how much would you say does each statement apply to this person?’ Both the self- and observer-reported D16 had good reliability ($\alpha \geq 0.88$); to create a single D score, both were scored by computing the mean

[2] In our preregistration, we had mistakenly stated that the images were openly accessible.

across items after recoding reverse-keyed items. After completing the observer report, participants received feedback about their own level of D relative to others.

3.2. Results and discussion

To test assumed similarity on D, we computed the zero-order correlation between participants' mean self- and observer-reported D scores. We found a significant, medium-to-large (Cohen, 1988) positive correlation between self- and observer-reported D, $r(700) = 0.40$, 95% CI [0.33, 0.46], $p < .001$. This suggests that participants assumed that random strangers have similar levels on D to themselves. Of note, the size of this assumed similarity effect is in line with assumed similarity correlations found for honesty–humility when rating strangers (e.g., $r = 0.43$; Thielmann et al., 2020). In addition, we found that self-reported D ($M = 2.0$, $SD = 0.7$) was significantly lower than observer-reported D ($M = 2.6$, $SD = 0.7$), $t(1394.30) = 16.23$, $p < .001$, $d = 0.87$, 95% CI [0.76, 0.98], suggesting that participants see themselves more positively than they see others.

4. Study 2

In Study 1, we found evidence for assumed similarity on D when raters were asked to imagine a typical unknown other. However, because there was no information about the target and imagining a 'typical other' is difficult, it is possible that participants thought of a specific typical person in their lives, such as a friend or colleague. This may have artificially inflated the assumed similarity correlation due to actual similarity between rater and target. Thus, to address these potential issues, in Study 2, we sought to replicate the assumed similarity effect on D when asking raters to judge an unknown target depicted on a photo.

Another potential confound of assumed similarity besides actual similarity is spurious similarity (Paunonen and Hong, 2013). Spurious similarity is present when raters and targets share a certain group membership (e.g., sex, education, ethnic background) and raters rely on this information about the targets' group membership when completing observer reports. This may also have inflated the assumed similarity correlation in Study 1: The raters themselves may have considered themselves to be a typical or normal person and therefore may have judged another member of their group in similar ways as they judged themselves. In that regard, it seems likely that men imagined another man, whereas women imagined another woman. This is problematic because males tend to score higher on D than females (Hartung et al., 2022). Hence, if there is congruency between rater and target sex, what appears to be assumed similarity might actually be spurious similarity due to the convergence between one's own level on D and the accurate rating of same-sex targets. Consequently, a more stringent test of assumed similarity on D is whether assumed similarity is found for sex-incongruent rater–target pairs. To investigate this possibility, in Study 2, we additionally explored whether assumed similarity on D was moderated by the congruence between raters' and targets' sex which, in turn, would test whether assumed similarity on D is present above and beyond spurious similarity.

4.1. Methods

4.1.1. Participants

Similar to Study 1, we recruited participants through https://darkfactor.org, and focused on those who completed the English version of the D16 between December 2020 and April 2021 and agreed to complete an additional task. A total of $N = 325$ (55% female) participants aged 18–75 ($M = 32.0$, $SD = 12.6$) years provided usable data for the current analyses, which provided us with satisfactory power (80%) to detect relatively small correlations of $r = 0.15$ with two-tailed $\alpha = 0.05$. Full demographic information of the sample can be found in Supplementary Tables S1 and S2.

https://doi.org/10.1017/jdm.2025.10013 Published online by Cambridge University Press

4.1.2. Procedures and measures

The procedure and measures are the same as in Study 1 with the exception that when completing the observer reports, we presented participants with a photograph of an unknown individual and asked them to rate this person. Photos were randomly selected from a pool of 60 photographs taken from a stock photo database (https://www.colourbox.com/; 30 male, 30 female). The photos presented the upper body of targets in a spontaneous pose before a neutral background. The photo was always visible to participants while responding to the items. Both the self- and observer-reported D16 had good reliability ($\alpha \geq 0.87$).

4.2. Results and discussion

Replicating the findings from Study 1, we found a significant, medium-sized (Cohen, 1988), zero-order correlation between self- and observer reports on D indicating assumed similarity, $r(323) = 0.34$, 95% CI [0.24, 0.44], $p < .001$. Also, we again found that self-reported D ($M = 2.3$, $SD = 0.7$) was significantly lower than observer-reported D ($M = 2.6$, $SD = 0.7$), $t(648) = 6.17$, $p < .001$, $d = 0.48$, 95% CI [0.33, 0.64].

To explore the effect of sex congruence, we first computed correlations between self- and observer reports separately for same-sex and different-sex rater–target dyads. Both correlations were comparable in size, yielding $r(158) = 0.30$, 95% CI [0.15, 0.44], $p < .001$, when sex was congruent, and $r(163) = 0.37$, 95% CI [0.23, 0.50], $p < .001$, when sex was incongruent. To explore these effects further, we ran nested multiple regression models. The first model (Model 1) tested the main effect of self-reported D predicting observer-reported D. The next model (Model 2) tested whether sex congruence moderated assumed similarity by testing the three-way interaction between the self-reported D, sex of the rater, and the sex of the target (for both factors, male was coded as −1 and female as 1). To assess whether participants assumed similarity on D, we only interpreted the main effect in Model 1, and to test the moderation of sex congruence, we only interpreted the three-way interaction effect in Model 2. For both models, self- and observer reports on D were standardized prior to analysis. As shown in Table 2, we found that higher self-reported D significantly predicted higher observer-reported D, indicating assumed similarity. As suggested by the similar effect sizes in the two sex-congruence conditions, the interaction between self-reported D, rater sex, and target sex was nonsignificant. Thus, participants' assumed similarity on D did not depend on whether targets shared or did not share the same sex. The remaining effects in the model were also all nonsignificant. Taken together, we again found evidence that people assume unknown others are similar to themselves in terms of D. This effect, in turn, was not moderated by congruence between raters' and targets' sex. This finding speaks against spurious similarity resulting from shared group membership in terms of sex to inflate the assumed similarity effect on D.

5. Study 3

Study 2 provided additional evidence that people assume unknown others to have similar levels on D, this time when they were shown a photograph of a stranger. Moreover, we found evidence that assumed similarity on D did not depend on the congruence between raters' and targets' sex. The goal of Study 3 was to replicate the findings in Study 2, this time by testing the moderation of sex congruence of assumed similarity in a within-subjects design. Given the plausibility that assumed similarity on D might be stronger when rating same-sex targets, in Study 3, we preregistered the hypothesis that assumed similarity on D would be moderated by sex congruency (https://osf.io/t4s3u/), despite our Study 2 findings.

Table 2. *Regression results predicting observer-reported D (Studies 2–5).*

Study	Model	Predictor	β	Lower CI	Upper CI	Partial R^2	R^2
Study 2	Model 1	Intercept	0	−0.10	0.10	0	0.12***
		Self-reported D	0.34***	0.24	0.44	0.12	
	Model 2	Intercept	−0.01	−0.12	0.10	0	0.14***
		Self-reported D	0.35***	0.25	0.46	0.12	
		Rater sex	0.09	−0.01	0.20	0.01	
		Target sex	−0.09	−0.20	0.02	0.01	
		Self-reported D × Rater sex	0.01	−0.10	0.11	0	
		Self-reported D × Target sex	0	−0.11	0.10	0	
		Rater sex × Target sex	−0.10	−0.20	0.01	0.01	
		Self-reported D × Sex congruence	−0.05	−0.16	0.05	0	
Study 3	Model 1	Intercept	0	−0.14	0.14	0.13	0.12/0.44
		Self-reported D	0.35***	0.21	0.49	0.13	
	Model 2	Intercept	0	−0.19	0.12	0.21	0.19/0.54
		Self-reported D	0.38***	0.23	0.53	0.15	
		Rater sex	0.13	−0.02	0.29	0.04	
		Target sex	0.20***	0.10	0.29	0.02	
		Self-reported D × Rater sex	0.03	−0.12	0.18	0.01	
		Self-reported D × Target sex	−0.05	−0.14	0.05	0.01	
		Rater sex × Target sex	0.07	−0.02	0.17	0	
		Self-reported D × Sex congruence	0.07	−0.02	0.17	0	
Study 4	Model 1	Intercept	0	−0.07	0.07	0	0.06***
		Self-reported D	0.25***	0.18	0.31	0.06	
	Model 2	Intercept	0.01	−0.04	0.06	0	0.52***
		Self-reported D	0.19***	0.14	0.23	0.07	
		Target attractiveness	−0.67***	−0.71	−0.62	0.48	
		Self-reported D × Target attractiveness	0.11***	0.07	0.15	0.03	
Study 5	Model 1	Intercept	0	−0.13	0.13	<0.001	0.01
		Self-reported D	0.09	−0.04	0.22	0.01	
	Model 2	Intercept	0.06	−0.12	0.24	<0.001	0.04*
		Self-reported D	−0.08	−0.26	0.10	<0.001	
		Target-attractiveness	−0.12	−0.38	0.13	<0.001	
		Self-reported D × Target attractiveness	0.35**	0.10	0.60	0.03	

Note. *** $p < .001$; ** $p < .01$; * $p < .05$; lower and upper CI's represent 95% confidence intervals; rater sex and target sex are coded as male = −1 and female = 1. For Study 3, sex congruence represents the terms rater sex and target sex. The R^2 for Study 3 represents the marginal R^2/conditional R^2. For Study 5, the target-attractiveness factor indexes whether participants were rating low attractiveness targets (coded as '−1') or high attractiveness targets (coded as '1'). Attractiveness, self-reported D, and observer-reported D were all standardized prior to analysis.

5.1. *Methods*

5.1.1. Participants

Our target sample size was determined based on a priori power analyses to detect small- to medium-sized effects (e.g., $r = 0.25$) with 80% power using a one-tailed test (see preregistration for more details). We ran separate power analyses for tests of the zero-order assumed similarity correlation and the interaction effect between observer report and sex congruence (i.e., multiple regression analysis) and ultimately based the required sample size on the larger resulting estimate ($N = 97$) while also accounting for potential participant exclusions. Thus, our goal was to recruit at least $N = 110$ participants.

Participants were recruited from the local university mailing list at a German university, social media, and through snowball sampling. As compensation for participation, university students were given course credits. We collected as many participants as possible between August and September 2021. Of the 180 participants that started the study, 132 completed it. We applied the following preregistered exclusion criteria: Failing at least one of three attention checks which were embedded in the questionnaires and simply asked participants to indicate a specific response (e.g., 'please select strongly agree'; $n = 15$), taking less than 2 seconds on average per D item ($n = 1$), and having low variance across D items (i.e., $SD < 0.3$; $n = 5$). We also checked if any participant recognized the photographed targets and whether they had a good grasp of the German language (the study was conducted in German); however, all participants passed these criteria. The final sample size was $N = 111$ (65% female) aged 18–73 ($M = 29.5$, $SD = 12.8$) years. Full demographic information can be found in Supplementary Tables S1 and S2.

5.1.2. Procedure and measures

Participants first provided informed consent and demographic information. They then completed the self-reported 35-item measure of D (D35; Bader et al., 2022) by indicating their agreement to the items on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). Next, participants saw the stimuli for the same-sex target condition and the different-sex target condition, in counterbalanced order. In both conditions, participants were shown a half-body photograph taken from a pool of 20 photographs (again taken from https://www.colourbox.com/; 10 male, 10 female). After seeing the stimuli for one condition, participants rated the photographed individual using the observer version of the D35. They were then shown the stimuli for the other condition and again rated the photographed individual using the same items. Both the self- and observer reports of the D35 showed high internal consistency ($\alpha \geq 0.88$).

5.2. *Results and discussion*

Consistent with our results from Studies 1 and 2, we found a significant, medium-sized, positive zero-order correlation between self- and observer-reported D for both same-sex, $r(109) = 0.35$, 95% CI [0.17, 0.50], $p < .001$, and different-sex targets, $r(109) = 0.37$, 95% CI [0.19, 0.52], $p < .001$.[3] We also found that self-reported D ($M = 1.9$, $SD = 0.4$) was significantly lower than observer-reported D ($M = 2.6$, $SD = 0.6$), $t(187.22) = 10.05$, $p < .001$, $d = 1.35$, 95% CI [1.06, 1.64].

As preregistered, we examined the effect of sex congruence on assumed similarity by running multilevel regressions with random intercepts for participants. We followed the same nested modelling approach as in Study 2 except that we used the *lme4* and *lmerTest* packages (Bates et al., 2015; Kuznetsova et al., 2017) to run mixed effects models for the within-person manipulation of sex congruence. Moreover, this modeling approach deviated from our preregistered analysis plan as we

[3] We also tested for the presence of regression to the mean by examining scatterplots of self-reported and observer-reported D for participants who saw the incongruent condition first and for those who saw the congruent condition first. If responses in the second condition that participants saw moved closer to the assumed similarity association in the first condition, this would have constituted evidence for regression to the mean. However, there was no evidence of this for either order (see Supplementary Figures S7 and S8).

https://doi.org/10.1017/jdm.2025.10013 Published online by Cambridge University Press

tested a three-way interaction between self-reported D, rater sex, and target sex (whereas we had preregistered an interaction between self-reported D and a factor which indexed whether participants rated a sex congruent or incongruent target). Importantly, results from the preregistered models led to the same conclusions (see Supplementary Table S3). As shown in Table 2, higher self-reported D was associated with higher observer-reported D, indicating assumed similarity. Once again, the three-way interaction between self-reported D, rater sex, and target sex was nonsignificant, indicating that assumed similarity on D was not moderated by sex congruence. In addition to that, there was a significant effect of target sex on observer-reported D which, surprisingly, indicated that female targets were rated slightly higher on D than male targets. All other effects were nonsignificant (see Table 2). In summary, we mostly replicated the pattern of results from Study 2, providing further evidence against spurious similarity resulting from sharing one's sex with a target to inflate assumed similarity on D.

6. Study 4

Thus far, we consistently found evidence for assumed similarity on D—in line with the theoretical notion that D represents the conjunction of exploitativeness and distrust. To further explore the robustness of this effect, we investigated whether it may be moderated by the interpersonal attraction of targets. Specifically, interpersonal attraction is defined as 'a positive attitude or evaluation regarding a particular person' (Aron and Lewandowski, 2001, p. 7860). Thus, interpersonal attraction can result from likability or physical attractiveness. Past research has found that assumed similarity tends to be stronger when raters are more attracted to targets (for a review, see Thielmann and Hilbig, 2022). For example, greater interpersonal attraction is associated with higher assumed similarity of attitudes (Marks and Miller, 1982; Mashman, 1978) and personality traits (Marks et al., 1981; Miyake and Zuckerman, 1993).

There are several potential explanations for why interpersonal attraction might moderate assumed similarity. First, assuming similarity highlights the commonalities between the rater and the target, thereby enhancing one's own self-image if the target is likable/attractive (Marks and Miller, 1987). Second, assuming that people have a positive self-view, people are motivated to maintain cognitive balance (Heider, 1958), which can be achieved by bringing into balance one's personality with a likable or an attractive other. Finally, assuming similarity for likable/attractive targets could increase feelings of belongingness with these targets (Machunsky et al., 2014). Following these notions, the assumed similarity effect on D may be stronger with increasing interpersonal attraction between rater and target. Then again, if assumed similarity on D exclusively serves to justify one's own exploitativeness, interpersonal attraction should be irrelevant for assumed similarity to occur. Thus, in Study 4, we sought to explore whether interpersonal attraction—operationalized as likeability—moderated the association between self- and observer reports on D.

6.1. Methods

6.1.1. Participants

The data used for this research came from the Prosocial Personality Project (PPP), a multiwave online study conducted via a panel provider in Germany. A detailed documentation of the project including participant compensation, variables assessed, a priori exclusion criteria, sample sizes and compositions per wave, and prior publications using part of the PPP data is available online (https://osf.io/m2abp/). None of the prior publications has studied assumed similarity; thus, the observer reports on D have not been used in any published analyses yet.

Wave 1 of the PPP base sample included $N = 4585$ participants. Of these, $N = 862$ completed the observer report on D about an unknown target presented on a photo. However, 28 participants indicated that they knew the target presented in the photograph and their data were therefore not included in the analyses. Thus, the final sample for the current investigation included $N = 834$ (46% female)

https://doi.org/10.1017/jdm.2025.10013 Published online by Cambridge University Press

participants aged 19–74 (M = 45.6, SD = 12.0) years. A sensitivity analysis suggests that this sample size gives us 80% power to detect an effect of r = 0.10. Full demographic information can be found in Supplementary Tables S1 and S2.

6.1.2. Procedure and measures

At wave 1 of the PPP base study, which was conducted in 2019, participants began by giving their informed consent and demographic information before completing different self-report measures including the D70 (Bader et al., 2022). This measure asked participants to indicate their agreement to 70 items on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree; α = 0.95). In a follow-up wave collected in 2021 (i.e., termed follow-up 2021-01b within the PPP), participants were shown a randomly selected photograph from the same pool used in Study 2 (half-body; 30 male, 30 female) and asked to rate the photographed individual using the observer-report form of the D35 (α = 0.95). Afterward, they were presented with two items measuring interpersonal attraction which asked participants to rate 'how likeable do you find the person' (1 = very likeable to 6 = not at all likeable) and 'would you like to get to know the person' (1 = yes, very much to 6 = no, not at all). We reversed scored both items so that higher scores indicated higher likability. Both items were strongly correlated, $r(832)$ = 0.80, 95% CI [0.77, 0.82], p < .001, and were thus combined to an interpersonal attraction score by computing a mean for each participant. Finally, participants were asked to indicate whether they knew the target person (yes vs. no). Subsequently, participants responded to additional self-report scales that are not pertinent to the present investigation.

6.2. Results and discussion

We again found a significant, this time small- to medium-sized, positive zero-order correlation between self- and observer reports on D indicating assumed similarity, $r(832)$ = 0.25, 95% CI [0.18, 0.31], p < .001. We again also found that self-reported D (M = 2.1, SD = 0.4) was significantly lower than observer-reported D (M = 2.5, SD = 0.6), $t(1592.50)$ = 12.02, p < .001, d = 0.57, 95% CI [0.48, 0.67].

To examine whether interpersonal attraction moderated this effect, we ran nested multiple linear regression models similar to our previous studies. However, Model 2 tested whether interpersonal attraction moderated assumed similarity by including the attraction score and the interaction between self-reported D and attraction as predictors.[4] All variables were standardized prior to analysis. As shown in Table 2, we found the interaction between self-reported D and attraction predicting observer-reported D to be significant. To follow-up on this effect, we ran a simple slopes analysis using the *interactions* package (Long, 2024) and found that at lower levels of attraction (−1 SD) there was a weaker, yet still significant, assumed similarity effect, β = 0.08, SE = 0.03, 95% CI [0.02, 0.14], p = .013. However, at mean, β = 0.19, SE = 0.02, 95% CI [0.14, 0.23], p < .001, and high levels of attraction (+1 SD), β = 0.30, SE = 0.03, 95% CI [0.23, 0.36], p < .001, the assumed similarity effect was more pronounced (see Figure 1). Thus, the more likable targets were perceived, the more stronger assumed similarity on D. In addition, we also found a main effect for interpersonal attraction showing that more likable targets were rated lower on D. That is, raters ascribed more less socially undesirable characteristics to targets they considered more likable.[5]

The interaction effect is in line with previous research showing that assumed similarity is stronger when rating targets one feels attracted to (Marks et al., 1981; Marks and Miller, 1982; Mashman, 1978;

[4] To test the robustness of the interaction between self-reported D and interpersonal attraction, we also specified a model in which we added the quadratic effects for self-reported D and interpersonal attraction. Supporting the results, the linear interaction effect between self-reported D and interpersonal attraction remained significant when including the quadratic effects (see Supplementary Table S4 for results of this analysis).

[5] We also tested whether sex-congruency moderated the assumed similarity effect, and whether there was a significant interaction between sex congruency and interpersonal attraction on assumed similarity. As reported in Supplementary Table S5, these effects were all nonsignificant.

https://doi.org/10.1017/jdm.2025.10013 Published online by Cambridge University Press

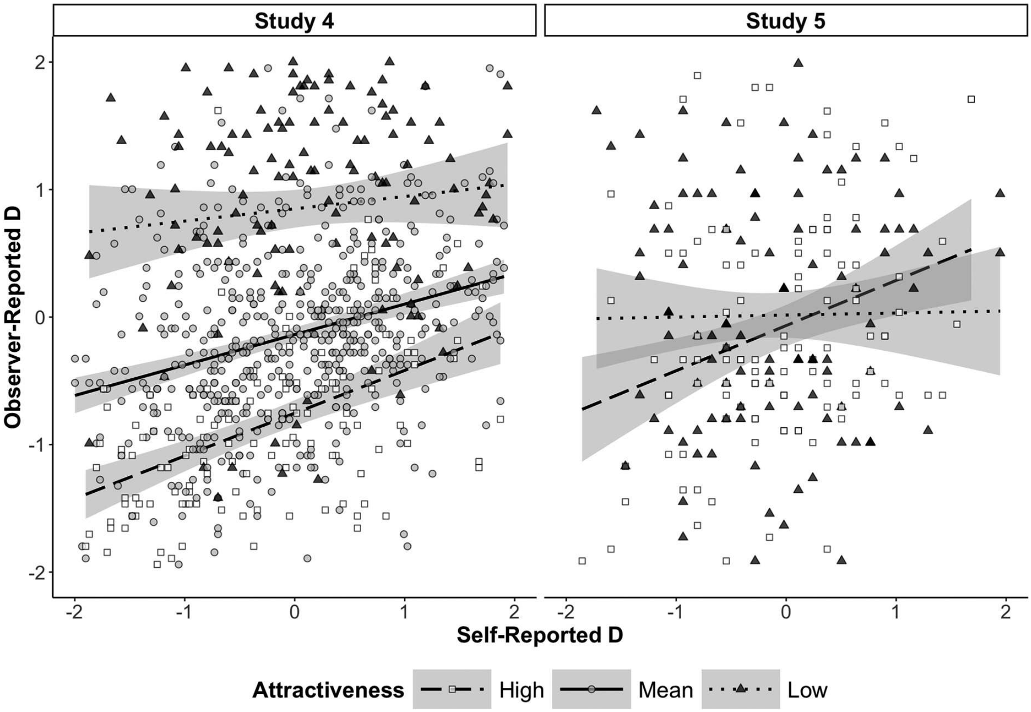

Figure 1. *Scaled and mean-centered self- and observer-reported D split by levels of target interpersonal attraction (Studies 4 and 5). Note. For Study 4, high- and low-target interpersonal attraction corresponds to attractiveness rating at +1 SD and −1 SD from the mean, respectively; For Study 5, high- and low-target interpersonal attraction corresponds to the targets that were selected for participants to rate; The shaded area represents the 95% confidence interval. Although there appears to be a ceiling effect for observer-reported D in Study 4 for those who reported low target attractiveness, there is no evidence of this when examining histograms of this variable (see Supplementary Figure S5).*

Miyake and Zuckerman, 1993). Moreover, it is consistent with theorizing that one is incentivized to assume similarity with others because it highlights ones commonalities and belongingness with interpersonally attractive targets (Machunsky et al., 2014; Marks and Miller, 1987), and to maintain cognitive balance (Heider, 1958). However, this finding does not follow from the theory of D itself. Indeed, it appears that there is a motivational component to assumed similarity on D that goes beyond mere justification for exploitativeness. The main effect of attractiveness predicting lower observer ratings of D is consistent with halo effects suggesting that attractive individuals are ascribed with more socially desirable traits than less attractive individuals (Batres and Shiramizu, 2023; Klebl et al., 2022).

7. Study 5

Study 4 showed assumed similarity on D to be moderated by interpersonal attraction yielding stronger assumed similarity for more likable targets. Study 5 sought to replicate and extend this finding. Specifically, because we did not experimentally manipulate interpersonal attraction in Study 4, it is unclear whether perceived likability drove assumed similarity or vice versa. Thus, in Study 5, we experimentally manipulated interpersonal attraction—this time operationalized as physical attractiveness—so that participants were asked to rate either a target they considered physically attractive or a target they considered physically unattractive.

https://doi.org/10.1017/jdm.2025.10013 Published online by Cambridge University Press

7.1. Methods

7.1.1. Participants

We initially recruited N = 314 participants from a local university mailing list, social media, and through snowball sampling. For university students, course credits were given as compensation for participation. Remaining participants did not receive compensation. After excluding participants for not completing the study (n = 69) and for failing an attention check (i.e., ‘Please select “strongly agree” here. This is to check your attention’; n = 7), the final sample contained N = 238 participants (68% female) aged between 18 and 77 (M = 29.8, SD = 12.7) years, 58% of whom were students. A sensitivity analysis suggests that with this sample size we have 80% power to detect an effect of r = 0.18. For full demographic information, see Supplementary Tables S1 and S2.

7.1.2. Procedure and measures

The study was conducted in German. First, participants were asked to give their informed consent and answer demographic questions. Next, they completed the self-report version of the D16 (Moshagen et al., 2020a) before they were asked to provide interpersonal attraction ratings of 10 photographed individuals. These photographs were randomly sampled from a pool of 27 photographs (13 female and 14 male) taken from the same online image database as before (https://www.colourbox.com/). Interpersonal attraction ratings were captured by asking participants to indicate how attractive they perceived each target to be using a slider from ‘very unattractive’ (1) to ‘very attractive’ (100). Next, of the 10 photographed individuals, we randomly selected a same-sex target that participants had either rated as the most attractive or the least attractive before, depending on which condition (attractive vs. unattractive target) a participant was randomly assigned to. In the event of a tie, meaning that two or even more pictures received an identical (highest or lowest) attractiveness score, we randomly selected one of these pictures for the rating (n = 1). Participants then completed the observer report form of the D16 for the selected target. Both the self- and observer report of the D16 had satisfactory reliability ($\alpha \geq 0.83$).

7.2. Results and discussion

Unlike in the previous studies, this time we did not find a significant zero-order correlation between self- and observer-reported D across both attractive and unattractive targets, $r(236) = 0.09$, 95% CI [−0.04, 0.22], $p = .159$. However, this was attributable to the fact that for unattractive targets, there was no evidence for assumed similarity, and even a slight tendency for assumed *dis*similarity, $r(117) = -0.08$, 95% CI [−0.26, 0.10], $p = .390$. For attractive targets, by contrast, we again found a significant, medium-sized, positive correlation between self- and observer-reported D and, thus, assumed similarity, $r(117) = 0.27$, 95% CI [0.10, 0.43], $p = .003$. Finally, we found that self-reported D (M = 1.9, SD = 0.5) was significantly lower than observer-reported D (M = 2.5, SD = 0.7), $t(427.03) = 12.21$, $p < .001$, $d = 1.12$, 95% CI [0.93, 1.31].

To examine the effect of interpersonal attraction on assumed similarity more directly, we ran nested multiple linear regressions as in our previous studies. However, this time we added a *target-attractiveness* factor that indexed whether participants rated an attractive (coded as ‘1’) or unattractive target (coded as ‘−1’) and the interaction between self-reported D and target attractiveness in Model 2. Both self- and observer-reported D were standardized. In line with the apparent differences in assumed similarity correlations between conditions, there was a significant interaction between self-reported D and target-attractiveness (see Table 2 and Figure 1). The two main effects were not significant (see Table 2).

In summary, we again found that participants reported that they were lower on D than others. This finding is consistent with evidence on the better than average effect, which describes how people tend to see themselves more positively than they see others (Zell et al., 2020). We also again found that assumed similarity was stronger when rating more interpersonally attractive targets. Thereby,

https://doi.org/10.1017/jdm.2025.10013 Published online by Cambridge University Press

Study 5 extends our findings from Study 4 to physical attractiveness, and—because targets were experimentally assigned to attractiveness conditions—supported that attractiveness ratings influenced assumed similarity. As such, Study 5 provides further evidence for an interesting boundary condition of assumed similarity on D, suggesting that assumed similarity on D not only is more than a simple expression of (dis)trust but also has a motivational component. Nevertheless, there were some inconsistencies in our findings when compared to Study 4. We did not find a significant assumed similarity effect nor a main effect of interpersonal attraction. This could be explained by several methodological differences. Namely, whether interpersonal attraction was measured before or after observer-reported D, our measure of interpersonal attraction (liking vs. physical attractiveness), or the experimental manipulation in Study 5 which meant that no participant rated moderately attractive targets (but only targets of high or low attractiveness).

8. General discussion

Navigating social dilemmas in terms of cooperation or defection is important for the functioning of societies. Social dilemmas afford the expression of exploitativeness and distrust (Thielmann et al., 2020b), and the general disposition toward both these expressions are argued to be captured by the Dark Factor of Personality, D (Hilbig et al., 2022; Moshagen et al., 2018). Consequently, this implies that those high on D will assume others are likewise willing to maximize their utility by exploiting others. We tested this conjecture through the lens of person perception, namely, the phenomenon of assumed similarity (Cronbach, 1955). Specifically, across five studies, we examined whether people assume similarity on D to test the substantive definition of D that exploitativeness and distrust are two sides of the same (personality) coin.

9. Assumed similarity on D

Across studies, we found that people's judgments of unacquainted others on D were positively related to their self-perception on this very characteristic. These findings are consistent with the notion that D not only comprises the tendency to engage in socially/ethically aversive behavior, but also the beliefs that serve as justifications of this behavior (Moshagen et al., 2018). Specifically, we had reasoned that those high on D would be more likely to believe that others are untrustworthy as this can serve to justify their own exploitation of others (Hilbig et al., 2022). Indeed, these distrust-related beliefs are a critical element of the substantive theory of D that sets it apart from other traits capturing individual differences in aversive behavior (e.g., honesty–humility; Ashton et al., 2014). Thus, our finding of assumed similarity on D is in line with the conceptualization of D as comprising dispositions toward exploitativeness *and* distrust and in line with the finding that D accounts for behavior both in economic games that allow for the expression of exploitativeness and those that tap into distrust (Hilbig and Thielmann, 2025).

We also found evidence against the influence of spurious similarity on assumed similarity on D. Despite gender differences in D that could have contributed to spurious similarity (Hartung et al., 2022), congruency in sex between raters and targets did not moderate assumed similarity on D. Nevertheless, we did find evidence for (low) interpersonal attraction as a boundary condition: Assumed similarity on D was weaker when rating targets who were less liked and perceived as less physically attractive. On one hand, these findings are consistent with the theory that D comprises the conjunction of exploitativeness and distrust if we assume that attractive targets engender distrust from those high on D and trust from those low on D. On the other hand, weaker assumed similarity on D when rating interpersonally unattractive targets does not directly follow from the theory of D. Rather, it suggests that assumed similarity on D also has a motivational component—as has been argued for assumed similarity on other personality traits as well (e.g., Collisson and Howell, 2014).

https://doi.org/10.1017/jdm.2025.10013 Published online by Cambridge University Press

Indeed, the moderating effect of interpersonal attraction is consistent with theories arguing that individuals assume similarity with others because they wish to highlight commonalities and belongingness with attractive others (Machunsky et al., 2014; Marks and Miller, 1987), and to maintain cognitive balance (Heider, 1958). Our results suggest that part of the assumed similarity effect on D also serves this purpose. Our findings have implications for research that finds people are more generous to attractive individuals in economic games (Shang and Zhang, 2024). Specifically, those that are higher on D would be less likely to cooperate with attractive targets because they judge attractive targets to be more similar to themselves and thus also higher in D and thereby more untrustworthy. Hence, the relationship between cooperation and attractiveness may be conditional on D. Nevertheless, future research is required to test the role of assumed similarity of D in the tendency to cooperate with more attractive individuals.

10. *The value account of assumed similarity*

Our finding that people assume similarity on D is also consistent with the value account of assumed similarity, which posits that people assume similarity of these traits because they are linked to values that people care about (Lee et al., 2009; Thielmann et al., 2023; Thielmann et al., 2020a). Given that D is associated with self-transcendence versus self-enhancement values (García-Fernández et al., 2025; Moshagen et al., 2018), our findings provide further support of the value account of assumed similarity. Indeed, our focus on D proverbially shines a light on a gap in the literature, as little research has looked at the ‘dark side’ of the value-related trait continuum when studying assumed similarity.

11. *Limitations and future directions*

Although our research and findings have important implications for the theory of D and assumed similarity, the following limitations should be acknowledged. First, we only focused on two potential moderators of assumed similarity on D. In addition to sex congruency and interpersonal attraction, there are other perceiver and target characteristics that could potentially moderate assumed similarity on D. For example, assumed similarity has been found to increase with relationship closeness even after controlling for actual similarity (Thielmann and Hilbig, 2022). Thus, to further test the robustness of the assumed similarity effect on D, future research should test other perceiver–target characteristics such as relationship closeness.

Second, our research was limited by the fact that we focused on judgments of strangers that were displayed in photos—with the exception of Study 1. Although this focus had methodological benefits (e.g., being able to systematically vary target information), most real-life interactions involve more than just viewing a picture of an unknown individual, and whether assumed similarity holds when assessing acquainted targets on D is an open question. Thus, future research is needed to examine whether assumed similarity on D extends to situations with more involved interactions (e.g., brief interactions with targets) or, similar to the first limitation mentioned, judgments of known others.

Third, a limitation of our research is the lack of direct evidence for the specific underlying mechanisms driving assumed similarity. As is common for research on assumed similarity, it is not possible to determine whether participants judge their own characteristics to be common (i.e., the false-consensus effect), project unique characteristics onto others (i.e., social projection), want to think that others share their values (i.e., motivational mechanism), or some combination of these. Possibly, participants even anchored their subsequent responses to their initial responses (given that participants always self-reported D before completing corresponding observer reports). Somewhat reassuringly, prior research has found evidence of assumed similarity even when participants completed observer reports before self-reports of honesty–humility (e.g., Thielmann et al., 2020a). Moreover, we found the assumed similarity effect in Study 4 despite a 2-year interval between self-reported D and observer-reported D which ought to be sufficiently long to rule out anchoring. Nevertheless, future research

https://doi.org/10.1017/jdm.2025.10013 Published online by Cambridge University Press

should counterbalance the order in which participants complete self- and observer reports of D to test for anchoring effects.

Finally, our reliance on correlational data and because we did not connect our findings to actual behavior in social dilemmas limits our interpretations. Specifically, causal interpretation such as assumed similarity on D serving as a justification for individuals' own aversive behavior or influencing cooperation in social dilemmas, should be interpreted with caution.

While D is theorized to be the common core of all socially aversive traits, examining assumed similarity on different themes of D (Bader et al., 2021) might yield a more nuanced picture. For instance, although Webster and Campbell (2023) found assumed similarity on all four dark tetrad traits, this effect was stronger for machiavellianism and sadism than for narcissism and psychopathy. Hence, a promising avenue of future research could be to explore the relative strength of assumed similarity for different themes of D.

12. Conclusion

The conjunction of exploitativeness and distrust is argued to be captured by D which implies that people would assume others have similar levels on D. Indeed, we found that people do assume similarity on D when rating unacquainted others and that this effect was unaffected by sex congruency between raters and targets. These findings support the notion that D jointly captures tendencies for exploitation and distrust of others. However, our findings that assumed similarity on D is weaker for less interpersonally attractive targets suggests that (low) interpersonal attraction might be a boundary condition for assumed similarity on D. In conclusion, people appear to assume similarity on D consistent with D comprising the tendency to exploit and be distrustful of others, but this effect seems not to hold when judging others one feels unattracted to. Thus, assumed similarity on D seems to serve at least two purposes: justifying one's (im)moral behavior and feeling connected with interpersonally attractive others.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/jdm.2025.10013.

Acknowledgments. We thank all members of the Personality, Identity, and Crime team for their invaluable advice at various points in the research process. We are also grateful to Moritz Pischel for their help formatting the tables included in the manuscript and supplemental materials.

Data availability statement. All materials, data, and analysis scripts used in the present research can be accessed at https://osf.io/t4s3u/.

Competing interest. We declare no competing interests.

References

Aron, A., & Lewandowski, G. (2001). Interpersonal attraction, psychology of. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 7860–7862). Pergamon. https://doi.org/10.1016/B0-08-043076-7/01787-3.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*(2), 150–166. https://doi.org/10.1177/1088868306294907.

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*(2), 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.

Bader, M., Hartung, J., Hilbig, B. E., Zettler, I., Moshagen, M., & Wilhelm, O. (2021). Themes of the dark core of personality. *Psychological Assessment*, *33*(6), 511–525. https://doi.org/10.1037/pas0001006.

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, predictive validity, and self-other agreement. *Journal of Personality Assessment*, *104*(5), 660–673. https://doi.org/10.1080/00223891.2021.1984931.

Balliet, D., & Van Lange, P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, *139*(5), 1090–1112. https://doi.org/10.1037/a0030939.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Batres, C., & Shiramizu, V. (2023). Examining the "attractiveness halo effect" across cultures. *Current Psychology*. *42*(29), 25515–25519.

Bruins, J. J., Liebrand, W. B. G., & Wilke, H. A. M. (1989). About the saliency of fear and greed in social dilemmas. *European Journal of Social Psychology*, *19*(2), 155–161. https://doi.org/10.1002/ejsp.2420190207.

Cohen, J. (1988). *Statistical power analysis for the Behavioral sciences*. Routledge Academic.

Collisson, B., & Howell, J. L. (2014). The liking-similarity effect: Perceptions of similarity as a function of liking. *The Journal of Social Psychology*, *154*(5), 384–400. https://doi.org/10.1080/00224545.2014.914882.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity. *Psychological Bulletin*, *52*, 177–193. https://doi.org/10.1037/h0044919.

DeYoung, C. G. (2015). Openness/intellect: A dimension of personality reflecting cognitive exploration. In M. Mikulincer, P. R. Shaver, M. L. Cooper, & R. J. Larsen (Eds.), *APA handbook of personality and social psychology, Vol. 4. Personality processes and individual differences* (pp. 369–399). American Psychological Association. https://doi.org/10.1037/14343-017.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the big five. *Journal of Personality and Social Psychology*, *93*(5), 880. https://doi.org/10.1037/0022-3514.93.5.880.

García-Fernández, J., Postigo, Á., González-Nuevo, C., Cuesta, M., & Moshagen, M. (2025). Unethical vs. human values: Relationship between the D factor and Schwartz's theory of human values. *Journal of Individual Differences*. *46*(2), 89–96. https://econtent.hogrefe.com/doi/10.1027/1614-0001/a000439.

Hartung, J., Bader, M., Moshagen, M., & Wilhelm, O. (2022). Age and gender differences in socially aversive ("dark") personality traits. *European Journal of Personality*, *36*(1), 3–23. https://doi.org/10.1177/0890207020988435.

Heider, F. (1958). Perceiving the other person. In *The psychology of interpersonal relations* (pp. 20–58). John Wiley & Sons Inc. https://doi.org/10.1037/10628-002.

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), 2. 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*(11), 2730–2755. https://doi.org/10.1037/xge0001232.

Hilbig, B. E., & 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*, *20*, e16. https://doi.org/10.1017/jdm.2025.1.

Hilbig, B. E., Thielmann, I., Zettler, I., & Moshagen, M. (2023). The dispositional essence of proactive social preferences: The dark core of personality vis-à-vis 58 traits. *Psychological Science*, *34*(2), 201–220. https://doi.org/10.1177/09567976221116893.

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.

Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., & Van Lange, P. A. M. (2003). *An atlas of interpersonal situations*. Cambridge University Press. https://doi.org/10.1017/CBO9780511499845.

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

Klebl, C., Rhee, J. J., Greenaway, K. H., Luo, Y., & Bastian, B. (2022). Beauty goes down to the core: Attractiveness biases moral character attributions. *Journal of Nonverbal Behavior*, *46*(1), Article 1. https://doi.org/10.1007/s10919-021-00388-w.

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.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Lee, K., Ashton, M. C., Pozzebon, J. A., Visser, B. A., Bourdage, J. S., & Ogunfowora, B. (2009). Similarity and assumed similarity in personality reports of well-acquainted persons. *Journal of Personality and Social Psychology*, *96*, 460–472. https://doi.org/10.1037/a0014059.

Long, J. A. (2024). jtools: Analysis and presentation of social scientific data. *Journal of Open Source Software*, *9*(101), 6610.

Lüdecke, D. (2023). *sjPlot: Data visualization for statistics in social science. R package version 2.8. 10. sjPlot: data visualization for statistics in social science. R package version 2.8. 14*.

Machunsky, M., Toma, C., Yzerbyt, V., & Corneille, O. (2014). Social projection increases for positive targets: Ascertaining the effect and exploring its antecedents. *Personality and Social Psychology Bulletin*, *40*(10), 1373–1388. https://doi.org/10.1177/0146167214545039.

Marks, G., & Miller, N. (1982). Target attractiveness as a mediator of assumed attitude similarity. *Personality and Social Psychology Bulletin*, *8*, 728–735. https://doi.org/10.1177/0146167282084020.

https://doi.org/10.1017/jdm.2025.10013 Published online by Cambridge University Press

Marks, G., & Miller, N. (1987). Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin*, *102*(1), 72–90. https://doi.org/10.1037/0033-2909.102.1.72.

Marks, G., Miller, N., & Maruyama, G. (1981). Effect of targets' physical attractiveness on assumptions of similarity. *Journal of Personality and Social Psychology*, *41*(1), 198–206. https://doi.org/10.1037/0022-3514.41.1.198.

Mashman, R. C. (1978). The effect of physical attractiveness on the perception of attitude similarity. *The Journal of Social Psychology*, *106*(1), 103–110. https://doi.org/10.1080/00224545.1978.9924150.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633–644. https://doi.org/10.1509/jmkr.45.6.633.

Miyake, K., & Zuckerman, M. (1993). Beyond personality impressions: Effects of physical and vocal attractiveness on false consensus, social comparison, affiliation, and assumed and perceived similarity. *Journal of Personality*, *61*(3), 411–437. https://doi.org/10.1111/j.1467-6494.1993.tb00287.x.

Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, *125*, 656–688. https://doi.org/10.1037/rev0000111.

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

Moshagen, M., Zettler, I., Horsten, L. K., & Hilbig, B. E. (2020b). Agreeableness and the common core of dark traits are functionally different constructs. *Journal of Research in Personality*, *87*, 103986. https://doi.org/10.1016/j.jrp.2020.103986.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, *66*(6), 1025–1060. https://doi.org/10.1111/1467-6494.00041.

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), Article 6. https://doi.org/10.1016/S0092-6566(02)00505-6.

Paunonen, S. V., & Hong, R. Y. (2013). The many faces of assumed similarity in perceptions of personality. *Journal of Research in Personality*, *47*(6), 800–815. https://doi.org/10.1016/j.jrp.2013.08.007.

Prentice, M., Jayawickreme, E., Hawkins, A., Hartley, A., Furr, R. M., & Fleeson, W. (2019). Morality as a basic psychological need. *Social Psychological and Personality Science*, *10*(4), 449–460. https://doi.org/10.1177/1948550618772011.

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (Version 4.3.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, *18*(5), 429–434. https://doi.org/10.1111/j.1467-9280.2007.01917.x.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology* (Vol. 25, pp. 1–65). Academic Press.

Shang, J., & Zhang, Y. (2024). Influence of male's facial attractiveness, vocal attractiveness and social interest on female's decisions of fairness. *Scientific Reports*, *14*(1), 16778. https://doi.org/10.1038/s41598-024-67841-w.

Thielmann, I., Böhm, R., Ott, M., & Hilbig, B. E. (2021). Economic games: An introduction and guide for research. *Collabra: Psychology*, *7*(1), 19004. https://doi.org/10.1525/collabra.19004.

Thielmann, I., & Hilbig, B. E. (2022). Assumed similarity. In *Cognitive illusions* (pp. 272–286). Routledge.

Thielmann, I., & Hilbig, B. E. (2023). Generalized dispositional distrust as the common Core of populism and conspiracy mentality. *Political Psychology*, *44*(4), 4. https://doi.org/10.1111/pops.12886.

Thielmann, I., Hilbig, B. E., & Zettler, I. (2020a). Seeing me, seeing you: Testing competing accounts of assumed similarity in personality judgments. *Journal of Personality and Social Psychology*, *118*, 172–198. https://doi.org/10.1037/pspp0000222.

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

Thielmann, I., Rau, R., & Locke, K. D. (2023). Trait-specificity versus global positivity: A critical test of alternative sources of assumed similarity in personality judgments. *Journal of Personality and Social Psychology*, *124*(4), 828–847. https://doi.org/10.1037/pspp0000420.

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

Webster, G. D., & Campbell, J. T. (2023). Personality perception in game of thrones: Character consensus and assumed similarity. *Psychology of Popular Media*, *12*(2), 207–218. https://doi.org/10.1037/ppm0000398.

Zell, E., Strickhouser, J. E., Sedikides, C., & Alicke, M. D. (2020). The better-than-average effect in comparative self-evaluation: A comprehensive review and meta-analysis. *Psychological Bulletin*, *146*(2), Article 2. https://doi.org/10.1037/bul0000218.

Cite this article: Tan, N. P.-J., Hilbig, B., Moshagen, M., Zettler, I., Payer, S., and Thielmann, I. (2025). Strangers in the dark: assumed similarity in judgments of unknown others on aversive personality. *Judgment and Decision Making*, e38. https://doi.org/10.1017/jdm.2025.10013