Examining the accuracy of lay beliefs about the effects of personality on prosocial behavior

Stavrova, Olga; Evans, Anthony M.; Sleegers, Willem; van Beest, Ilja

Published in:
Journal of Behavioral Decision Making

DOI:
10.1002/bdm.2282

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Received: 7 September 2021 | Revised: 17 February 2022 | Accepted: 24 March 2022

DOI: 10.1002/bdm.2282

RESEARCH ARTICLE

WILEY

Examining the accuracy of lay beliefs about the effects of personality on prosocial behavior

Olga Stavrova[1] | Anthony M. Evans[2] | Willem Sleegers[3] | Ilja van Beest[1]

[1]Department of Social Psychology, Tilburg University, Tilburg, The Netherlands

[2]Allstate Corporation, Northfield Township, Illinois, USA

[3]Rethink Priorities, San Francisco, California, USA

Correspondence
Olga Stavrova, Tilburg University, Warandelaan 2, 5000 LE Tilburg, The Netherlands.
Email: o.stavrova@uvt.nl

Abstract

Prior research on personality and prosocial behavior has focused on actor-level effects of personality by examining which personality traits predict individuals' prosocial behavior. But do lay people take into account others' personality when making predictions of others' future prosocial behavior? The present research was designed to answer this question. We focused on two interpersonal traits from the Big Five model, agreeableness and extraversion, and examined whether people have accurate lay beliefs about the effects of these traits on prosocial behavior. The results of four studies showed that participants consistently attributed agreeableness (and to a lesser extent, extraversion) a greater role in predicting others' prosocial behavior compared with the role that it plays in reality. Results were consistent in studies of zero-acquaintance interactions and close relationships and when people predicted both single instances and aggregated measures of others' prosocial behavior. Our results did not depend on participants' awareness of research hypotheses and persisted even when they were explicitly warned that the information about others' personality might not be accurate. These findings inform the literature on social perception and stereotype accuracy and contribute to our understanding of how people make future-oriented predictions of others' behavior.

KEYWORDS
agreeableness, cooperation, prosocial behavior, social dilemmas, social perception, stereotype accuracy, trust, trustworthiness

1 | INTRODUCTION

What factors predict prosocial behavior (e.g., trustworthiness, help, and cooperation) among strangers? Research suggests that individual differences in personality are key to answering this question (Volk et al., 2012). Prior research on individual differences in social dilemmas, and prosocial behavior more broadly, has focused on the actor-level effects of personality traits; for example, previous research asked which personality traits are significantly correlated with individual decisions to cooperate with strangers, or provide help to others in times of need (Balliet et al., 2009; Hilbig & Zettler, 2009; Thielmann et al., 2020). With some exceptions (Cooper et al., 2015; Kugler et al., 2014), the exploration of *interpersonal* (or partner-level) consequences of personality traits has not received much attention. What traits do people value (in others) when faced with dilemmas of help and cooperation, and do people have accurate lay beliefs about the effects of traits on prosocial behavior? The present research was designed to answer these questions.

We focus on personality inferences related to two interpersonal dimensions of the Big Five: extraversion and agreeableness (McCrae & Costa, 1989; Tov et al., 2016). Extraversion plays an important role in success in social life: Extraverts are more popular in their social networks (Feiler & Kleinbaum, 2015), are more liked and praised by others, including their colleagues and bosses (Judge et al., 2002). At the same time, when making first impressions, it is others' agreeableness, rather than extraversion, that people are mostly eager to learn about (Ames & Bianchi, 2008) and agreeableness seems to be particularly important as a predictor of prosocial behavior (Graziano et al., 1997). Here, we ask how perceptions of others' extraversion and agreeableness influence expectations of others' prosocial behavior. Importantly, we explore to what extent people's expectations about the prosociality of agreeable (versus disagreeable) and extraverted (versus introverted) others are in line with reality. In other words, we investigate whether people over-rely on others' agreeableness and extraversion as predictors of prosocial behavior.

2 | INTERPERSONAL EFFECTS OF PERSONALITY ON PROSOCIAL BEHAVIOR

Personality traits are important for how people perceive and evaluate each other (Back et al., 2011). We focus on two traits that represent the main interpersonal/social dimensions of the Big Five (McCrae & Costa, 1989; Tov et al., 2016) and could affect individuals' expectations of others' prosociality: agreeableness and extraversion.

When it comes to prosocial behavior, agreeableness stands out as a potential predictor. Agreeableness implies prosociality by definition: Agreeable individuals are described as "having a soft heart," "taking time out for others," and "sympathizing with others' feelings" (Goldberg, 1992). Not surprisingly, higher levels of agreeableness predict likeability, peer acceptance, and higher chances of being selected as friends (Jensen-Campbell et al., 2002; Selfhout et al., 2010; Stopfer et al., 2014; van der Linden et al., 2010). In economic games, providing information related to other player's agreeableness shapes individuals' decisions: For example, people are less likely to take on risks when playing against partners described as more (vs. less) aggressive (Kugler et al., 2014). More (vs. less) agreeable peers are perceived as more trustworthy in interdependent work groups (Naber et al., 2018) and being matched with a disagreeable partner in a public goods game leads to less cooperation (at least, among agreeable individuals; Drouvelis & Georgantzis, 2019). Finally, higher levels of agreeableness are positively associated with prosocial behaviors (Graziano et al., 2007) and to the extent that lay people are aware of these links, they will expect more (vs. less) agreeable others to display more prosocial behaviors. Taken together, these findings suggest that people might expect more prosociality from agreeable (vs. disagreeable) others.

Agreeableness, however, is not the only Big Five trait relevant to interpersonal behavior. Others' extraversion might also be important in shaping expectations of prosociality. Previous research has highlighted that higher levels of extraversion are associated with peer acceptance, sociometric popularity, and likeability in children and adolescents (Jensen-Campbell et al., 2002; Stopfer et al., 2013; van der Linden et al., 2010; Wolters et al., 2014). At the same time, social dominance and "getting ahead" motivations might render extraverts more prone to conflicts and bullying (Bono et al., 2002; Mitsopoulou & Giovazolias, 2015). Indeed, people believe some facets of extraversion (assertiveness, excitement seeking) are negatively related to cooperation (Cooper et al., 2015). Taken together, these findings suggest that extraversion might have both positive and negative effects on expectations of prosociality. Therefore, while we expected extraversion to affect expectations of prosociality, we did not have a prediction regarding the direction of this effect, assuming that people could be equally likely to expect extraverts to be more or less prosocial than introverts.

3 | CALIBRATING EXPECTATIONS OF PROSOCIALITY

People are prone to errors when making predictions about others' prosociality. For example, people overestimate the extent to which financial incentives boost blood donations (Miller & Ratner, 1998) and hold inaccurate beliefs about the effects of others' emotions on their trustworthiness (Kausel & Connolly, 2014). Here, we explore whether people err by over-relying on others' personality traits. People might generally over-estimate the importance of others' personality when making predictions about their prosocial behavior. For example, studies on the Fundamental Attribution Error (FAE) showed that when presented with others' behavior, people tend to ignore potential situational explanations and instead attribute behavior to dispositional causes (Gilbert & Malone, 1995; Ross, 1977; Ross & Nisbett, 1991). If people make (exaggerated) inferences about others' dispositions from observing their behavior, they might also over-use their beliefs about others' personality traits (such as extraversion and agreeableness) when making predictions of their future prosocial behavior.

The question of whether people over-estimate or under-estimate the prosociality of more or less agreeable and extraverted others is also related to work on social perception and stereotyping (Jussim et al., 2016). This literature has explored to what extent people's beliefs about groups correspond to actual group differences. People do not necessarily perceive the differences between individuals with different characteristics, such as gender, social class, or cultural background, as larger than they really are (Jussim et al., 2015). In fact, it has been suggested that people are even more likely to under- rather than over-estimate group differences (Jussim et al., 2015). At the same time, when the predicted behavior (e.g., prosociality) represents a particularly accessible or even defining attribute of a psychographic group (e.g., agreeable people), people might over-estimate this attribute's importance (Epley & Eyal, 2019). For example, people tend to exaggerate gender differences in attributes that are more central to

gender stereotypes (e.g., mind reading ability) than less central (e.g., happiness) (Eyal & Epley, 2017). Also, people believe that others described as high in warmth, and low in excitement-seeking, assertiveness and angry-hostility are more likely to show cooperative behaviors in economic games, while in reality, these traits were unrelated to cooperation (Cooper et al., 2015). Hence, in predicting others' prosociality, people might be more likely to over-rely on the traits most central to social behavior – agreeableness and extraversion (McCrae & Costa, 1989; Tov et al., 2016), than on other less central traits (openness, conscientiousness, neuroticism).

4 | THE PRESENT RESEARCH

We examined whether people overweight others' agreeableness and extraversion when predicting their prosocial behavior. In Study 1, using a fully incentivized real-time prisoner's dilemma, we examined people's expectations regarding the cooperation of agreeable (vs. disagreeable) and extraverted (vs. introverted) partners, as well as actual cooperation rates of agreeable (vs. disagreeable) and extraverted (vs. introverted) individuals. In Study 2, we tested whether people overestimate the role of agreeableness and extraversion when predicting others' trustworthy behavior in a trust game and explored the potential role of the experimenter demand effects. In Study 3, we examined whether people overestimate the prosociality of agreeable and extraverted partners when predicting both single decisions as well as average decisions across different social dilemma situations. In Study 4, we explored predictions of prosociality in the absence of explicit personality information by studying whether individuals who know each other (romantic partners) overestimate how much each other's personality shapes their prosocial behavior. In all studies, we focused on the two interpersonal dimensions of the Big Five—agreeableness and extraversion; additionally, to test the specificity of our predictions, Studies 1 and 4 included all the Big Five traits.

Study 1 was exploratory; Studies 2–4 were preregistered. All materials, data, and computer code can be accessed at the study's OSF page: https://osf.io/xu4mz/?view_only=1ea2d4a97eac42e8bece0ccba591eb20.

5 | STUDY 1

In Study 1, participants completed all measures of the Big Five, were randomly assigned to dyads, and played a fully incentivized real-time prisoner's dilemma. Before deciding to cooperation or defect, we provided participants with a complete picture of their partner's personality, including their standing on all the Big Five traits. This design allowed us to examine whether participants give the information about their partner personality more weight than warranted by reality: for example, by choosing to cooperate with more (vs. less) agreeable partners to a greater extent than warranted by the actual cooperation rate displayed by more (vs. less) agreeable partners.

5.1 | Method

5.1.1 | Design and procedure

Participants were recruited on MTurk and completed the study via oTree, an online platform for conducting controlled behavioral experiments with multiple participants in real time (Chen et al., 2016). After participants were paired another player, they completed the Big Five scales using the mini-IPIP inventory (Donnellan et al., 2006) that assesses each Big Five trait with four items, using a 5-point response scale (1 = *very inaccurately*, 5 = *very accurately*). All scales showed good reliabilities (α between .80 and .86). After completing the scales, both players were shown a brief description of each trait (see Supporting Information) and their own scores on all five traits, in percentiles, for example, "Your score in extraversion is higher than X% of people." Percentiles were based on a representative U.S. sample that completed the mini-IPIP (Donnellan, personal communication). The order in which the five traits were presented was randomized.

Next, participants were introduced to a prisoner's dilemma. They learned that they would be randomly paired with another participant and that both of them would simultaneously and privately choose Keep or Transfer. Their payoff was determined as a function of their decisions (see Table 1). Ten points equaled $0.10.

Before making their decision, participants were shown their partner's scores on the Big Five traits (e.g., "Your partner's score in extraversion is higher than X% of people"), and they also learned that their scores will be shown to their partner too. Participants then responded to the two key dependent variables. First, they made their Keep (= 0) versus Transfer (= 1) decision. Second, using an 11-point scale, they indicated how likely it was that their partner chose Transfer (1 = *Definitely chose Keep*, 10 = *Definitely chose Transfer*). At the end, participants were informed about their partner's decision and, consequently, their payoff, and filled in basic socio-demographic information. All participants received their payoff within a couple of days using the MTurk bonus system.

5.1.2 | Participants

We recruited 382 (191 pairs) participants. Of those, 59 had missing values on the key variables and were removed, resulting in a final sample of 323 participants (M_{age} = 35.68, SD_{age} = 11.24, 45.8% female). Participants received $2 as a compensation for participation and could earn bonus payment, depending on their and their partner's decisions

TABLE 1 Pay-off structure for the dyadic prisoner's dilemma, Study 1

		The other participant	
		Transfer	**Keep**
You	Transfer	20 points, 20 points	0 point, 30 points
	Keep	30 points, 0 point	10 points, 10 points

TABLE 2 Means, standard deviations, and zero-order correlations, Study 1

		M	*SD*	1	2	3	4	5	6	7	8	9	10	11
1	Cooperation	0.52	0.50	-	-	-	-	-	-	-	-	-	-	-
2	Cooperation expectation	4.70	2.66	.53***	-	-	-	-	-	-	-	-	-	-
3	Partner agreeableness	3.72	0.97	.14*	.27***	-	-	-	-	-	-	-	-	-
4	Partner extraversion	2.65	1.08	.09	.16**	.27***	-	-	-	-	-	-	-	-
5	Partner openness	3.96	0.94	.09	.17**	.31***	.20***	-	-	-	-	-	-	-
6	Partner conscientiousness	3.75	0.87	.10+	.12*	.19***	.11*	.16***	-	-	-	-	-	-
7	Partner emotional stability	3.63	1.01	.05	.12*	.09	.29***	.16**	.45***	-	-	-	-	-
8	Actor agreeableness	3.74	0.97	.05	−.02	.03	.07	.00	−.02	−.03	-	-	-	-
9	Actor extraversion	2.62	1.08	−.14*	−.01	.11*	.08	.02	−.04	−.06	.25***	-	-	-
10	Actor openness	3.97	0.95	.02	−.04	.03	.02	.10+	.03	.05	.32***	.21***	-	-
11	Actor conscientiousness	3.77	0.88	−.05	−.04	−.04	−.04	.05	.10+	.08	.14*	.10+	.15***	-
12	Actor emotional stability	3.63	1.00	−.11+	−.05	−.05	−.04	.06	.06	.07	.10+	.28***	.16**	.46***

Note: Actor and partner Big Five scores represent self-reports.
+$p < .10$.
*$p < .05$. **$p < .01$. ***$p < .001$.

in the game. A sensitivity power analysis conducted with g*power 3.1 (Faul et al., 2009) showed that this sample size would allow us to detect a correlation of $r = .16$ with a power of 80% and $\alpha = .05$ (two-tailed test).

5.2 | Results

5.2.1 | Zero-order correlations

Descriptive information and zero-order correlations are shown in Table 2. Participants were more likely to cooperate with more (vs. less) agreeable partners ($r = .14$, $p = .013$, 95% CI [0.03, 0.25]). In contrast, partner extraversion was not significantly related to participants' cooperation decision ($r = .09$, $p = .10$, 95% CI [−0.02, 0.20]).

Participants' expectations regarding their partner's cooperation showed positive associations with all of their partner's Big Five traits, with the values ranging from $r = .12$, $p = .03$, 95% CI [0.01, 0.23] (conscientiousness and emotional stability), to $r = .27$, $p < .001$, 95% CI [0.17, 0.37] (agreeableness).

Participants' own Big Five scores were not related to their own cooperation decision, except for extraversion, with more extraverted people being less likely to cooperate ($r = -.14$, $p = .014$, 95% CI [−0.25, −0.03]).

5.2.2 | Dyadic analyses

Next, to examine the effects of both actor and partner traits on actor cooperation simultaneously, we used an actor-partner interdependence model (APIM; Kenny et al., 2006). APIM allows estimating the effects of the characteristics of both members of a dyad on each member's decision, while taking into account the non-independence of observations resulting from a dyadic data structure. As our data included indistinguishable dyads, we opted for a multilevel regression with participants nested within dyads. Following the recommendations in the literature (Kenny & Kashy, 2011), we allowed for correlated errors among members of the same couple. We used gls function of the nlme package for continuous outcome (cooperation expectation) and glmmPQL function of the MASS package for binary outcome (cooperation) in R (Venables & Ripley, 2002). The models included a random intercept at the level of dyads. The results of these analyses are shown in Table 3.

Expectations of cooperation

Participants expected agreeable and extraverted partners to be more cooperative than disagreeable and introverted partners (Model 1: $b = 0.75$, $p < .001$, 95% CI [0.46, 1.04] and Model 2: $b = 0.41$, $p = .003$, 95% CI [0.14, 0.68]). Partner openness, conscientiousness, and emotional stability were also positively associated with expectations of cooperation (Tables 2 and S1). However, when all actor and

TABLE 3 Dyadic (multilevel) model results, Study 1

	DV: actor expectations of partner cooperation, *b*/β		
	Model 1	**Model 2**	**Model 3**
Fixed effects			
Partner characteristics			
Agreeableness	0.75/.27***	-	0.59/.21***
Extraversion	-	0.41/.17**	0.17/.07
Openness	-	-	0.23/.08
Conscientiousness	-	-	0.13/.04
Emotional stability	-	-	0.14/.05
Actor characteristics			
Agreeableness	-	-	−0.03/−.01
Extraversion	-	-	−0.02/−.01
Openness	-	-	−0.12/−.04
Conscientiousness	-	-	−0.05/−.02
Emotional stability	-	-	−0.08/−.03
	DV: actor cooperation, odds ratios		
	Model 1	**Model 2**	**Model 3**
Fixed effects			
Partner characteristics			
Agreeableness	1.32*	-	1.26+
Extraversion	-	1.22+	1.12
Openness	-	-	1.10
Conscientiousness	-	-	1.21
Emotional stability	-	-	0.95
Actor characteristics			
Agreeableness	-	-	1.21
Extraversion	-	-	0.73**
Openness	-	-	1.07
Conscientiousness	-	-	0.95
Emotional stability	-	-	0.87

Note: β were obtained by standardizing the variables.
+$p < .10$.
*$p < .05$. **$p < .01$. ***$p < .001$.

partner Big Five traits were included, only partner agreeableness predicted cooperation expectations (Model 3: $b = 0.59$, $p < .001$, 95% CI [0.27, 0.91]).

Cooperation

Partner agreeableness was associated with a higher likelihood of actor cooperation (Model 1: OR = 1.32, $p = .017$, 95% CI [1.05, 1.68]), while the effect of partner extraversion was close to significance (Model 2: OR = 1.22, $p = .059$, 95% CI [0.99, 1.51]). Model 3 included the remaining partner and actor Big Five scores. Of all partner traits, only agreeableness was close to significance (OR = 1.26, $p = .087$, 95% CI [0.97, 1.65]). Of the actor traits, only actor extraversion (negatively) predicted actor cooperation (OR = 0.73, $p = .011$, 95% CI [0.58, 0.92]): More extraverted individuals were less likely to cooperate.

5.3 | Discussion

In a fully incentivized, real-time prisoner's dilemma, participants believed agreeable (and, at zero-order level, extraverted) partners to be more likely to cooperate. In reality however, more agreeable partners were not more likely to cooperate than less agreeable partners, and more extraverted partners were even less likely to cooperate than less extraverted partners. This provides initial evidence of a mismatch

between lay beliefs about the role of traits in prosocial (e.g., cooperation) behavior and the reality.

6 | STUDY 2

Study 2 made several contributions. First, the mismatch between lay beliefs and reality observed in Study 1 could be driven by (1) participants' lack of understanding that, in reality, personality scores are subject to measurement error and can therefore lack accuracy and/or (2) experimenter demand effects. Therefore, in Study 2, when providing participants with personality information about the target, we explicitly warned them that this information was not perfectly accurate. In addition, to rule out experimenter demand effects, we tested whether the effect of target personality on expectations of prosociality depends on participants' awareness of our hypotheses.

Second, in Study 1, the description of trait agreeableness provided to the participants did not fully reflect the mini-IPIP items that were used to measure agreeableness. Also, before completing the personality measures, participants learned that their scores (although anonymized) would be revealed to other player, which could have induced socially desirable responding. Taken together, these features could have contributed to a mismatch between expectations and reality. In Study 2, we addressed these limitations: by using standardized instructions to measure personality ("Describe yourself as you honestly see yourself") and we used identical wording in the measures presented to targets and the descriptions provided to raters.

Third, we extended the investigation to another social dilemma: the trust game. We recruited a group of trustors and a group of trustees. Trustors indicated how much money they expected agreeable (vs. disagreeable) and extraverted (vs. introverted) trustees to send back to them in a trust game. Trustees filled in measures of agreeableness and extraversion and indicated how much money they decided to send back to trustors. We compared trustors' expectations regarding how much money agreeable (vs. disagreeable) and extraverted (vs. introverted) trustees will send back to them with the actual amounts of money that agreeable (vs. disagreeable) and extraverted (vs. introverted) trustees decided to send.

Measures, data collection, and analyses were preregistered: https://aspredicted.org/47V_H4T.

6.1 | Method

6.1.1 | Design and procedure

Participants were asked to imagine that they were matched with another participant of this study in a decision-making task (the trust game). The trustor's decision was binary and the trustee's decision was continuous (ranging from £0 to £3). The game instructions are presented in the Supporting Information. Participants were randomly assigned to the role of either trustors or trustees.

Trustors

Trustors read the trust game instructions and answered a comprehension check question (see Supporting Information). Before making their decision, trustors learned about two fictional traits: v-/z-dominance and front-/back-brainedness (we used fictional traits to avoid demand effects; the labels were borrowed from Critcher et al., 2015). v-/z-dominance described individual differences in extraversion, and front-/back-brainedness described individual differences in agreeableness. Specifically, participants learned that "people high in Z-dominance are more likely to be sociable and outgoing, to be full of energy, and to generate a lot of enthusiasm than people low in Z-dominance," while "people high in Front-brainedness are more likely to be helpful and unselfish with others, to have a forgiving nature, and less likely to be cold or aloof than people low in Front-brainedness." These descriptions were developed based on the items used as indicators of agreeableness and extraversion in the BFI-44 inventory (John & Srivastava, 1999). Trustors further read that "although psychological tests do pretty well at determining one's level of Front-brainedness and Z-dominance, they are not 100% accurate."

Trustors were then asked to estimate the amount of money Person 2 would send them back (between 0 and $3) if Person 2 had a higher (vs. lower) level of Z-dominance and front-brainedness than most other participants of this study. For example, "Suppose you had sent £1 (so Person 2 received £3). How much money do you think Person 2 would send back to you (from 0 to £3), if Person 2 had a HIGHER level of Front-brainedness than most other participants of this study?" Using this question format, participants provided their estimates of the trustworthiness (amount of money sent back) of others with higher and lower levels of Z-dominance and front-brainedness, resulting in four variables: trustworthiness expectations for high versus low Z-dominance and high versus low front-brainedness.

Trustees

Participants assigned to the role of trustees read the descriptions of v- and z-dominant and front-and back-brained people and were asked to indicate whether they thought that they had a higher or a lower level of Z-dominance and front-brainedness than most other participants of this study ("I'm more z-dominant" vs. "I'm less Z-dominant"; "I'm more front-brained" vs. "I'm less front-brained"). Asking participants to provide their relative standing on each trait relative to others using a dichotomous measure allowed us to compute the average trustworthiness of more and less agreeable and extraverted trustees and compare it to trustors' expectations directly. Note that the instructions to personality scales often ask participants to rate themselves relative to others (e.g., Goldberg, 1992; IPIP: https://ipip.ori.org/new_ipip-50-item-scale.htm) and personality scales that use dichotomous response options do not necessarily have worse psychometric properties than the ones that use Likert-scale responses (Hilbert et al., 2016). Afterwards, trustees were introduced to the trust game, answered the comprehension check question and indicated how much money they would like to transfer back to Person 1 (between 0 and €3) if Person 1 decides to send them their money. Participants' decisions in this study were not incentivized.

We counterbalanced the order in which participants in both roles read about each trait. At the end, both trustors and trustees completed the Perceived Awareness with Research Hypotheses scale (Rubin, 2016) (4 items, e.g., "I knew what the researchers were investigating in this research"; 1 = *strongly disagree*, 7 = *strongly agree*; Cronbach's α = .88; $M = 4.64$, $SD = 1.44$).

Participants filled in basic socio-demographic information and were debriefed about the fictional nature of v-/z-dominance and front-/back-brainedness.

6.1.2 | Participants

Participants were recruited on Prolific. We recruited 200 participants (100 trustors and 100 trustees). A priori power analysis conducted with g*power (Faul et al., 2009) showed that this sample size would allow us to detect effects of $d = .28$ (for both one-sample and independent sample *t* tests, 80% power, α = .05, two-tailed). To compensate for participants failing the understanding question, we decided to collect at least 300. Overall, 301 participants finished the study. Thirty-two participants responded incorrectly to the question testing their understanding of the trust game, resulting in a total sample of 269 participants ($M_{age} = 24.99$, $SD_{age} = 7.11$, 50.7% male).

6.2 | Results

The results are presented in Figure 1.

6.2.1 | Trustors

Trustors expected agreeable trustees to return more money ($M = 1.90$ out of €3, $SD = 0.52$) than disagreeable trustees ($M = 0.54$ out of €3, $SD = 0.61$), $t(105) = 15.484$, $p < .001$, $d = 1.5$. Trustors also expected extraverted trustees to return more money ($M = 1.53$, $SD = 0.64$) than introverted trustees ($M = 0.96$, $SD = 0.66$), $t(102) = 5.49$, $p < .001$, $d = 0.54$. The presentation order (agreeableness first vs. extraversion first) did not moderate these effects (agreeableness level × order interaction: $p = .81$; extraversion level × order interaction: $p = .49$).

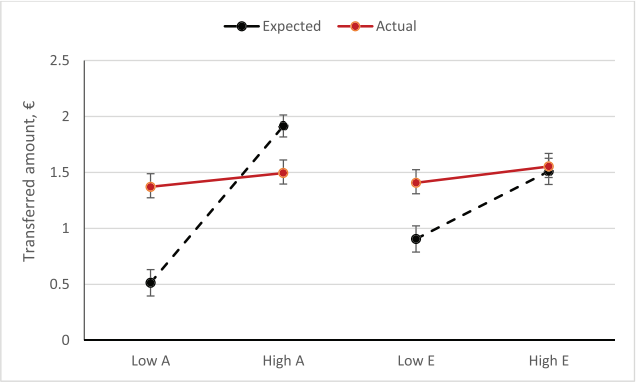

FIGURE 1 Expected and actual amounts (in €) that trustees high (vs. low) in agreeableness (A) and extraversion (E) sent back; error bars represent standard errors; Study 2

6.2.2 | Trustees

Trustee extraversion and trustee agreeableness were associated with somewhat higher trustworthiness (i.e., more money returned), however none of these associations reached the conventional level of significance (extraversion: $r = .15$, $p = .080$; agreeableness: $r = .13$, $p = .112$). These associations did not depend on whether trustees responded to extraversion versus agreeableness measures first (agreeableness × order interaction: $p = .68$; extraversion × order interaction: $p = .83$).

6.2.3 | Comparing trustors' expectations with trustees' decisions

We conducted a series of one-sample *t* tests to compare trustors' expectations of different personality profiles against their actual behavior:

Trustors expected an agreeable trustee to send them back more money ($M = 1.91$, $SD = 0.52$) than agreeable trustees actually did (1.49), $t(107) = 8.35$, $p < .001$, $d = 0.81$. Trustors expected a disagreeable trustee to send them back much less money ($M = 0.51$, $SD = 0.60$) than disagreeable trustees actually did (1.37), $t(117) = 15.48$, $p < .001$, $d = 1.43$.

Trustors did not significantly under- or over-estimate the trustworthiness of extraverted trustees ($M = 1.51$, $SD = 0.54$), test value: 1.55; $t(106) = 0.70$, $p = .49$, but they underestimated the trustworthiness of introverted trustees: They expected an introverted trustee to send them back less money ($M = 0.91$, $SD = 0.66$) than introverted trustees actually did, 1.41, $t(113) = 8.09$, $p < .001$, $d = 0.76$.

Overall, as shown in Figure 1, the amount of money trustees sent back (solid red lines) was barely affected by trustee personality. In contrast, the amount of money others expected trustees to send back (dashed black lines) varied greatly as a function of trustee personality.

6.2.4 | Assessing demand effects

Following the pre-registered analysis plan, we examined whether trustors' estimates of trustee returns were affected by participants' PARH. In the sample of trustors, we used repeated-measures ANOVA with trustee agreeableness (low vs. high) as a within-subject factor, participants' PARH (z-standardized) as a covariate and trustworthiness expectations as the dependent variable. The model specified the interaction between trustee agreeableness and PARH. Replicating the

results of the paired-sample *t* test reported above, there was a significant main effect of the within-subject factor, $F(1, 104) = 247.34$, $p < .001$, partial $\eta^2 = .79$. The interaction between trustee agreeableness and PARH was not significant, $F(1, 104) = .76$, $p = .39$. We then repeated these analyses with trustee extraversion (instead of agreeableness). The main effect of trustee extraversion was significant, $F(1, 101) = 29.95$, $p < .001$, partial $\eta^2 = .23$, while the interaction between trustee extraversion and PARH was not, $F(1, 101) = .06$, $p = .80$. Hence, regardless of trustors' awareness of research hypotheses, their expectations regarding how much money trustees will send them back were shaped by their beliefs about trustee personality, suggesting that the results are unlikely to be driven by demand effects.

Finally, in addition to these pre-registered analyses, we compared the level of PARH in trustors and trustees. If the effect of target personality on expectations is stronger than on actual behavior due to demand effects, participants asked to provide expectations of others' decisions (trustors) should show more awareness of the research hypotheses than participants asked to make decisions (trustees). However, trustors and trustees did not significantly differ in PARH, $t(266) = .36$, $p = .72$.

6.3 | Discussion

Study 2 showed that people over-estimate the importance of others' agreeableness and (to a lesser extent) extraversion when making predictions about their trustworthy behavior. This effect did not depend on participants' awareness of research hypotheses and persisted even when they were explicitly warned that the information about their partner personality is not perfectly accurate. Taken together, these findings provide initial evidence against experimenter demand effects. We will return to the potential role of demand effects in Study 4.

7 | STUDY 3

Study 3 extended the findings of Studies 1 and 2 in multiple ways. First, although Study 2's design allowed us to directly compare expected versus actual trustworthiness, this comparison was restricted to individuals who described themselves as having above versus below average agreeableness and extraversion scores. In Study 3, we explored the entire spectrum of agreeableness and extraversion by making use of a yoked control design: Each target filled in agreeableness and extraversion scales using a standard 5-point Likert scale and was paired with a rater who received the information about the matched target's agreeableness and extraversion scores.

Second, we examined the effects of partner agreeableness and extraversion across a variety of social dilemma situations. Prior research has shown that personality is a better predictor of behavior in aggregate than of any single instance of behavior (Fleeson, 2004). Hence, people might be wrong to rely on personality when predicting any single behavior, but might be right to do so when predicting average behavior measured across different contexts. Study 2 tested this possibility.

The study was conducted in two stages: In Stage 1, participants (targets) filled in agreeableness and extraversion scales and made decisions without feedback in four different social dilemma situations: the dictator game, the trust game, the public goods game, and the give some game (a continuous version of the prisoner's dilemma; Lee et al., 2013). Each game had five rounds with different pay-offs, resulting in 20 decisions overall. In Stage 2, a new group of participants (raters) were shown targets' agreeableness and extraversion scores and made predictions regarding targets' decisions in each round of each game (20 different predictions), as well as an average prediction of targets' behavior (across all 20 trials). We examined the associations between target personality and expected (by raters) as well as actual trial-level and average prosocial behavior of targets.

Measures, data collection, and analyses were preregistered (http://aspredicted.org/blind.php?x=nr2ut8).

7.1 | Method

7.1.1 | Design and procedure

Targets

The study consisted of two parts where we measured participants' personality traits and decisions in social dilemma games. The order in which participants completed these two parts was randomized.[1] As the order did not affect targets' decisions and did not interact with any variables of interest (all *p*s > .32), we do not discuss it further.

Participants completed *agreeableness* and *extraversion* scales of the IPIP-50 (Goldberg, 1992). Each trait was measured with 10-items and responses were given on a 5-point scale (1 = *very inaccurate*, 5 = *very accurate*). Both scales were reliable (extraversion: $\alpha = .90$; agreeableness: $\alpha = .85$).

In each game, participants learned that they would be matched with other participants. The payoff for each decision was given in points (1 point = $0.01). We paid out one randomly selected decision of 10 randomly selected participants. For each game, there was at least one question checking participants' understanding of the rules of the game (see Supporting Information). As preregistered, if participants responded incorrectly (to at least one question), their decisions in this particular game (but not in others) was removed from the analyses.

To enable the comparison of targets' decisions in each trial with raters' predictions of targets' decisions in each trial and on average across trials, in all games, targets indicated their decisions (and raters indicated their predictions) in percentages (e.g., what percentage of their endowment they would like to transfer to the other player). We used a slider that ranged from 0% to 100% and participants could

[1]Due to a programming error, participants who filled in the decision making part first completed the survey on January 20, and participants who filled in the personality scales first completed the survey on January 21.

make decisions in 5% steps. Across the four games, the rounds differed in participants' initial endowments (ranging from 10 points in round 1 to 500 points in round 5).

The structures of the four games are illustrated in Figure 2; the rules are summarized in Appendix A and the complete instructions are provided in the Supporting Information.

Across the games, we refer to participants' decisions in each round as *trial-level prosocial behavior*. As a measure of *average prosocial behavior*, we computed the average of participants' decisions across the 20 trials.

Raters

Raters were given the information about the targets' agreeableness and extraversion scores, and learned that their task would be to predict the targets' decisions in different situations. The predictions were incentivized by assigning a bonus of $5 to the five most accurate guessers.

Following the procedure of Study 1, participants read descriptions of extraversion and agreeableness (we used the same descriptions). Then they were provided with the targets' percentile ranks in each trait. For example, they read that the target's "score in agreeableness is higher than 75% of people," which means that the target "is more agreeable than 75% of people." The order in which the information about the target's extraversion and agreeableness was presented was counterbalanced between the participants. As the order did not affect raters' estimates ($p = .536$) and did not interact with any variables of interest ($p = .063$ and $p = .647$), we do not discuss it further. As a manipulation check, participants indicated how agreeable and how extraverted the target was (1 = *not at all*, 7 = *very*).

Afterwards, participants learned that the targets made 20 financial decisions and were asked to provide their best estimates of the targets' decisions. Raters were provided with the same description of the games and were asked to complete the same understanding check questions as the targets. As preregistered, if participants responded incorrectly (to at least one question), their estimates in this particular game (but not in others) were removed from the analyses. In each round, raters indicated what percentage of the points they thought the target transferred to the other player (or contributed to the group in the public goods game), from 0 = 0% to 100 = 100%, in 5% increments. After making an estimate in each of 20 rounds (*trial-level expected prosocial behavior*), raters were asked to consider all 20 decisions at once and provide an estimate of the percentage of the total amount of points that the target contributed/gave to other players, on average across all games and rounds (*average expected prosocial behavior*). We used the same slider measure ranging from 0% to 100% (in 5% increments).

Participants

To be able to detect a small effect (e.g., correlation between traits and behaviors: $r = .10$), we decided to collect 700 targets (and 700 raters in stage 2). To compensate for participants failing the understanding questions, we collected an additional 50 responses (e.g., 750 targets). Both targets and raters were recruited on MTurk.

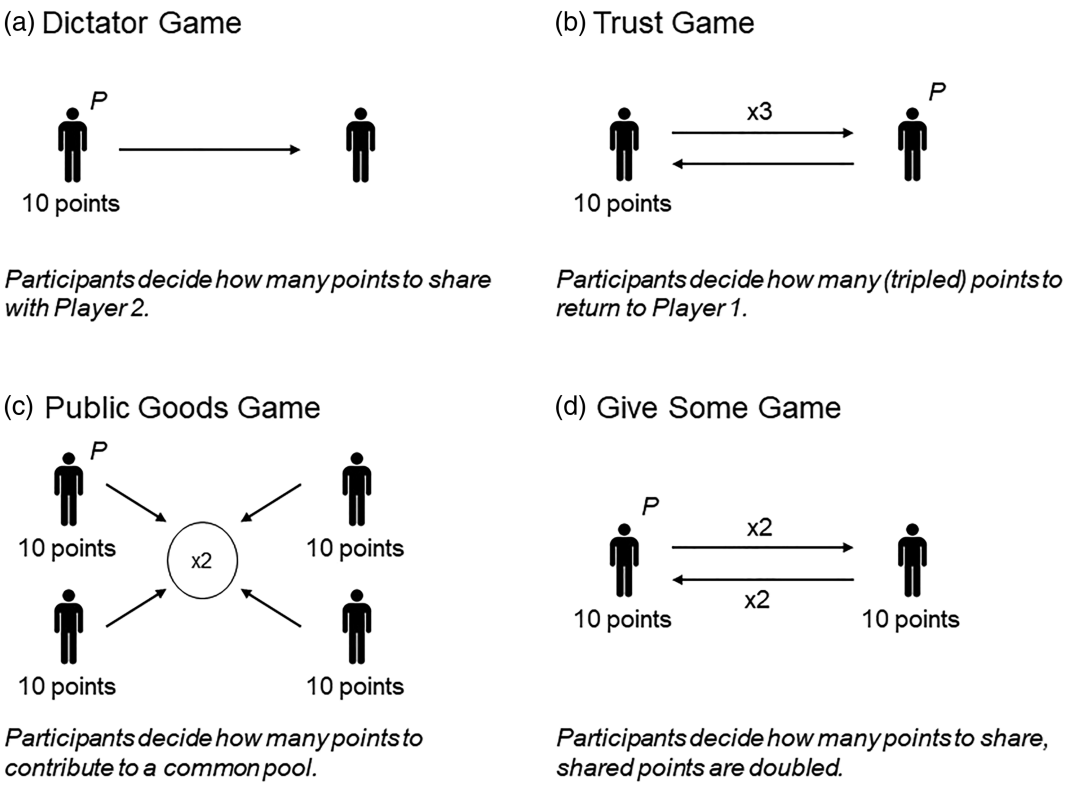

FIGURE 2 In Study 2, participants made decisions in four social dilemma games. The decisions made by participants are indicated with *P*

We received 751 complete responses from targets, of whom 108 failed the understanding check questions of all games and two had missing values on agreeableness and/or extraversion, resulting in the final sample of 643 ($M_{age} = 38.53$, $SD_{age} = 11.56$, 44.8% female) targets.

To match each target with one rater, we opened the study to 643 participants and made sure that each target is matched with only one rater. Seventy raters failed to correctly respond to the understanding questions in all four games. As a result, we re-opened the study to get the remaining 70 targets matched with raters. After going through this procedure a couple of times, all 643 targets were matched with a rater.[2] The raters had similar demographics as the targets ($M_{age} = 36.83$, $SD_{age} = 10.93$, 45.9% female).

7.2 | Results

7.2.1 | Manipulation check

We conducted multiple regression analyses with target agreeableness and extraversion as independent variables and rater perceptions of target agreeableness and extraversion as dependent variables. The results showed that raters perceived more (vs. less) agreeable targets as being more agreeable ($\beta = .63$, $p < .001$, 95% CI [0.56, 0.69]) but not more extraverted ($\beta = .009$, $p = .767$, 95% CI [−0.05, 0.07]); similarly, more (vs. less) extraverted targets were perceived as being more extraverted ($\beta = .73$, $p < .001$, 95% CI [0.67, 0.79]) but not more agreeable ($\beta = -.05$, $p = .102$, 95% CI [−0.10, 0.01]). This suggests that the manipulations of agreeableness and extraversion were successful.

7.2.2 | Associations of personality with actual versus expected prosocial behavior

Figure 3 shows the associations of target agreeableness and extraversion with both actual and expected prosocial behavior. Across all games, target agreeableness was related to rater expectations of target prosocial behavior (on average, $r = .36$, $p < .001$, 95% CI [0.29, 0.43]) and to target actual prosocial behavior (on average, $r = .24$, $p < .001$, 95% CI [0.17, 0.31]). For extraversion, there were positive associations for expected prosociality (on average, $r = .13$, $p = .001$, 95% CI [0.05, 0.21]) and close-to-zero associations for actual prosociality (on average, $r = -.09$, $p = .06$, 95% CI [−0.18, 0.002]).

7.2.3 | Comparing the effect of personality on expected versus actual trial-level prosocial behavior

As decisions were nested within participants and games, we used multilevel regression with participants and games included as random effects.[3] The effects of agreeableness and extraversion were allowed to vary randomly across the games (that is, they were included as random slopes) and the effect of decision type (actual vs. expected) was allowed to vary across participants. We used lmer function of the lme4 package in R.

Agreeableness

The interaction effect reached significance ($b = -7.15$, $p < .001$, 95% CI [−10.64, −3.66], $\beta = -.08$). Raters overestimated the prosociality of more agreeable (+ 1 *SD*; $b = -6.00$, $p = .002$, 95% CI [−8.62, −1.13], $\beta = -.09$) targets and underestimated the prosociality of less agreeable targets (−1 *SD*; $b = 4.88$, $p = .011$, 95% CI [1.13, 8.62], $\beta = .07$), see Figure 4.

Extraversion

The interaction was significant as well ($b = -4.48$, $p = .001$, 95% CI [−7.13, −1.82], $\beta = -.06$). Figure 4 shows that raters significantly overestimated the prosociality of more extraverted (+ 1 *SD*; $b = -5.05$, $p = .009$, 95% CI [−7.65, −0.16], $\beta = -.07$) targets and underestimated the prosociality of less extraverted targets (−1 *SD*; $b = 3.90$, $p = .042$, 95% CI [0.14, 7.65], $\beta = .06$).

7.2.4 | Comparing the effect of personality on expected versus actual *average* prosocial behavior

We used a similar procedure to examine whether target agreeableness and extraversion predict raters' tendency to over- (or under-)estimate target prosocial behavior across trials. We calculated the *average actual transfer* by averaging the percentage of transferred points by the target across the trials. We used raters' estimate of the percentage of transferred points by the target across the trials as the indicator of the *average expected transfer*. We regressed average prosociality on target agreeableness (mean centered), decision type (actual = 1, expected = 0) and their interaction. The interaction was significant ($b = -5.63$, $p = .008$, 95%CI [−9.77, −1.49], $\beta = .07$). Figure 4 shows that raters accurately estimated the prosociality of least (minimum empirical score) agreeable ($b = 12.07$, $p = .054$, 95%CI [−0.19, 24.32], $\beta = .21$) but overestimated the prosociality of above average (+ 1 *SD*) agreeable targets ($b = -8.24$, $p < .001$, 95%CI [−12.67, −3.82], $\beta = -.14$).

[2]This process resulted in 41 targets being matched with two raters (that is, the final number of raters we recruited was 684). We used the raters' estimates that were obtained first in the main analyses. Additional analyses using alternative raters for these 41 targets provided identical results (see Supporting Information).

[3]Our original plan (see preregistration) was to estimate the effect of personality on the difference score for each trial separately (20 regressions). Later on, we decided that the multilevel analyses would be a better option. Also, the average regression coefficient across the 20 trials were nearly identical as the regression coefficients from the multilevel analyses (e.g., .14 vs. .13 for agreeableness and .15 vs. .11 for extraversion). We present the results of the 20 regression analyses in the Supporting Information.

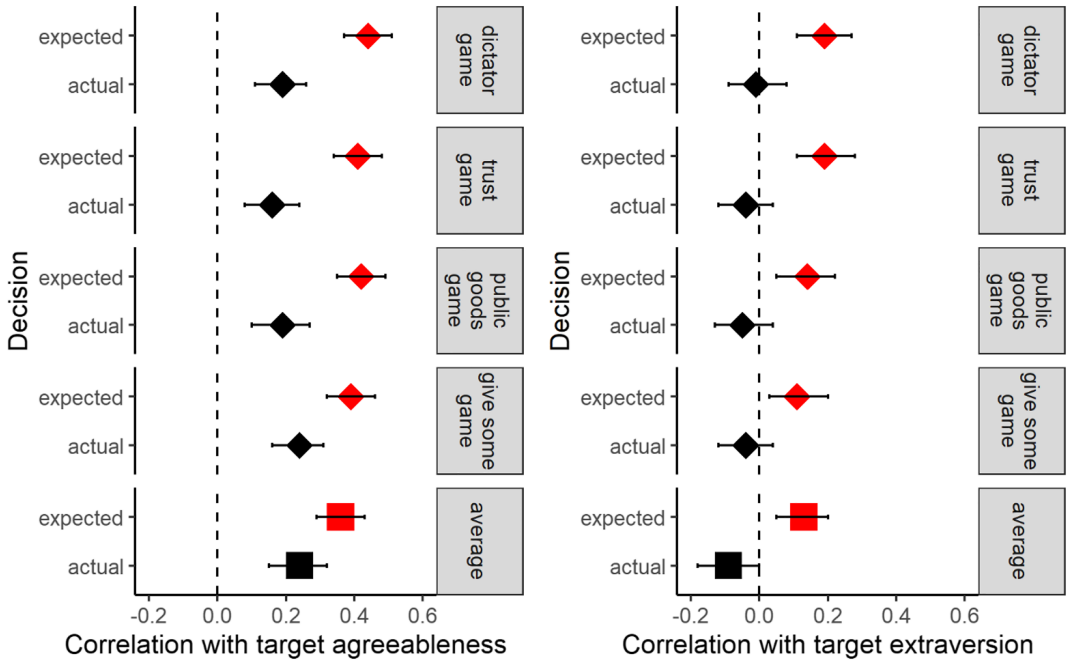

FIGURE 3 Target agreeableness (left) and extraversion (right) and actual versus expected prosociality, Study 3

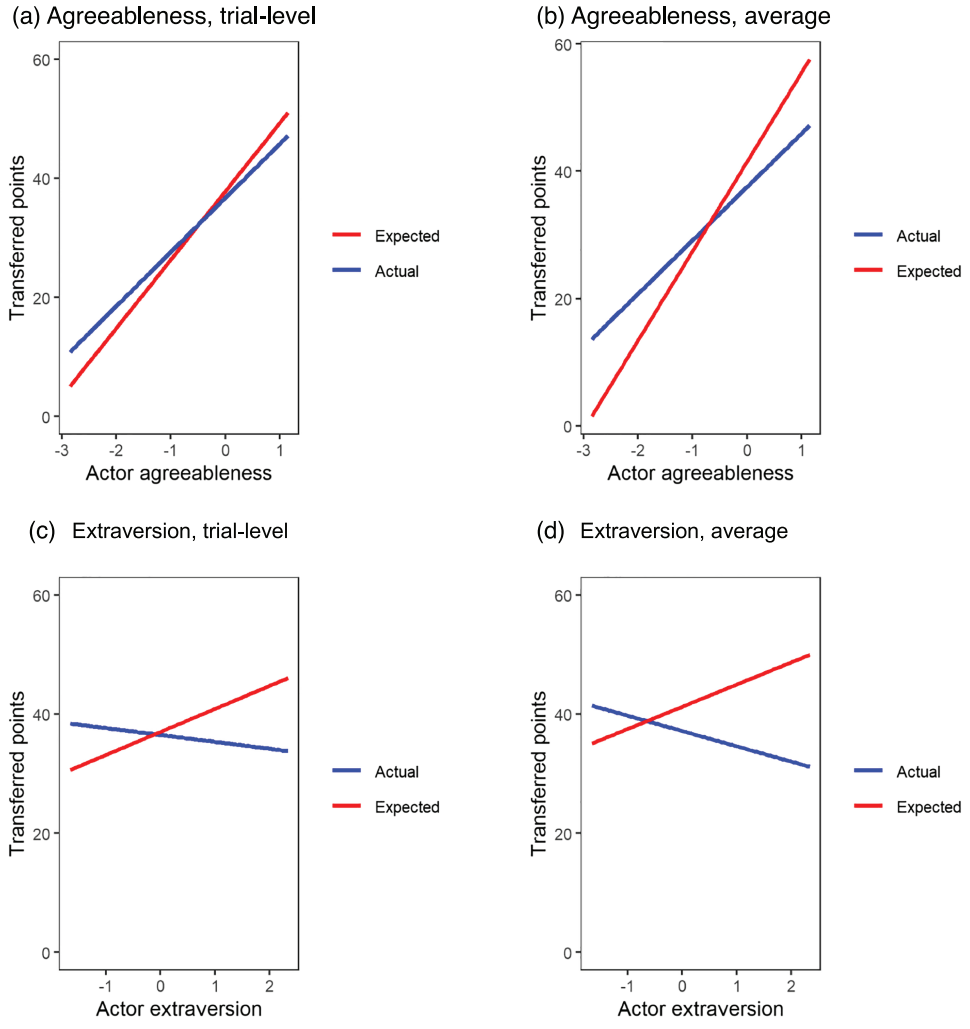

FIGURE 4 Expected and actual percentage of contributed points, model estimates, Study 3

Target extraversion also predicted raters' tendency to overestimate target transfers (extraversion × decision type interaction: $b = -6.36$, $p < .001$, 95%CI [−9.49, −3.22], $\beta = -.11$). Raters significantly overestimated the average transfers of above average extraverted (+1 *SD*) targets ($b = -10.47$, $p < .001$, 95%CI [−15.02, −5.92], $\beta = -.18$), were relatively accurate when predicting the average transfer of below average (− 1 *SD*) extraverted targets ($b = 2.24$, $p = .31$, 95%CI [−2.12, 6.61], $\beta = .04$), and underestimated the average transfer of least (minimum empirical score) extraverted targets ($b = 6.44$, $p = .035$, 95%CI [0.46, 12.42], $\beta = .11$).

Further analyses (see Supporting Information) comparing the interaction effects between target personality and decision type (actual vs. expected) on prosocial behavior at the trial level and on average showed that raters overestimated the importance of target agreeableness when predicting both trial-level and average prosocial behavior to the same extent. Hence, people tend to overestimate the effect of others' personality traits when trying to predict any single decision, as well as when predicting average decisions across multiple contexts.

7.3 | Discussion

Across different social dilemma situations, people over-relied on other personality (i.e., agreeableness and extraversion) when making predictions about their prosocial behavior. This over-reliance pattern emerged when predicting others' decisions in a single context as well as when predicting their average decisions across different contexts.

8 | STUDY 4

In Studies 1–3, participants were provided with explicit information about others' agreeableness and extraversion scores. It is possible that participants "over-used" this information when making behavioral predictions, as they had little or no other information available. In Study 4, we address this limitation by exploring whether the over-reliance pattern emerges in the absence of explicit personality information and in the presence of other information. Specifically, we recruited romantic couples who were likely to already have some insight into each other's personality traits and past behavior. We asked both partners to rate themselves on the Big Five scales. We used a variety of measures of actual prosocial behavior: Participants were assigned to interact with strangers (not each other), made decisions in several economic games and business ethics dilemmas and were asked to donate to charities. Participants were also asked to predict how their romantic partner would respond to each of these measures. All decisions (except for business ethics dilemmas) were incentivized, with one randomly selected decision being paid out as a bonus to 10 randomly selected participants.

The study was preregistered (https://aspredicted.org/blind.php?x=q7q7yw).

8.1 | Method

8.1.1 | Design and procedure

The procedure was identical for both partners in a couple. Only participants who asserted (by checking the box) that they would complete the study independently from their partner were allowed to continue. The study consisted of two parts where we measured Big Five traits and prosocial behaviors, in a randomized order.

Big Five

We used the same scale as in Study 1: mini-IPIP scale (Donnellan et al., 2006; 1 = *very inaccurate*, 5 = *very accurate*). Participants rated themselves and their partner. All scales were reliable (self-ratings: α between .67 and .84; partner ratings: α between .69 and .84). The order in which participants rated their own versus their partner personality was counterbalanced. The order had no main or interaction effects on any of the dependent measures (all *p*s > .09; see Supporting Information for details).

Trust game

We used the same trust game instructions as in Study 3 and assigned all participants to the role of trustees. Participants were told that they would be matched with an anonymous other player (not their romantic partner). As preregistered, participants who responded incorrectly ($n = 239$ or 22.5%) to the comprehension check question (same as in Study 3) were removed from the analyses of trustworthiness (but kept in the analyses of other measures). We used the same measure of trustworthiness as in Study 3 (a slider that ranged from 0% to 100%; 5% increments). Participants were told that their romantic partner will make a decision in the same situation and were asked to predict what their decision will be (i.e., how many points their romantic partner will transfer, using the same scale, ranging from 0% to 100%). Here and with respect to all the following measures of prosocial behavior, the order in which participants provided own decisions versus partner predictions was counterbalanced. The order had no main or interaction effects on any of the dependent measures (all *p*s > .08; see Supporting Information for details).

Ultimatum game with an uncertain pie size

We used a version of the ultimatum game that allowed to measure both cooperative behavior and honest (vs. deceptive) behavior (Moran & Schweitzer, 2008). Participants learned that they would be randomly assigned to another player (not their romantic partner) and would receive a number of points that can range between 10 and 300. They can then offer some of the points to the other player. The other player can accept or reject their offer, and in case of rejection none of the players will receive anything. Critically, participants

learned that the other player would not know how many points participants had at their disposal; they would only know that any number between 10 and 300 is equally likely. As preregistered, participants who responded incorrectly to a comprehension check question (see Supporting Information; $n = 39$ or 3.7%) or indicated implausible values (e.g., transferred 300 points while only having 100 points, $n = 36$ or 3.3%) were removed from the analyses of this game (but kept in the analyses of other measures).

Participants learned that the number of points they have is 100 and were asked to communicate two numbers to the other player: (1) How many points they would like to offer them and (2) the total amount of points they received. The first number is used as a measure of cooperative behavior. The second number is used as a measure of honest behavior: By underreporting the number of points they received, participants could make lower offers appear more acceptable to the other player. Following the standard procedure (Moran & Schweitzer, 2008), for each participant, we subtracted the actual amount (100) from the amount they reported they received. Lower values on the obtained difference score reflect more deceptive behavior and higher values reflect more honest behavior. Participants were also asked to think about how their romantic partner would respond to both measures and provided two numbers: their expectation of their romantic partner's cooperation (how many points their partner would transfer to the other player) and honesty (the amount of points their partner would report they had received).

Charity donation

Participants were reminded that one randomly selected decision they made so far will be paid out as a bonus (reaching up to 3 pounds). They were then asked to indicate what percentage of their bonus (from 0% to 100%) they would like to donate to Doctors Without Borders, a charitable organization. They also made predictions about how much their partner will donate, using the same scale (0% to 100%).

Business ethics dilemmas

Participants were shown three ethical dilemmas (taken from Ashton & Lee, 2008) in which financial interests are pitted against ethical concerns (e.g., harming others). For example, participants were asked whether they would market an extremely profitable product even though the product has known health risks (1 = *definitely not*, 100 = *definitely yes*; reverse-coded such that higher values represent more ethical decisions). The scenarios are presented in Appendix B. For each dilemma, participants indicated their decision and made predictions about how their partner will respond to the same scenarios, using the same scale. Participants' responses to three dilemmas were combined into one scale (own decisions: $\alpha = .73$, partner predictions: $\alpha = .76$).

Overall prosociality

We standardized the five different measures of prosocial behavior described above and averaged them into an index of overall prosociality. We also report the results for each measure separately (here and in Supporting Information, Table S3).

8.1.2 | Participants

We recruited 529 romantic couples (1058 unique participants; $M_{age} = 29.93$, $SD_{age} = 8.31$, 49.3% male) on Prolific Academic. The majority (89%) of the couples were heterosexual and were living together (33% married, 2.1% in a registered partnership, 39.9% cohabiting). All participants indicated that they have filled in the questionnaire independently of their partners and did not discuss it with them. A sensitivity power analysis conducted with g*power 3.1 (Faul et al., 2009) showed that this sample size would allow us to detect a correlation of $r = .09$ with a power of 80% and $\alpha = .05$ (two-tailed test).

8.2 | Results

Zero-order correlations among the variables are shown in Table S2 (Supporting Information).

8.2.1 | Associations of personality with actual versus expected prosocial behavior

Figure 5 shows the associations of actor personality (self-ratings of agreeableness and extraversion) with both actor prosocial behavior and partner expectations of actor prosocial behavior. With an exception of the responses to business ethics dilemmas, actor agreeableness appears to be more strongly related to partner expectations of actor prosocial behavior (on average, $r = .30$, $p < .001$, 95% CI [0.24, 0.35]) than to actor actual prosocial behavior (on average, $r = .21$, $p < .001$, 95% CI [0.15, 0.27]).

A somewhat similar pattern emerged for extraversion, although it was not significantly associated with either expected ($r = .05$, $p = .11$, 95% CI [−0.01, 0.11]) or actual ($r = -.03$, $p = .39$, 95% CI [−0.09, 0.03]) prosocial behavior. Of the remaining Big Five traits, openness and neuroticism were positively associated with both actual (openness: $r = .10$, $p < .001$, 95% CI [0.04, 0.16], neuroticism: $r = .07$, $p = .031$, 95% CI [0.01, 0.13]) and expected (openness: $r = .08$, $p = .009$, 95% CI [0.02, 0.14], neuroticism: $r = .11$, $p < .001$, 95% CI [0.05, 0.17]) prosociality (see Table S2; note that only agreeableness and neuroticism significantly predicted expectations when all traits were used as predictors).

Comparing the effect of personality on expected versus actual prosocial behavior

We used multilevel analyses with participants nested within couples. Following the recommendations in the literature (Kenny & Kashy, 2011), we allowed for correlated errors among members

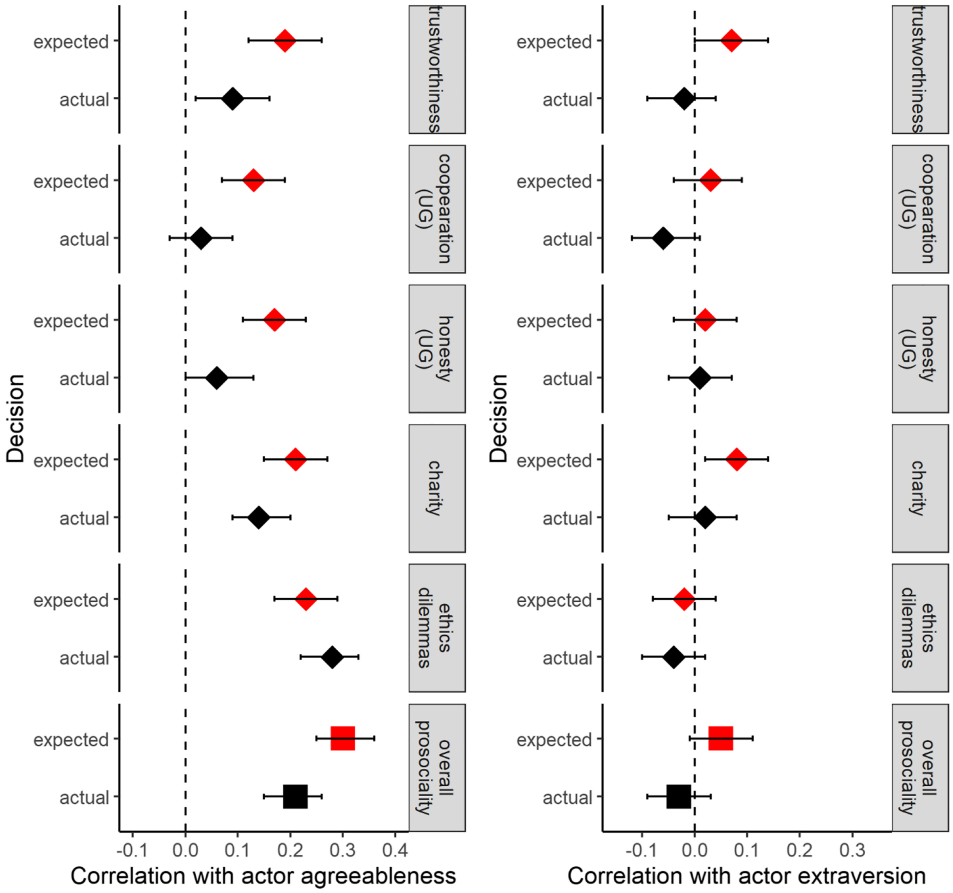

FIGURE 5 Correlations of actor agreeableness (left) and extraversion (right) with actor prosocial behavior (black points) and partner expectations of actor prosocial behavior (red points), Study 4

of the same couple. We used gls function of the nlme package[4] in R.

Agreeableness. We regressed the overall index of prosocial behavior on actor agreeableness (centered), decision type (actual = 1, expected = 0) and their interaction. The interaction was significant ($b = -.09$, $p = .002$, 95% CI [−0.15, −0.03], $\beta = -.06$). A simple slope analysis (Figure 6) shows that individuals underestimated the prosociality of less (−1 *SD*) agreeable partners (effect of decision type: $b = .08$, $p = .036$, 95% CI [0.01, 0.15], $\beta = .07$) and overestimated the prosociality of more (+ 1 *SD*) agreeable partners (effect of decision type: $b = -.07$, $p = .049$, 95% CI [−0.15, −0.003], $\beta = -.06$). We also analyzed each indicator of prosocial behavior separately. We obtained the same results for four out of five indicators of prosociality: One exception were responses to business ethics dilemmas where agreeableness had a similar effect on both actual and expected prosociality (see Table S3).

Extraversion. We conducted the same analysis with extraversion. The extraversion × decision type interaction was significant ($b = -.05$, $p = .033$, 95% CI [−0.10, −0.004], $\beta = -.04$). However, the simple slope analysis revealed only marginally significant over- and underestimation of prosociality for the most (effect of decision type: $b = -.11$, $p = .067$, 95% CI [−0.24, 0.01], $\beta = -.09$) and the least (effect of decision type: $b = .10$, $p = .064$, 95% CI [−0.01, 0.21], $\beta = .09$) extraverted individuals in the sample.

None of the remaining Big Five dimensions showed a significant interaction with decision type (all *p*s > .11; see Supporting Information for details).

8.2.2 | Exploratory analyses

As suggested by an anonymous reviewer, we repeated the analyses using partner-ratings of actor personality (instead of actor self-ratings). The results showed that a significant personality × decision type interaction for agreeableness ($b = -.10$, $p = .0002$, 95% CI [−0.16, −0.05], $\beta = -.08$) but not for extraversion. Hence, when predicting partner prosocial behavior, participants over-relied on their partner agreeableness as reflected in partner self-ratings and in their own ratings of partner agreeableness. Further details of these analyses are reported in the Supporting Information.

[4]Our original plan (see preregistration) was to compute a difference score between actual and expected prosociality and to regress it on actor personality using multilevel analysis. Later on, we were made aware of the limitations of using differences scores (Krueger & Wright, 2011) and also realized that the difference score in this study represents a combination of measures obtained from both members of a dyad, which would artificially eliminate differences between couples (variance at the couple level). Therefore, instead of using difference scores, we decided to estimate the interaction between actor personality and decision type (actual vs. expected). Importantly, the pre-registered analyses yielded the same conclusions and can be consulted in Supporting Information.

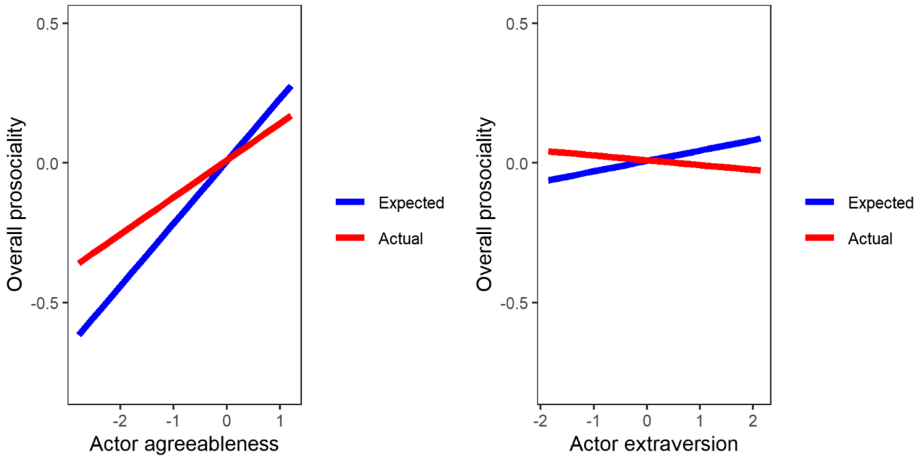

FIGURE 6 Actual versus expected prosocial behavior as a function of actor agreeableness, Study 4

8.3 | Discussion

Studies 4 showed that people give others' personality too much weight when making predictions about their prosocial behavior even in the absence of explicit information about others' personality. When predicting their romantic partners' prosocial behavior, people overestimated the prosociality of more agreeable (and, to a smaller extent, more extraverted) partners and underestimated the prosociality of less agreeable (and, to a smaller extent, less extraverted) partners. Note that participants' predictions regarding the effects of others traits—neuroticism, openness, and conscientiousness—were accurate. In other words, participants' prosociality predictions were guided by agreeableness (and extraversion) more than warranted by reality. Our finding that participants overestimated the importance of their partner personality without being explicitly informed about it further suggests that the discrepancy between lay beliefs and reality is unlikely to be driven by demand effects.

The over-reliance pattern emerged across several measures of prosocial behavior. One exception were decisions in business ethics dilemmas, where agreeableness predicted both expected and actual prosociality. We speculate that in contrast to more context-free measures of prosociality in social dilemma games, more (vs. less) prosocial decisions in business ethics dilemmas do not only reflect one's underlying prosocial dispositions but reveal a variety of context-specific attitudes and values (e.g., attitudes towards violent sports or healthy foods), making it possible for a partner to use their knowledge of these attitudes and values (rather than agreeableness) to make predictions. In future studies, it might be interesting to extend the present findings to a wide range of situations, including hypothetical and real-life prosocial behaviors, context-specific and context-free prosocial choices, prosocial behavior towards different targets (e.g., close vs. distant others) and different life areas (e.g., work vs. family).

9 | GENERAL DISCUSSION

What traits do people value (in others) when faced with dilemmas of help and cooperation, and do people hold accurate lay beliefs about the role of personality in prosociality? Herein, we explored these questions with respect to two *interpersonal* dimensions of the Big Five: agreeableness and extraversion. Our results suggest that people consider agreeableness to be particularly important for prosociality: People expect greater prosociality from more (vs. less) agreeable others and are more likely to cooperate with more (vs. less) agreeable partners. Although people tend to use information about others' agreeableness as a trustworthiness cue, our results indicate that this might not be an effective strategy. In all studies, participants believed the effect of agreeableness on prosocial behavior to be larger than it actually was. For example, in Study 3, even the *largest* correlation between agreeableness and *actual behavior* ($r = .23$) was smaller than the *smallest* correlation between agreeableness and *expected behavior* ($r = .32$). Participants consistently overestimated the prosociality of more agreeable others and underestimated the prosociality of less agreeable others.

While people had strong and consistent beliefs regarding the prosociality of more (vs. less) agreeable others, lay beliefs about extraversion were somewhat mixed: Participants linked extraversion to more prosociality only in two out of four studies. It is possible that lay beliefs about extraversion and prosociality are more complex than initially assumed: For example, people might link some facets of extraversion (e.g., warmth, positive emotions) to more prosociality and other facets (e.g., assertiveness, excitement-seeking) to less prosociality. Hence, when it comes to lay beliefs about extraversion and prosociality, future studies might benefit from a facet-level analysis.

Similarly, when making predictions about others' prosocial behavior, participants consistently gave agreeableness more weight than warranted by reality; the mismatch between expectations and reality was however less pronounced and less consistent for extraversion. At a broader theoretical level, this finding is consistent with the recent insights from the literature on stereotype accuracy (Epley & Eyal, 2019; Jussim et al., 2015). This literature suggested that people are more likely to exaggerate differences between groups in the dimensions that define these groups and distinguish them from others (Eyal & Epley, 2017). As *prosocial* behavior could be considered more central to agreeableness, people tend to exaggerate the differences in

prosociality between agreeable/disagreeable (but less so between extraverted/introverted, and not at all between more or less conscientious, neurotic or open) people.

Could the over-reliance on personality demonstrated in the present studies be just a methodological artifact, such as experimenter demand effects? In Study 2, the effect of target personality on predictions of trustworthiness did not depend on participants' awareness of research hypotheses and persisted even when they were explicitly warned that the information about their partner personality is not perfectly accurate. The results of Study 4 further speak against the possibility of demand effects. In this study, people overestimated the effect of agreeableness on others' prosociality even in the absence of any explicit personality information.

Prior research has shown that personality may represent a worse predictor of single instances of behavior than of the behavior in aggregate (Diener & Larsen, 1984; Epstein, 1979; Fleeson, 2001, 2004). Can averaging predictions of behavior across multiple contexts improve prediction accuracy? Relying on personality might lead to less errors when predicting average behavior across different situations than when predicting single instances of behavior. However, the results of Study 3 showed that participants equally overestimated the importance of personality when predicting any single decision as well as average behavior across trials. It is possible that different social dilemmas in this study do not represent sufficiently diverse situations. Hence, it should be further explored whether relying on personality might be a more effective and less error-prone strategy when predicting average behavior not only across different social dilemmas but also across different everyday real-life behaviors measured over time (see Fleeson, 2001).

9.1 | Limitations and future research

We found that individuals overestimate the effect of agreeableness and extraversion on others' prosociality when given explicit information about others' personality scores (Studies 1–3). Situations like this are common: For example, companies make hiring and promotion decisions based on candidates' personality profiles, and online dating websites often provide users with the personality scores of potential partners. Yet, there are many contexts where people do not receive explicit information about others' personality traits.

Study 4 suggests that the over-reliance pattern can emerge even in the absence of explicit personality information by demonstrating it in romantic couples. Do people over-rely on agreeableness and extraversion when predicting strangers' prosociality as well? The literature on the accuracy of personality perceptions at zero acquaintance has shown that people accurately detect personality in strangers based on brief conversations, social media profiles or brief silent videos (Connelly & Ones, 2010; Tskhay & Rule, 2014). It has been suggested that individuals' personality finds an expression in their behavior (e.g., online posts, consumption patterns), emotional and facial expressions, body language and even looks (Gosling et al., 2002). Future studies should address whether people tend to pick up on these cues in zero-acquaintance situations and misestimate the prosociality of agreeable and extraverted others, even in the absence of explicit knowledge about their actual level of agreeableness and extraversion.

Also, our finding that individuals' expectations regarding the prosociality of agreeable versus disagreeable and extraverted versus introverted individuals do not match the reality raises the question of long-term effects of partner agreeableness and extraversion on trust and cooperation. Are agreeable and extraverted people able to maintain a reputation as particularly trustworthy colleagues and partners over time? And if yes, how do they manage to do that? Do people ever learn that agreeableness and extraversion cues are not as reliable as they initially thought? These questions present exciting opportunities for future studies.

9.2 | Theoretical contributions

The present research contributes to several streams of literature. First, our findings contribute to the literature on social perception and stereotype accuracy (Jussim et al., 2015). Specifically, we showed that people are subject to attribution errors not only when making inferences about others' dispositions from learning about their behavior (known as Fundamental Attribution Error), but also when making inferences (i.e., predictions) about others' behavior from learning about their dispositions. That is, people relied on personality (e.g., agreeableness) more than warranted by reality when predicting others' prosocial behavior. Also, the fact that people over-estimate the effect of most social (but not other dimensions) of the Big Five provides further support to a recently expressed idea that people exaggerate group differences only with respect to the dimensions that are most central/defining of the groups (Epley & Eyal, 2019).

Second, our findings contribute to the literature on personality predictors of prosociality (e.g., Graziano et al., 2007). Even though agreeableness implies an elevated concern for others by definition, agreeableness was not consistently related to prosocial choices in the present studies (overall $N = 2993$). These results are consistent with some prior studies that failed to detect a significant association between agreeableness and trustworthiness as well (e.g., Evans & Revelle, 2008). This could be explained by a broad nature of agreeableness as a construct (it includes several facets, not all of which are directly linked to cooperation). Indeed, using a six-factor model of personality (HEXACO) that differentiates between agreeableness (forgiveness, temper-control) and honesty-humility (selflessness and modesty), revealed that only the latter consistently predicts trustworthiness (Thielmann & Hilbig, 2015). We hope that future studies will extend the present investigation into the effects of partner personality to honesty-humility, as well as other prosocial personality traits, such as moral identity (Aquino et al., 2009) or justice sensitivity (Stavrova & Schlösser, 2015). Such investigations would provide valuable insights about the generalizability of our findings of a mismatch between the importance of personality as driving prosocial behavior in lay beliefs versus reality.

ACKNOWLEDGMENT

There is no funding to report.

CONFLICT OF INTEREST

There is no conflict of interest.

AUTHOR CONTRIBUTIONS

OS: conceptualization, methodology, data curation (Studies 2–4), formal analysis, writing – original draft; AME: conceptualization, methodology, writing – review and editing; WS: data curation (Study 1), writing – review and editing; IVB: conceptualization, methodology, writing – review and editing.

DATA AVAILABILITY STATEMENT

The data are available at https://osf.io/xu4mz/?view_only=1ea2d4a97eac42e8bece0ccba591eb20.

ORCID

Olga Stavrova https://orcid.org/0000-0002-6079-4151
Anthony M. Evans https://orcid.org/0000-0003-3345-5282
Willem Sleegers https://orcid.org/0000-0001-9058-3817
Ilja van Beest https://orcid.org/0000-0003-2855-3638

REFERENCES

Ames, D. R., & Bianchi, E. C. (2008). The Agreeableness Asymmetry in First Impressions: Perceivers' Impulse to (Mis)judge Agreeableness and How It Is Moderated by Power. *Personality and Social Psychology Bulletin*, *34*(12), 1719–1736. https://doi.org/10.1177/0146167208323932

Aquino, K., Freeman, D., Reed, A. II, Lim, V. K. G., & Felps, W. (2009). Testing a social-cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. *Journal of Personality and Social Psychology*, *97*(1), 123–141. https://doi.org/10.1037/a0015406

Ashton, M. C., & Lee, K. (2008). The prediction of honesty–humility-related criteria by the HEXACO and five-factor models of personality. *Journal of Research in Personality*, *42*(5), 1216–1228. https://doi.org/10.1016/j.jrp.2008.03.006

Back, M. D., Schmukle, S. C., & Egloff, B. (2011). A closer look at first sight: Social relations lens model analysis of personality and interpersonal attraction at zero acquaintance. *European Journal of Personality*, *25*(3), 225–238. https://doi.org/10.1002/per.790

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

Bono, J. E., Boles, T. L., Judge, T. A., & Lauver, K. J. (2002). The role of personality in task and relationship conflict. *Journal of Personality*, *70*(3), 311–344. https://doi.org/10.1111/1467-6494.05007

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, *136*(6), 1092–1122. https://doi.org/10.1037/a0021212

Cooper, D. A., Connolly, T., & Kugler, T. (2015). Lay personality theories in interactive decisions: Strongly held, weakly supported. *Journal of Behavioral Decision Making*, *28*(3), 201–213. https://doi.org/10.1002/bdm.1842

Critcher, C. R., Dunning, D., & Rom, S. C. (2015). Causal trait theories: A new form of person knowledge that explains egocentric pattern projection. *Journal of Personality and Social Psychology*, *108*(3), 400–416. https://doi.org/10.1037/pspa0000019

Diener, E., & Larsen, R. J. (1984). Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. *Journal of Personality & Social Psychology*, *47*(4), 871–883. https://doi.org/10.1037/0022-3514.47.4.871

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, *18*(2), 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Drouvelis, M., & Georgantzis, N. (2019). Does revealing personality data affect prosocial behaviour? *Journal of Economic Behavior & Organization*, *159*, 409–420. https://doi.org/10.1016/j.jebo.2019.02.019

Epley, N., & Eyal, T. (2019). Chapter Two—Through a looking glass, darkly: Using mechanisms of mind perception to identify accuracy, overconfidence, and underappreciated means for improvement. In J. M. Olson (Ed.), *Advances in experimental social psychology* (Vol. 60, pp. 65–120). Academic Press. https://doi.org/10.1016/bs.aesp.2019.04.002

Epstein, S. (1979). The stability of behavior: I on predicting most of the people much of the time. *Journal of Personality and Social Psychology*, *37*(7), 1097–1126. https://doi.org/10.1037/0022-3514.37.7.1097

Evans, A. M., & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*, *42*(6), 1585–1593. https://doi.org/10.1016/j.jrp.2008.07.011

Eyal, T., & Epley, N. (2017). Exaggerating accessible differences: When gender stereotypes overestimate actual group differences. *Personality and Social Psychology Bulletin*, *43*(9), 1323–1336. https://doi.org/10.1177/0146167217713190

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavioral Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Feiler, D. C., & Kleinbaum, A. M. (2015). Popularity, similarity, and the network extraversion bias. *Psychological Science*, *26*(5), 593–603. https://doi.org/10.1177/0956797615569580

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027. https://doi.org/10.1037/0022-3514.80.6.1011

Fleeson, W. (2004). Moving personality beyond the person-situation debate:The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, *13*(2), 83–87. https://doi.org/10.1111/j.0963-7214.2004.00280.x

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21–38. https://doi.org/10.1037/0033-2909.117.1.21

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, *4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, *82*(3), 379–398. https://doi.org/10.1037/0022-3514.82.3.379

Graziano, W. G., Hair, E. C., & Finch, J. F. (1997). Competitiveness mediates the link between personality and group performance. *Journal of Personality and Social Psychology*, *73*(6), 1394–1408.

Graziano, W. G., Habashi, M. M., Sheese, B. E., & Tobin, R. M. (2007). Agreeableness, empathy, and helping: A person × situation perspective. *Journal of Personality and Social Psychology*, *93*(4), 583–599. https://doi.org/10.1037/0022-3514.93.4.583

Hilbert, S., Küchenhoff, H., Sarubin, N., Nakagawa, T. T., & Bühner, M. (2016). The influence of the response format in a personality questionnaire: An analysis of a dichotomous, a Likert-type, and a visual analogue scale. *TPM-Testing, Psychometrics, Methodology in Applied Psychology*, *23*(1), 3–24.

Hilbig, B. E., & Zettler, I. (2009). Pillars of cooperation: Honesty-humility, social value orientations, and economic behavior. *Journal of Research in Personality*, *43*(3), 516–519. https://doi.org/10.1016/j.jrp.2009.01.003

Jensen-Campbell, L. A., Adams, R., Perry, D. G., Workman, K. A., Furdella, J. Q., & Egan, S. K. (2002). Agreeableness, extraversion, and peer relations in early adolescence: Winning friends and deflecting aggression. *Journal of Research in Personality*, *36*(3), 224–251. https://doi.org/10.1006/jrpe.2002.2348

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). Guilford Press.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*(4), 765–780. https://doi.org/10.1037/0021-9010.87.4.765

Jussim, L., Crawford, J. T., Anglin, S. M., Chambers, J. R., Stevens, S. T., & Cohen, F. (2016). Stereotype accuracy: One of the largest and most replicable effects in all of social psychology. In *Handbook of prejudice, stereotyping, and discrimination* (2nd ed., pp. 31–63). Psychology Press.

Jussim, L., Crawford, J. T., & Rubinstein, R. S. (2015). Stereotype (in)accuracy in perceptions of groups and individuals. *Current Directions in Psychological Science*, *24*(6), 490–497. https://doi.org/10.1177/0963721415605257

Kausel, E. E., & Connolly, T. (2014). Do people have accurate beliefs about the behavioral consequences of incidental emotions? Evidence from trust games. *Journal of Economic Psychology*, *42*, 96–111. https://doi.org/10.1016/j.joep.2014.02.002

Kenny, D. A., & Kashy, D. A. (2011). Dyadic data analysis using multilevel modeling. In *Handbook for advanced multilevel analysis* (pp. 335–370). Routledge/Taylor & Francis Group.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. Guilford Press.

Krueger, J. I., & Wright, J. C. (2011). Measurement of self-enhancement (and self-protection). In M. D. Alicke & C. Sedikides (Eds.), *Handbook of self-enhancement and self-protection*. Guilford.

Kugler, T., Neeman, Z., & Vulkan, N. (2014). Personality traits and strategic behavior: Anxiousness and aggressiveness in entry games. *Journal of Economic Psychology*, *42*, 136–147. https://doi.org/10.1016/j.joep.2014.02.006

Lee, J. J., Knox, B., Baumann, J., Breazeal, C., & DeSteno, D. (2013). Computationally modeling interpersonal trust. *Frontiers in Psychology*, *4*(893). https://doi.org/10.3389/fpsyg.2013.00893

McCrae, R. R., & Costa, P. T. Jr. (1989). The structure of interpersonal traits: Wiggins's circumplex and the five-factor model. *Journal of Personality and Social Psychology*, *56*(4), 586–595. https://doi.org/10.1037/0022-3514.56.4.586

Miller, D. T., & Ratner, R. K. (1998). The disparity between the actual and assumed power of self-interest. *Journal of Personality and Social Psychology*, *74*(1), 53–62. https://doi.org/10.1037/0022-3514.74.1.53

Mitsopoulou, E., & Giovazolias, T. (2015). Personality traits, empathy and bullying behavior: A meta-analytic approach. *Aggression and Violent Behavior*, *21*, 61–72. https://doi.org/10.1016/j.avb.2015.01.007

Moran, S., & Schweitzer, M. E. (2008). When better is worse: Envy and the use of deception. *Negotiation and Conflict Management Research*, *1*(1), 3–29. https://doi.org/10.1111/j.1750-4716.2007.00002.x

Naber, A. M., Payne, S. C., & Webber, S. S. (2018). The relative influence of trustor and trustee individual differences on peer assessments of trust. *Personality and Individual Differences*, *128*, 62–68. https://doi.org/10.1016/j.paid.2018.02.022

Ross, L. (1977). The intuitive psychologist and his short-comings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). Academic. https://doi.org/10.1016/S0065-2601(08)60357-3

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. McGraw-Hill.

Rubin, M. (2016). The Perceived Awareness of the Research Hypothesis Scale: Assessing the influence of demand characteristics. Figshare. https://doi.org/10.6084/m9.figshare.4315778

Selfhout, M., Burk, W., Branje, S., Denissen, J., van Aken, M., & Meeus, W. (2010). Emerging late adolescent friendship networks and Big Five personality traits: A social network approach. *Journal of Personality*, *78*(2), 509–538. https://doi.org/10.1111/j.1467-6494.2010.00625.x

Stavrova, O., & Schlösser, T. (2015). Solidarity and social justice: Effect of individual differences in justice sensitivity on solidarity behaviour. *European Journal of Personality*, *29*(1), 2–16. https://doi.org/10.1002/per.1981

Stopfer, J. M., Egloff, B., Nestler, S., & Back, M. D. (2013). Being popular in online social networks: How agentic, communal, and creativity traits relate to judgments of status and liking. *Journal of Research in Personality*, *47*(5), 592–598. https://doi.org/10.1016/j.jrp.2013.05.005

Stopfer, J. M., Egloff, B., Nestler, S., & Back, M. D. (2014). Personality expression and impression formation in online social networks: An integrative approach to understanding the processes of accuracy, impression management and meta-accuracy. *European Journal of Personality*, *28*(1), 73–94. https://doi.org/10.1002/per.1935

Thielmann, I., & Hilbig, B. E. (2015). The traits one can trust: Dissecting reciprocity and kindness as determinants of trustworthy behavior. *Personality and Social Psychology Bulletin*, *41*(11), 1523–1536. https://doi.org/10.1177/0146167215600530

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

Tov, W., Nai, Z. L., & Lee, H. W. (2016). Extraversion and agreeableness: Divergent routes to daily satisfaction with social relationships. *Journal of Personality*, *84*(1), 121–134. https://doi.org/10.1111/jopy.12146

Tskhay, K. O., & Rule, N. O. (2014). Perceptions of personality in text-based media and OSN: A meta-analysis. *Journal of Research in Personality*, *49*, 25–30. https://doi.org/10.1016/j.jrp.2013.12.004

van der Linden, D., Scholte, R. H. J., Cillessen, A. H. N., Nijenhuis, J. t., & Segers, E. (2010). Classroom ratings of likeability and popularity are related to the Big Five and the general factor of personality. *Journal of Research in Personality*, *44*(5), 669–672. https://doi.org/10.1016/j.jrp.2010.08.007

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer. https://doi.org/10.1007/978-0-387-21706-2

Volk, S., Thöni, C., & Ruigrok, W. (2012). Temporal stability and psychological foundations of cooperation preferences. *Journal of Economic Behavior & Organization*, *81*(2), 664–676. https://doi.org/10.1016/j.jebo.2011.10.006

Wolters, N., Knoors, H., Cillessen, A. H. N., & Verhoeven, L. (2014). Behavioral, personality, and communicative predictors of acceptance and popularity in early adolescence. *The Journal of Early Adolescence*, *34*(5), 585–605. https://doi.org/10.1177/0272431613510403

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Stavrova, O., Evans, A. M., Sleegers, W., & van Beest, I. (2022). Examining the accuracy of lay beliefs about the effects of personality on prosocial behavior. *Journal of Behavioral Decision Making*, *35*(5), e2282. https://doi.org/10.1002/bdm.2282

APPENDIX A: DESCRIPTION OF THE GAMES, STUDY 3

In the *dictator game*, participants were given a certain amount of points and could decide how many of these points to transfer to Player 2. In each round, they were asked to indicate what percentage of their points they would like to transfer to Player 2. In the *trust game*, participants were assigned to the role of Players 2 (trustees). They learned that Player 1 received some points from the experimenter and could choose to transfer them to Player 2 (i.e., to them), in which case the transferred points would be tripled. In each round, they were asked to indicate what percentage of the points they would like to transfer back to Player 1. In the *public goods game*, participants learned that they are randomly assigned to interact with three other participants of the study, each of whom are given a certain amount of points for each interaction. Each member of the group decides what percentage of their points to keep for themselves, and what (if any) to contribute to the groups common project. All points contributed to the common project are doubled, and then split evenly among the four group members. In each round, they were asked to indicate what percentage of their points they would like to contribute. Finally, in the *give some game*, participants were told that they would be randomly assigned to another participant of the study and that both of them would receive a certain amount of points. Participants could give some, all, or none of their points to their partner, and their partner had the same decision to make. The amount the participant decided to give his/her partner is doubled by the experimenters before being handed over to the partner; and the amount the partner decides to give to the participant is doubled before handed over too. In each round, they indicated what percentage of their points they give to their partner.

APPENDIX B: BUSINESS ETHICS DILEMMAS, STUDY 4

Source: Ashton and Lee (2008)

B.1 | Dilemma 1

Suppose that you are in charge of new products for a food processing company. Your research-and-development team has come up with a new snack food, "Tastee Nuggets," that has received high marks in preliminary "taste tests." Part of the reason for the good taste of Tastee Nuggets is the use of some flavorful new artificial sweeteners and oils. However, some laboratory tests performed by your company suggest that these sweeteners and oils are likely to have addictive properties similar to those of some drugs, and are also likely to increase the risks of obesity, heart disease, and cancer in people who consume large amounts of those substances. Projections by your company's marketing team suggest that this product will be extremely profitable, and this will almost certainly lead to a major raise and promotion for you personally. It is now your decision as to whether or not Tastee Nuggets should be added to your company's product line, so that advertising and sales can soon begin.

B.2 | Dilemma 2

Suppose that you are managing a pension fund and are looking for good new investments. Recently, a violent new sport called TotalFighting has become fairly popular, with many people watching televised championship fights. Following the past few championship fights, rates of assault and homicide increased about 10%, nationwide, for several days. The company that runs the sport of TotalFighting has become very profitable, and is likely to become even more profitable in the future as similar sports are introduced into the market. Your pension fund now has the opportunity to buy some shares in this company, which would likely result in major gains in the value of the pension fund and also in your own commission payments.

B.3 | Dilemma 3

Suppose that you are a lawyer for an industrial products company that sells equipment used in drilling for oil and natural gas. You are aware that the country of Petronia is interested in buying large amounts of equipment from your company. However, because Petronias government has a very poor human rights record, it is illegal for any company from your country to do business with Petronia. Despite the laws against doing business with Petronia, you have discovered a legal loophole. If your company sets up a subsidiary company overseas—for example, in a small Caribbean island—then you can sell the equipment to Petronia through this company, and thereby avoid being prosecuted by your own government for breaking the law. This would result in large profits for your company, and also a large raise and promotion for yourself.