Personality Testing and the Public Goods Game

Click here for the latest version

Daniel Woods*

April 2, 2025

Abstract

Personality tests are commonly used to hire suitable employees but this process is susceptible to strategic misrepresentation by job-seekers. This paper uses a lab experiment as an analogy of such a hiring process by using the Public Goods Game (PGG) as a proxy for a cooperative work environment. Subjects first complete a Big Five personality test, focusing on the trait of "Agreeableness", which previous studies have linked to prosocial cooperation in the PGG. Two groups are formed: a high Agreeableness group and a low Agreeableness group. The high Agreeableness group should contribute more to the public good. The experiment manipulates the timing of revealing the group formation rule, as knowing the rule before the personality test allows for misrepresentation of Agreeableness. I find no evidence of misrepresentation when the group formation rule is revealed before the personality test. I also find that Agreeableness group formation increases contributions for both high and low groups, but only when it is described to subjects before the PGG. Contrary to the existing literature, I find no evidence that Agreeableness influences contributions in the PGG.

*MQBS Experimental Economics Laboratory, Macquarie Business School, Macquarie University, Sydney, Australia. Email: daniel.woods@mq.edu.au. Funding from the IFREE Small Grants Program is gratefully acknowledged. I thank Annika Kieninger and Matthias Waldauf for excellent research assistance. Helpful comments and feedback were received from Tim Cason, Raphael Epperson, David Gill, Elisabeth Gsottbauer, Stanton Hudja, Christian König-Kersting, and Andrew McGee. An exhaustive pre-registration (conducted in the spirit of a Registered Report) is publicly available at https://doi.org/10.17605/OSF.IO/YWM64 and the associated project is at http://doi.org/10.17605/OSF.IO/MDB7A.

1 Introduction

Psychometric tests, designed to measure a person's personality or other latent aspects that cannot be directly observed, are an established standard in many firms' hiring procedures. Firm hiring processes are so intertwined with psychometric testing, it even pervades its dictionary definition.[1] Psychometric tests are used on approximately 60 to 70% of US job-seekers (Weber & Dwoskin, 2014), and 75% of international firms either use or plan to use them in the future (Kantrowitz, Tuzinski, & Raines, 2018). Psychometric tests are a multi-billion dollar industry globally, with expenditure reaching 12.32 billion USD in 2021 and forecast to hit 23.28 billion in 2030 (Emergen Research, 2022). It remains an open question whether this expenditure is justified, as job-seekers have an incentive to misrepresent their true personality in these tests in order to increase the likelihood of getting a job. It has also been suggested that psychometric testing is unfair or discriminatory against minorities or those with disabilities (Weber & Dwoskin, 2014; Hawkins & Monroe, 2021; McGee & McGee, 2025). All of these elements illustrate the economic importance of psychometric testing, and why it is crucial to understand their efficacy, impacts, and any unintended consequences.

In this paper, I design a laboratory experiment to evaluate under what conditions personality testing may be effective. Lab experiments are becoming a common method to help inform firm personnel and hiring processes. Experiments are a cost-effective tool to evaluate potential firm policies and why they work, without confounds like employee self-selection and while still retaining some external validity (Villeval, 2016). I design the lab experiment to be analogous to 'hiring' using personality testing, and the subsequent 'work effort', at least in so far as a lab experiment permits. The experiment consists of two main parts, a personality test followed by a cooperation task. For the personality test, I elicit the 'Big Five' personality traits, which are widely employed in both hiring procedures and academic research in economics.[2] For the cooperation task, I use the Public Goods Game (PGG) as an representation of a cooperative work environment. In the PGG subjects can make socially-optimal contributions to a public good, but face a personal incentive to free-ride

[1] *'Psychometric test: A test that is designed to show someone's personality, mental ability, opinions, etc., often used by companies when they are deciding whether or not to employ someone.'* (Cambridge Business English Dictionary, 2023)

[2] The psychometric testing firms Big Five Assessments, Hogan Assessments, and SHL, among others, incorporate elements of the Big Five as part of their battery of psychometric testing services that they offer to firms. For a variety of examples of the Big Five in economics research, see (Bartling, Fehr, Maréchal, & Schunk, 2009; Fréchette, Schotter, & Trevino, 2017; Donato, Miller, Mohanan, Truskinovsky, & Vera-Hernández, 2017; Gill & Prowse, 2016; Holmén, Holzmeister, Kirchler, Stefan, & Wengström, 2023).

and contribute less. I interpret contributions to the public good as 'work effort', which is something an employer would like to encourage. I focus on the Big Five personality trait of 'Agreeableness', the tendency to act in a cooperative, unselfish manner, as research finds it positively impacts contributions in the PGG and other similar social dilemmas (Perugini, Tan, & Zizzo, 2010; Volk, Thöni, & Ruigrok, 2012; Kagel & McGee, 2014; Thielmann, Spadaro, & Balliet, 2020). I sort subjects into groups for the PGG based on their Agreeableness score, to mimic the role of an employer hiring based on personality tests in an attempt to maximize their firm's success.

The crucial treatment dimension in the experiment is the timing of information about the purpose of the initial personality questionnaire, i.e., the group formation rule for the PGG. There are three treatments on the time dimension, *Before* the personality test, *After* the personality test (but before the PGG), and *Never*. In the *Before* treatment, subjects have an incentive to misrepresent their personality in order to try and get into a more cooperative group. Whereas in the *Never* treatment, subjects are never informed about how groups are formed, and therefore have no material incentive to misrepresent their personality. Finally, in the *After* treatment, subjects also have no material incentive to misrepresent their personality as the group formation rule is only revealed directly after the personality test. Additionally, subjects know about the group formation rule, meaning they know they are grouped with similarly cooperative people in the PGG. Strategic misrepresentation should reduce the effectiveness of the group formation rule in increasing contributions in the PGG due to the compression of Agreeableness scores and potential mistrust. Whereas, knowledge of the rule could increase its effectiveness by reducing uncertainty about group members' cooperativeness, suggesting *After* > *Never* > *Before* in terms of effectiveness.

The second treatment dimension is the group formation rule itself. Groups are typically randomly assigned in economics experiments, which makes a *Random* treatment a natural baseline for the *Agreeableness* group formation rule. The experiment is a 3x2 design, so subjects in the *Random* treatment also have the group formation rule revealed to them either *Before* or *After* the personality test, or *Never*. With this battery of treatments, I aim to address the following research questions:

Question 1 *To what extent do individuals misrepresent their personality when they have strategic reasons to do so?*

Question 2 *Can personality tests be effective in encouraging cooperative behavior in the PGG?*

Question 3 *Does using personality tests in an unexpected way influence responses in later tests?*

I address Question 1 by comparing the responses to the personality test between the treatment with an *Agreeableness* group formation rule that is revealed *Before* to all other treatments, as strategic misrepresentation can only be present in the former. I find no evidence of misrepresentation of any personality trait in the *Before* treatment.

I address Question 2 by comparing the impact of each treatment dimension on contributions in the PGG while holding the other dimension fixed. This approach allows me to isolate and identify the most significant empirical factors influencing behavior. I find that the *Agreeableness* group formation rule increases contribution rates in the *Before* and *After* treatments. However, I also find that contributions increase regardless of whether the group is of high or low Agreeableness, and I find no evidence that the *Agreeableness* group formation rule is effective in the *Never* treatment. Therefore, sorting groups by Agreeableness is ineffective by itself. Rather, it is the knowledge that groups will be formed by Agreeableness that drives increased contributions.

I answer Question 3 by conducting another personality test after the PGG, and focus on subject responses when the *Agreeableness* group formation rule is revealed *After* the personality tests, as these subjects have their personality responses used in an unannounced way. Question 3 addresses an important methodological question in experimental economics: whether the unexpected use of previous responses changes how subjects behave in the future. I find no evidence that withholding information about the group formation rule until *After* the personality test affects responses to subsequent personality tests. Unless contradicted by future evidence, this design feature remains an appropriate option for experimental economists when their research question requires it.

1.1 Contribution to the Literature

I contribute to the voluminous literature on the PGG, and the related Repeated Prisoner's Dilemma (RPD).[3] A typical pattern of behavior in the PGG starts out with average contributions to the public good of around 50%, which decays steadily over time (Ledyard, 1995; Chaudhuri, 2011; Villeval, 2020). The socially optimal contribution level is 100%, but subjects face an individual incentive to free-ride off the contributions of others.

[3]When I refer to the PGG in this paper, I am using this as a shorthand reference for the commonly studied Linear Voluntary Contribution Mechanism. It is worth noting other forms exist (Ledyard, 1995).

One area of research in the PGG has centered on mechanisms or interventions aimed at enhancing contributions in the PGG. Examples include allowing for punishment (Fehr & Gächter, 2000) or facilitating endogenous group formation (Ahn, Isaac, & Salmon, 2009; Charness & Yang, 2014). Charness, Cobo-Reyes, and Jiménez (2014) find that inducing group identity through a team-building word task increases contributions in a PGG with endogenous group formation. Grouping by personality could also induce a group identity, and so this paper permits a conceptual replication by using an alternative method to induce group identity with exogenous fixed groups in the PGG. Drouvelis, Metcalfe, and Powdthavee (2015) find that priming subjects by using a word search task that contains words relating to cooperation is effective in increasing contributions in a one-shot PGG. In my experiment, subjects could be primed by the positive words used to describe personality traits, so this paper contributes by considering the persistence of the effect of an alternative method of priming in a repeated PGG.[4] I contribute to this overall 'intervention' strand of the PGG literature by considering whether personality sorting can increase contributions with exogenous fixed groups over multiple rounds.

Prior studies on exogenously imposed groups in the PGG have sorted subjects based on their previous contribution behavior, and found that this type of sorting is effective (Burlando & Guala, 2005; Gächter & Thöni, 2005; Gunnthorsdottir, Houser, & McCabe, 2007; Ones & Putterman, 2007). I contribute to this line of literature by sorting subjects in the PGG by something other than their contribution rates, namely their personality traits. In the RPD, Proto, Rustichini, and Sofianos (2019) find that sorting by the Big Five trait of Conscientiousness is effective in encouraging cooperative behavior, suggesting something similar should be possible in the PGG. Typically in these experiments that exogenously sort subjects, information about the sorting rule is withheld. In all of the mentioned research, the information (or lack thereof) provided about the sorting rule remains constant across all treatments. I contribute to this literature by varying the timing and presence of information about the sorting rule, and examining how this affects contributions in the PGG.

Another strand of the PGG literature considers the effects that individual characteristics have on contribution behavior in the PGG. Of particular interest to the current paper are studies that elicit Big Five personality characteristics.[5] The Big Five personality trait of Agreeableness has

[4]Note that the links of this paper to group identity and priming in the PGG were formed based on the results, and were not part of the initial pre-registered motivation or experimental design goals.

[5]Some other relevant papers on individual characteristics and the PGG are (Anderson, Mellor, & Milyo, 2004; Carpenter, Daniere, & Takahashi, 2004; Catola, D'Alessandro, Guarnieri, & Pizziol, 2021).

been found to be a significant predictor of contribution behavior in the PGG. Volk et al. (2012) and Perugini et al. (2010) find Agreeableness to be correlated with contributions in the PGG, although the latter only observe this relationship in men. In terms of Agreeableness in the RPD, Kagel and McGee (2014) find a positive correlation with cooperation, while Proto et al. (2019) observe this only in early periods. Additionally, Gill and Rosokha (2024) find that the trust facet of Agreeableness facilitates cooperation through learning. There are also non-incentivized studies that find a positive relationship between Agreeableness and prosocial actions in cooperative games (for a meta-analysis that includes both incentivized and non-incentivized studies, see Thielmann et al. (2020)). I aim to tackle the logical next question in this line of research: given our understanding that Agreeableness influences contributions, how can we leverage this insight? Creating PGG groups by Agreeableness in order to improve contributions is a natural next step, and is analogous to role of employers using personality testing to select well-suited employees.

The most closely related paper is by McGee and McGee (2024) (henceforth MM). In their experiment, they first elicit subjects' Big Five personality traits in an initial baseline session. In a follow-up session a week later, subjects complete a second Big Five assessment. Before taking the second personality test, subjects are informed that they will receive an extra payment if they are 'hired' for a hypothetical job. The hiring process is based in part on their Big Five characteristics as elicited in the second personality test. Subjects are given a job description that is designed to indicate that Big Five personality trait of Extraversion would be ideal.[6] MM find that subjects misrepresent their personality in the presence of incentives.

I take a different approach from MM which makes a complementary but distinct contribution to the literature. Firstly, my focus is on how misrepresentation impacts subsequent behavior. To continue the job analogy, an employee is likely to behave differently if they suspect their colleagues are manipulative, dishonest, and/or ill-suited for their roles due to misrepresentation. MM focus on the magnitude of misrepresentation given the incentive of being hired for a job that is never undertaken. Whereas, I extend the analogy to include the ensuing job effort decisions, to consider the effects of misrepresentation. Secondly, I consider the misrepresentation of personality traits in a between-subject design rather than within-subjects as in MM. In this regard, I follow the experimental literature on dishonesty, which emphasizes that dishonest behavior is difficult to observe at the individual level (Fischbacher & Föllmi-Heusi, 2013). Comparing subject responses

[6]MM also use a job description aimed at Introversion as well as a neutral description as robustness checks.

across two personality tests introduces a potential confounding factor through subjects' concern about substantial misrepresentation being detected. Finally, this paper contributes by providing a conceptual replication of some of the elements in MM.

I also make two important methodological contributions with this paper. Firstly, my experimental design permits a test of whether 'unexpected data use' (Charness, Samek, & van de Ven, 2022) influences subjects' future decisions. Unexpected data use is when responses are used in a way not described to subjects when they provided that data. This could cause problems for a similar reason as to why deception is not used in economics experiments - a loss of control over subjects' beliefs and expectations (Cooper, 2014; Cason & Wu, 2019). If a subject does not believe all of what they are told in experiments, then they would not always reveal what they would do if the situation was exactly as described. Charness et al. (2022) find that researchers consider unexpected data use as useful and not deceptive, but that student subjects' views differ, making it important to examine whether subjects change their behavior after its use. Secondly, I used a unique form of pre-registration that consisted of a complete paper similar to how it is currently written, alongside the code that conducted the statistical tests and generated the tables and figures.[7] I also publicly uploaded data after each day of collection. Widespread use of this level of pre-registration could help reduce the current credibility crisis in the social sciences (Butera, Grossman, Houser, List, & Villeval, 2020).

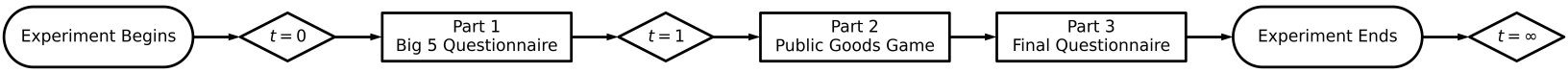
2 Experimental Design

I first provide a brief overview of the experiment, before describing the finer details. The experiment consists of three parts that are common to all treatments. Part 1 is a Big Five questionnaire, Part 2 is a PGG, and Part 3 is a short questionnaire that elicits four other personality traits. The first treatment dimension is how groups are formed in Part 2, the PGG. In the *Random* treatments, groups of three are formed randomly from all subjects in the session. In the *Agreeableness* treatments, subjects are first randomly shuffled into silos of six. Within each silo, the three subjects with the highest Agreeableness scores (as elicited in Part 1) are assigned to one group, while the remaining three are assigned to another group. The second treatment dimension is the timing of when information about the group formation rule is provided. This is either *Before* Part 1 ($t = 0$),

[7] Due to the nature of the publication process, multiple changes have been made. Major deviations are noted in the text, but see Appendix A for a more exhaustive list.

After Part 1 but before Part 2 ($t = 1$), or *Never* ($t = \infty$). An illustration of the timing of the experiment is presented in Figure 1, and a summary of the 3x2 design is provided in Table 1.

Figure 1: Timeline of Experiment



A diamond ($\diamond$) represents a possible treatment point at which the group assignment rule for Part 2 is revealed.

Table 1: Treatments

	Before ($t = 0$)	*After* ($t = 1$)	*Never* ($t = \infty$)
Agreeableness	*Agreeableness Before*	*Agreeableness After*	*Agreeableness Never*
Random	*Random Before*	*Random After*	*Random Never*

2.1 Part 1 - Big Five Elicitation

Part 1 consists of 50 questions designed to elicit the Big Five personality traits (McCrae & John, 1992). These traits are Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.[8] In all treatments, subjects are given a short overview of the PGG in Part 2 before completing the Part 1 questions.[9] If information about the group formation rule is provided (i.e. in *Before* or *After*), it is provided either directly before or directly after subjects complete the 50 questions. In the *Agreeableness Before* and *After* treatments, subjects are presented with the following message:

> For Part 2, you will be assigned to a group of three **based on your 'Agreeableness' score. Your Agreeableness score is determined by your responses to particular questions in Part 1.**
>
> Agreeableness is a personality trait where people high in Agreeableness are often described as *selfless, trusting, good-natured, generous, and forgiving.* (Costa, McCrae, & Dembroski, 1989)
>
> In scientific studies, **a high level of Agreeableness has been found to have a positive effect on group cooperation decisions** similar to the type in Part 2. [References button with pop-up window]

[8] See Appendix C for full details on the personality trait questions and their scoring.
[9] The instructions for Parts 1 and 3 are presented in Appendix E.

Each group of three is formed from six randomly selected subjects. **The three subjects with the highest Agreeableness scores will be assigned to one group, and the remaining three subjects to the other group.**

A high level of detail is provided to help subjects understand the specific personality trait that is being used in the group formation rule, and why it could be beneficial or desirable to be in the high Agreeableness group. In the *Agreeableness* treatments subjects are not told whether they are in the high or low Agreeableness group, even in *Before* and *After*, in order to facilitate comparisons with *Never*.[10]

2.1.1 Predictions: Misrepresentation

When it comes to strategic misrepresentation in the Big Five questionnaire of Part 1, there are three treatment groups of interest. The first are those that know in advance that their Part 1 responses will be used to form groups in Part 2 (*Agreeableness Before*). The second are those that know in advance that their Part 1 responses will not be used to form groups in Part 2 (*Random Before*). The final group are those that do not know in advance about the group formation rule in Part 2 ($t > 0$). The first two groups are aware of how their Part 1 responses affect Part 2 while answering Part 1, while the third group is unaware of this while answering Part 1.

I propose two behavioral channels that could influence Part 1 responses: the incentive to misrepresent Agreeableness, and the suspicion that Part 1 answers may be used in some way for Part 2. An incentive to misrepresent Agreeableness exists when it is known groups will be formed based on this trait. Suspicion occurs only when subjects are not aware of the purpose of the questionnaire. Subjects may believe (sometimes correctly) that the questionnaire will be used in some relevant way in the future, as they know there will be a following Part 2. Each of the comparisons between the relevant groups and the differences in operative channels between them are summarized in Table 2.

Table 2 demonstrates comparisons that isolate either channel: incentives through comparing *Agreeableness Before* and *Random Before*, and suspicion through *Random Before* and *After* & *Never*. Both channels have the potential to influence Agreeableness. Incentives should increase the reported Agreeableness scores, as subjects will prefer to be in H groups (or avoid L groups).

[10]The difference between *Agreeableness After* and *Agreeableness Never* treatments is knowledge of the group formation rule (see Table 3). Adding an additional 'knowledge of group type' would confound this comparison with an additional channel in which behavior could differ.

Table 2: Misrepresentation of Agreeableness - Treatment Comparisons

Treatment Comparison	Incentive	Suspicion
Agreeableness Before to *Random Before*	−	0
Random Before to $After$ & $Never$	0	+
Agreeableness Before to $After$ & $Never$	−	+

Going from the first treatment to the second, + indicates that channel has been added, 0 indicates no change, and − indicates that channel has been taken away. The $After$ & $Never$ grouping includes all treatments except for *Agreeableness Before* and *Random Before*.

I propose that suspicion leads to more socially desirable responses, thereby increasing reported Agreeableness scores. There are many possibilities of what a subject might be suspicious of, but the obvious candidates of group formation or having answers revealed to others in Part 2 would both suggest a tendency towards more socially desirable responses.[11] Hypotheses 1 and 2 formalizes the conjecture that the reported Agreeableness scores in Part 1 in the presence of incentives or suspicion respectively.

Hypothesis 1 *Agreeableness scores are higher in Agreeableness Before than in Random Before*

Hypothesis 2 *Agreeableness scores are higher in $After$ & $Never$ treatments than in Random Before*

2.2 Part 2 - Public Goods Game

Part 2 consists of a PGG adapted from the version used by Lugovskyy, Puzzello, Sorensen, Walker, and Williams (2017). Groups of three are assigned from silos of six subjects by the group formation rule (i.e. randomly or by Agreeableness). Each group of three remains together for 15 'group cooperation decisions'. In each decision, each subject has 25 tokens they can allocate to either a Private account or a 'Cooperation' account.[12] Each token a subject allocates to the Private account earns that subject 10 points. Each token a subject allocates to the Cooperation account earns each of the three group members (i.e. including the subject in question) 4 points each. In other words, one token allocated to the Cooperation account earns the group 12 points overall. I refer to tokens allocated to the Cooperation account as 'contributions'. The marginal per-capita

[11] Both incentives and suspicion also have the potential to influence the other personality traits. Suspicion because it is not known which traits could be used, and incentives if misrepresentation is unsophisticated. Section 3.1.1 also tests the other personality traits.

[12] Framing the PGG in terms of group cooperation is likely to increase contributions (Dufwenberg, Gächter, & Hennig-Schmidt, 2011). This should not be an issue as all treatments are framed in the same way.

return (the ratio of the private benefit of one token to the Cooperation account to the opportunity cost of that token) is $MPCR = \frac{4}{10} = 0.4$. For $MPCR = \frac{4}{10} = 0.4$, it is a well-replicated result that groups' average contribution rates typically start at around 50% and then decline steadily over time (Ledyard, 1995; Chaudhuri, 2011). This $MPCR$ was chosen so that any intervention that increases contributions would have plenty of room to do so without censoring at full contributions. Subjects make their decision by deciding how many tokens to allocate to the Cooperation account, with the remainder being allocated to their Private account. After making their decision, subjects are reminded of their own contribution, and also told the total group contribution in that round. These round summaries are also available at any time during Part 2 in a history table that is displayed at the bottom of the screen.

2.2.1 Predictions: Contributions

One important distinction to make is that in the *Agreeableness* treatments, one group will have higher Agreeableness than the other. The high group is predicted to have higher contributions than the low group. I therefore consider these two types of groups separately, as I would like to observe the positive effects of personality sorting.[13] I denote the two types of groups H and L for high and low Agreeableness respectively. In the following discussion, I take the viewpoint of the H group when describing potential effects.

I conjecture that there are three main factors at play here: the group formation rule itself, strategic misrepresentation of Agreeableness, and knowledge of the group formation rule. The Agreeableness group formation rule should be effective in increasing contributions, as this personality trait is linked with cooperation and generosity. Hypothesis 3 tests this conjecture under each timing condition.

Hypothesis 3 *The number of tokens contributed in Agreeableness H is greater than in Random.*
The number of tokens contributed in Random is greater than in Agreeableness L.
The number of tokens contributed in Agreeableness H is greater than in Agreeableness L.

However, the effectiveness of the Agreeableness group formation rule will differ depending on when information about the rule is revealed. Consider comparing *Before* to *After*, two treatments where subjects know the group formation rule before the PGG. In *Before* the group formation rule

[13] In the employment analogy, the low group would simply not be hired. However, given the expectations of lab subjects this is not practical to implement.

is known prior to when Agreeableness is measured. Subjects have an incentive to misrepresent themselves in the Agreeableness elicitation to try and be placed in the H group (or to avoid the L group). Agreeableness scores would be compressed and the end result would be more similar to random group formation in terms of each group's true level of Agreeableness. Whereas in *After*, the group formation rule is only revealed after the Agreeableness elicitation, precluding strategic misrepresentation. The Agreeableness group formation rule should be more effective in the absence of strategic misrepresentation. In terms of the Random group formation rule, I posit that the timing has no effect. Hypothesis 4 formalizes these conjectures.

Hypothesis 4 *The number of tokens contributed in Agreeableness Before H is lower than in Agreeableness After H.*

The number of tokens contributed in Random Before is the same as in Random After.

The number of tokens contributed in Agreeableness Before L is higher than in Agreeableness After L.

Now consider comparing *After* to *Never*, two treatments that do not have strategic misrepresentation but differ in whether subjects know the group formation rule prior to the PGG. Knowing that the Agreeableness group formation rule is in effect means that subjects are aware they are grouped with similarly cooperative people. Such confidence would increase initial contributions if subjects are concerned about being taken advantage of by lower contributors. Higher initial contributions would have a flow-on effect if subjects are conditional cooperators. Therefore, Agreeableness group formation should be more effective when the rule is known in the absence of strategic misrepresentation. Hypothesis 5 formalizes these conjectures.

Hypothesis 5 *The number of tokens contributed in Agreeableness After H is higher than in Agreeableness Never H*

The number of tokens contributed in Random After is the same as in Random Never

The number of tokens contributed in Agreeableness After L is lower than in Agreeableness Never L

Table 3 presents especially interesting treatment comparisons that isolate the impact of a particular effect while holding other factors constant. This assumes effects are additively separable, but potential interactions means the full 3x2 design is prudent.

Table 3: Efficiency - Selected Treatment Comparisons

Treatment Comparison	Incentive to misrepresent	Knowledge of group formation rule	Agreeableness group formation
Agreeableness Before to *Agreeableness After*	–	0	0
Agreeableness After to *Agreeableness Never*	0	–	0
Agreeableness Before to *Random Before*	–	0	–
Agreeableness After to *Random After*	0	0	–
Agreeableness Never to *Random Never*	0	0	–

Going from the first treatment to the second 0 indicates no change, and – indicates that channel has been taken away.

2.3 Part 3 - Final Questionnaire

In Part 3, subjects are first informed that they are to complete a final survey, and that their final earnings for the experiment have already been set. Subjects then answer 16 personality questions and a standard demographic questionnaire. The 16 questions elicit the three elements of the 'Dark Triad' (Paulhus & Williams, 2002), and the 'Sincerity' and 'Fairness' facets of the 'Honesty-Humility' trait from 'HEXACO' (Ashton & Lee, 2009). The three Dark Triad measures are 'Machiavellianism' (Christie & Geis, 1970), 'Narcissism' (Raskin & Hall, 1979), and 'Psychopathy' (Hare, 1985). Machiavellianism is marked by a calculating, manipulative, and deceitful nature towards other people. Narcissism is defined as being egotistic and prideful with limited empathy for others. Psychopathy is characterized by selfishness, impulsiveness and a lack of remorse for ones actions. Honesty-Humility is a personality trait where people avoid manipulating others for personal gain, and feel little temptation to break rules.

Part 3 provides a very conservative test of whether unexpected data use affects subjects' subsequent responses. It is conservative as subjects are explicitly informed that Part 3 is the last part of the experiment and that their final payments are already set. If this statement is taken seriously, then subjects have no material incentive to misrepresent their personality in their Part 3 responses. However, in the *Agreeableness After* treatment, information about how the earlier Part 1 responses would be used in Part 2 was initially withheld and then later disclosed to subjects. The unexpected data use from Part 1 may cause subjects to change their Part 3 responses in anticipation of additional unexpected data use, despite explicit statements to the contrary. It would be concerning if subjects in the *Agreeableness After* treatment responded in a different fashion than those in the other treatments, as it would imply a loss of experimental control. Such a finding would raise strong objections about using unexpected data use as a design feature in economics

experiments going forward.

The traits elicited in Part 3 all have a clear direction in terms of social desirability. Narcissism, Machiavellianism, and Psychopathy are clearly negative traits from the perspective of society, while Honesty/Humility is considered a positive trait. I propose that if a subject anticipates unexpected data use, then they would misrepresent themselves towards what is more socially desirable. I propose two channels that would influence a subject's beliefs that their Part 3 responses will be used to affect something in the experiment. The first channel is whether subjects are aware that the data from personality questions have been used for something in the experiment. These are subjects in the *Agreeableness Before* and *Agreeableness After* treatments, as they know the group formation rule in Part 2 was based on their Agreeableness score from Part 1. The subjects in the other treatments remain *Unaware* that personality responses could be used in other parts of the experiment. Subjects that know their personality questions in Part 1 were used in Part 2 could suspect that their personality responses in Part 3 are also used in some fashion, and misrepresent themselves accordingly. The second channel is whether the use of the personality data was unexpected. Subjects in *Agreeableness Before* expected this data use when completing Part 1, as they were told of the Agreeableness group formation rule in advance. Whereas, subjects in *Agreeableness After* did not expect it, but found out about it after completing Part 1. Subjects in *Agreeableness After* may think that their Part 3 responses will be used in some way that has not yet been revealed, and thus would be the most likely to misrepresent themselves in Part 3. Table 4 describes which channels are present between each group of treatments.

Table 4: Misrepresentation in Part 3 - Treatment Comparisons

Treatment Comparison	Unexpected Data Use Revealed	Knowledge of Personality Data Use
Agreeableness Before to *Agreeableness After*	+	0
Agreeableness Before to *Unaware*	0	−
Agreeableness After to *Unaware*	−	−

Going from the first treatment to the second, + indicates that channel has been added, 0 indicates no change, and − indicates that channel has been taken away. The *Unaware* grouping includes all treatments except for *Agreeableness Before* and *Agreeableness After*.

I aggregate each individual into one measure of 'Positive Perception', which positively weights Honesty/Humility and negatively weights the Dark Triad traits. Based on my previous reasoning, I posit the following Hypotheses about Positive Perception:

Hypothesis 6 *Reported Positive Perception is higher in Agreeableness After than in Agreeableness Before*

Hypothesis 7 *Reported Positive Perception is higher in Agreeableness After than in Unaware treatments*

Hypothesis 8 *Reported Positive Perception is higher in Agreeableness Before than in Unaware treatments*

2.4 Procedures

The data collection was conducted at the EconLab at the University of Innsbruck (UIBK) and the Vienna Center for Experimental Economics (VCEE) at the University of Vienna. 20 sessions (366 subjects) were run at the UIBK EconLab, and 4 sessions (66 subjects) were run at VCEE.[14] There are observations from 432 subjects, i.e., 144 groups of three. Each *Random* treatment has observations from 16 groups of three, and each *Agreeableness* treatment has observations from 32 groups of three. I collected a different number of groups by treatment as observations in the *Agreeableness* treatments are split between L and H groups.

Subjects were recruited using the online database hroot (Bock, Baetge, & Nicklisch, 2014) at the UIBK EconLab, and ORSEE at VCEE (Greiner, 2015). The experiment was computerized using oTree (D. L. Chen, Schonger, & Wickens, 2016). A session consisted of 6, 12, 18, or 24 subjects (depending on how many subjects showed up for a session), as multiples of six are required for the *Agreeableness* treatments.[15] All subjects within a session faced the same treatment, which was randomly assigned.[16] Subjects earned points over 15 rounds of the PGG described in Section 2.2, subjects remained in the same group of 3 for all 15 rounds, and subjects were paid based on their cumulative earnings. Points were converted at a rate of 1000 points = 3 Euros, and subjects received a show-up fee of 4 Euros. Subjects were required to correctly answer 7 comprehension questions about the PGG immediately before the PGG began. The experiment took 45-60 minutes and the average earnings were 16.03 Euros.

[14] I thank the lab managers and research assistants at VCEE. An additional lab was used due to exhausting the subject pool at UIBK. Using another German-speaking lab in this situation was pre-registered.

[15] *Random* sessions also used multiples of six for consistency despite only needing multiples of three.

[16] See the OSF project http://doi.org/10.17605/OSF.IO/MDB7A for the randomization implementation.

3 Results

A summary of key variables are reported in Table 5, while the others are reported in Table A5.

	Rand. Before	*Rand. After*	*Rand. Never*	*Agre. Before*	*Agre. After*	*Agre. Never*
PGG Contribution (Low Agre. High Agre.)	6.91 (8.16)	8.00 (8.90)	5.65 (6.80)	11.25 (9.29) 10.35 (9.45)	8.79 (9.40) 9.37 (8.92)	7.21 (8.72) 6.81 (7.81)
Agreeableness	105.02 (12.23)	106.29 (9.88)	104.33 (13.76)	99.04 (8.26) 115.79 (6.15)	94.06 (9.21) 112.65 (7.34)	99.08 (9.02) 113.81 (7.25)
Open Mindedness	21.69 (4.69)	21.17 (3.50)	20.71 (4.26)	22.54 (4.34) 23.06 (4.21)	22.27 (4.47) 23.42 (4.06)	21.00 (3.62) 21.62 (3.77)
Negative Emotionality	16.48 (5.39)	16.27 (4.83)	15.38 (4.57)	16.29 (5.13) 14.98 (5.13)	15.81 (3.76) 15.29 (4.57)	15.67 (5.03) 15.81 (4.94)
Extraversion	19.44 (4.05)	20.12 (5.32)	20.88 (4.22)	19.75 (4.69) 20.35 (4.41)	20.67 (4.55) 20.94 (4.28)	20.83 (3.59) 22.00 (3.84)
Conscientiousness	22.25 (3.89)	20.06 (4.15)	21.94 (4.62)	21.35 (4.07) 22.56 (3.63)	19.27 (4.52) 21.96 (3.92)	21.71 (4.13) 22.58 (3.34)
Earnings	€15.92 (1.01)	€16.01 (1.09)	€15.80 (0.88)	€16.30 (1.36) €16.22 (1.24)	€16.08 (1.52) €16.14 (1.15)	€15.94 (1.18) €15.90 (1.09)
Num. Subjects	48	48	48	48 48	48 48	48 48

Table 5: Summary Statistics by Treatment and Agreeableness Group Type

Standard deviations in parentheses. Low and High Agreeableness groups reported separately for *Agreeableness* treatments. PGG Contribution is reported at the individual per round level. Agreeableness $\in [26, 130]$ and all other personality traits $\in [6, 30]$. Abbreviations: *Rand.* = *Random*, *Agre.* = *Agreeableness*.

The statistical analysis proceeds as follows. Section 3.1 presents pre-registered analysis, while Section 3.2 reports exploratory analysis.[17]

[17]Section 3.1 presents all analysis that was in the main body of the pre-registration document, except for the tests

3.1 Pre-registered Analysis

3.1.1 Part 1: Strategic Misrepresentation of Agreeableness

For misrepresentation of Agreeableness, there are three comparison groups: *Agreeableness Before*, *Random Before*, and all $After$ & $Never$ treatments, because treatment differences can only impact Part 1 responses if they occur before Part 1. The outcome of interest is each subject's Agreeableness score, calculated from their responses to the Part 1 questions. The comparison between *Random Before* and *Agreeableness Before* tests Hypothesis 1, whether subjects misrepresent their Agreeableness when they have the incentive to do so. The comparison between *Random Before* and all *After & Never* treatments tests Hypothesis 2, whether subjects misrepresent their Agreeableness when they could be suspicious about how their personality responses are used. I report these results alongside summary statistics in Figure 2.[18]

Figure 2 shows that Agreeableness scores do not substantially differ between treatments. This is evidence against Hypothesis 1, which suggests subjects do not misrepresent their Agreeableness when they know they will be grouped by it in the PGG. *Agreeableness Before* looks slightly higher in the direction we would expect if there were strategic misrepresentation, however, the magnitude is small and not statistically significant.[19] Figure 2 also provides evidence against Hypothesis 2, which suggests that any suspicion about how their responses may be used does not impact how subjects answer Agreeableness questions. In Appendix B.3.1 I also consider 'unsophisticated' misrepresentation of the other Big 5 traits, but find no differences in most comparisons.[20]

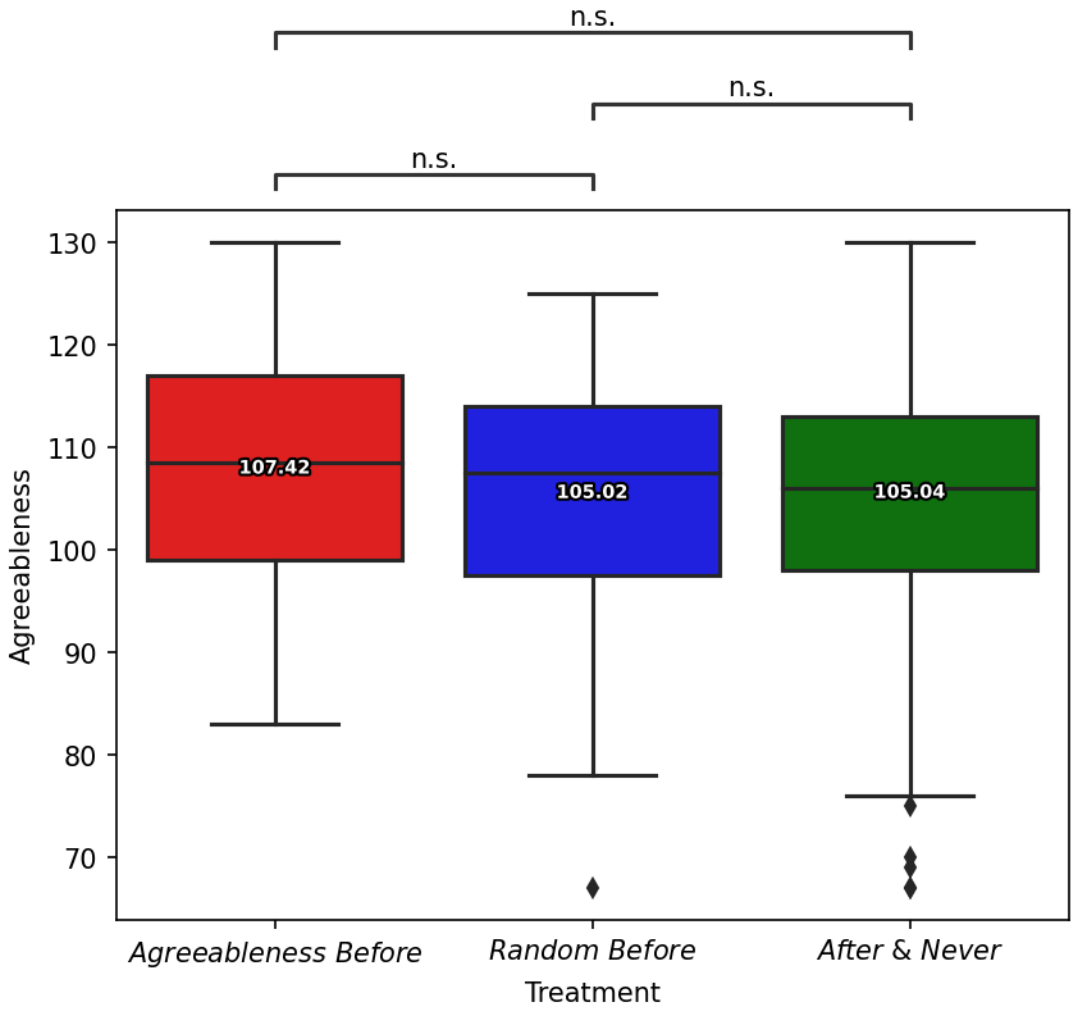
on unsophisticated misrepresentation, which is still discussed but the results are instead in Appendix B.3.1. Section 3.1 also discusses the effect of Agreeableness on PGG contributions and reports the associated Table 7 which was originally in the appendix of the pre-registration document.

[18]Throughout this paper I report p-values from conservative two-sided tests.

[19]Using a less conservative (and not pre-registered) Tobit regression yields p=0.25 and p=0.07 for *Agreeableness Before* to *Random Before* and *After & Never* respectively (Table A25). Low power may account for the failure to detect misrepresentation, but the small magnitude questions its empirical importance.

[20]Appendix B.3.2 also reports an exploratory robustness check of misrepresentation at the individual question level. Only one difference is significant after the large number of comparisons are corrected for.

Figure 2: Agreeableness by Treatment

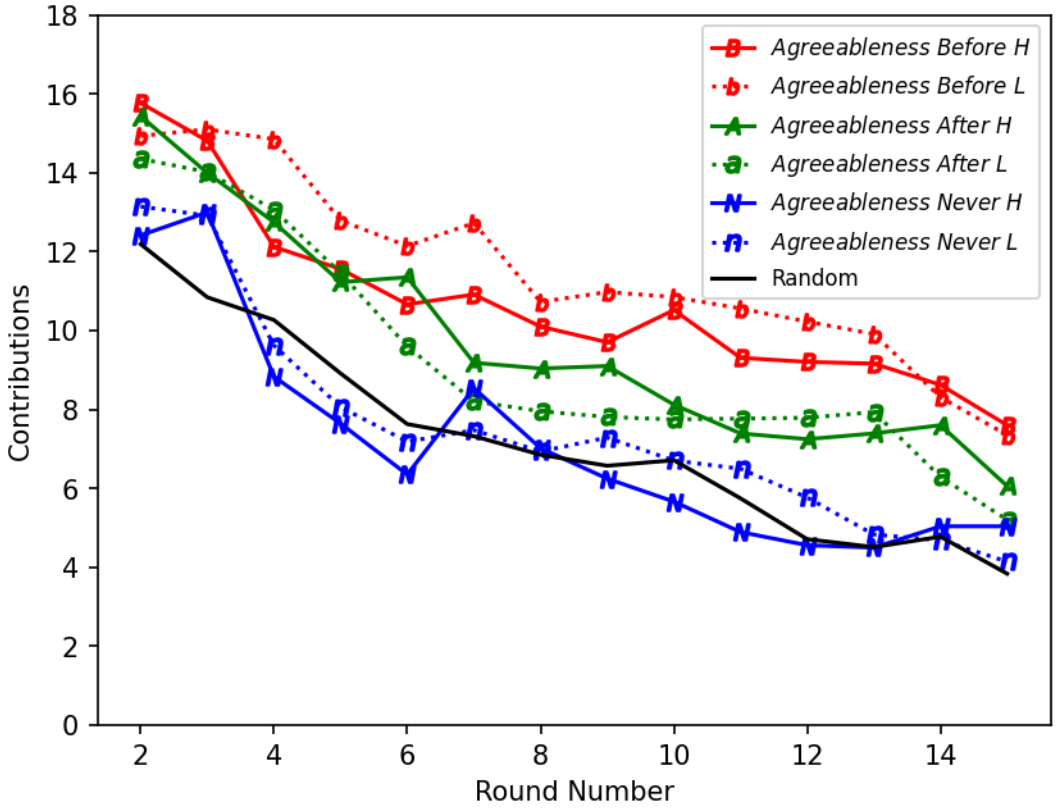


Mean Agreeableness overlaid. Statistical results are based on a Mann Whitney test. The three lines in the box are the 75%, 50% and 25% quartile when going from top to bottom, the top (bottom) whisker is the largest (smallest) value that is below (above) 1.5 times the difference between the 75% and 25% quartiles, and values outside this range are diamonds. ***=$p < 0.01$, **=$p < 0.05$, *=$p < 0.10$, and n.s.=$p \geq 0.10$.

3.1.2 Part 2: PGG Contributions

Figure 3 graphically illustrates average contributions over time by treatment. Figure 3 shows declining contributions over time, which is typical in this type of PGG. Figure 3 shows that *Agreeableness Before* and *Agreeableness After* increase contributions by about 2-4 tokens on average when compared to *Random*, and that this level shift is sustained over time. However, this increase is observed regardless of whether the group is of H or L Agreeableness. In addition, *Agreeableness Never* appears to be mostly ineffective as it is quite close to *Random*. From the patterns observed in Figure 3, it follows that the Agreeableness group formation rule does not increase contributions

Figure 3: Average Contributions by Round



by creating groups with high levels of Agreeableness. Rather, the increased contributions are driven by describing the personality trait of Agreeableness and how previous studies have found it to have a positive impact on prosocial actions.

Table 6 summarizes the results from the comparisons that are relevant for testing the numbered Hypotheses. Table 6 provides no support for Hypotheses 3, whether created H Agreeableness groups contribute more in the PGG, and whether created L Agreeableness groups contribute less. Although two treatment differences show statistical significance, their directions contradict the part of the hypothesis that states that L groups should contribute less. There is no evidence from *Agreeableness* groups to support Hypothesis 4, which states that contributions change when the group formation rule is known *Before* the personality test due the presence of strategic misrepresentation, or Hypothesis 5, which states that contributions change when the group formation rule is *Never* known due decreased confidence of being grouped with similarly cooperative individuals. For *Random* groups, there is support for the parts of Hypotheses 4 and 5 that posit that the timing of revealing the group formation rule is irrelevant in the *Random* treatments. Overall, Table 6 partially confirms what Figure 3 suggested, in that Agreeableness sorting can have a positive

impact regardless of group composition, but only when it is described before playing the PGG.

Table 6: Efficiency - Regressions

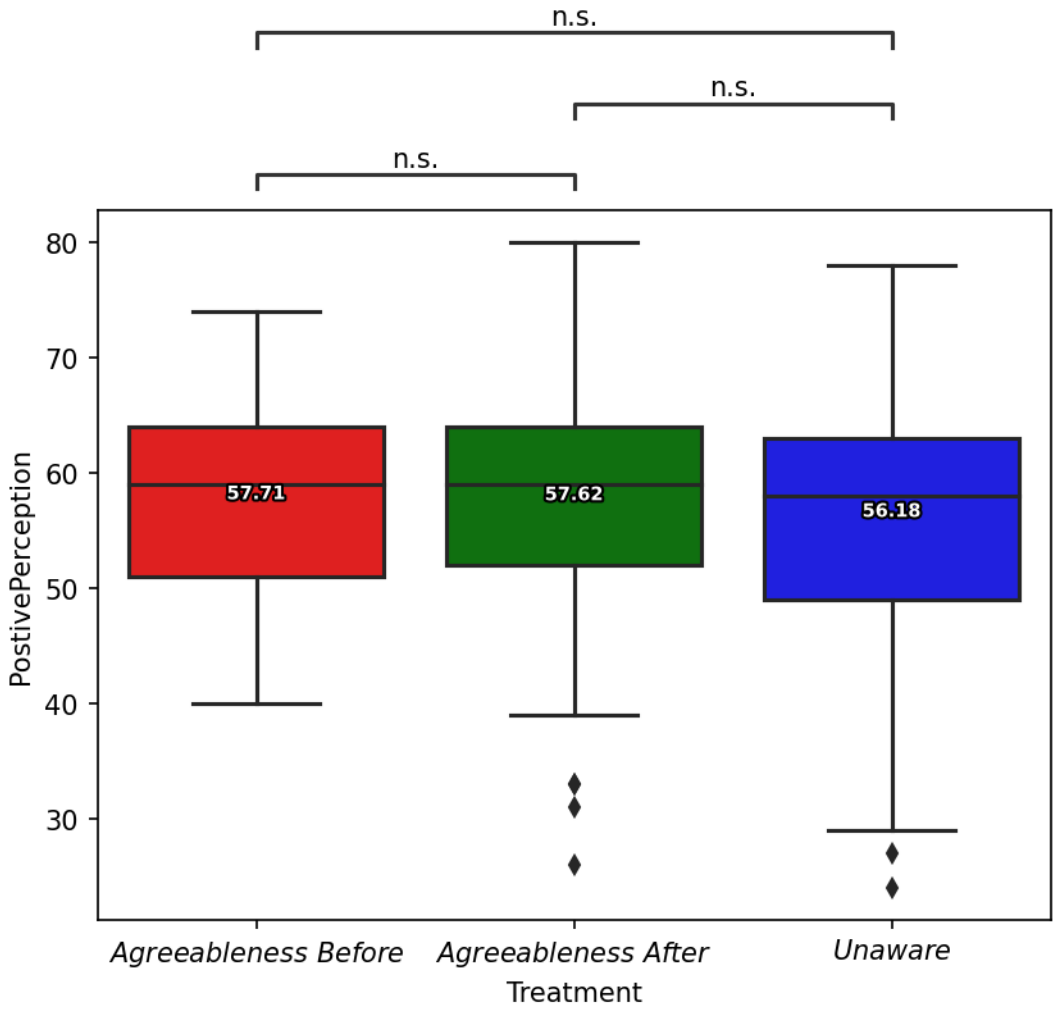
Pairwise Comparison	Coefficient	Hypothesis
Within Before		
Agreeableness Before H - Random Before	4.34*	H3 +
Agreeableness Before L - Random Before	5.19**	H3 −
Agreeableness Before H - Agreeableness Before L	-0.81	H3 +
Within After		
Agreeableness After H - Random After	2.18	H3 +
Agreeableness After L - Random After	1.36	H3 −
Agreeableness After H - Agreeableness After L	0.85	H3 +
Within Never		
Agreeableness Never H - Random Never	1.65	H3 +
Agreeableness Never L - Random Never	1.59	H3 −
Agreeableness Never H - Agreeableness Never L	0.16	H3 +
Within Agreeableness		
Agreeableness Before H - Agreeableness After H	1.26	H4 −
Agreeableness Before L - Agreeableness After L	2.97	H4 +
Agreeableness After H - Agreeableness Never H	2.78	H5 +
Agreeableness After L - Agreeableness Never L	2.19	H5 ∼
Within Random		
Random Before - Random After	-0.92	H4 ∼
Random After - Random Never	2.42	H5 ∼

Treatment coefficient (β_1) reported from panel Tobit regression at the group level of the form $AverageGroupContribution = \beta_0 + \beta_1 Treatment + \beta_2 Round$. Full regression output and details are in Appendix D.1.1. Second group in the pair is the omitted dummy. *Within X* are the groups of hypotheses holding X fixed. +, −, and ∼ indicate a positive, negative, or neutral predicted effect respectively. ***=$p < 0.01$, **=$p < 0.05$, and *=$p < 0.10$.

3.1.3 Part 3: Other Personality Measures

The main test in Part 3 is to detect whether the 'data use' of Part 1 responses to sort groups in Part 2 affects the responses in Part 3. There are three groups, *'Expected'* data use (of their Part 1 personality responses) (*Agreeableness Before*), *'Unexpected'* data use (*Agreeableness After*), and all treatments where subjects are *Unaware* of data use. I combine all of the personality traits elicited in Part 3 into one measure based on how likely it is they would be positively perceived by an observer. I call this combined measure 'Positive Perception'. I test for differences between the three relevant subject groups. Figure 4 summarizes the results of these comparisons. Figure 4 provides no support for Hypothesis 6, which states that revealing Part 1 personality responses were used in an unexpected way changes Part 3 personality responses (holding knowledge fixed:

Figure 4: Positive Perception by Treatment



Positive Perception $\in [16, 80]$, mean overlaid. Statistical results are from a Mann Whitney test. The three lines in the box are the 75%, 50% and 25% quartile when going from top to bottom, the top (bottom) whisker is the largest (smallest) value that is below (above) 1.5 times the difference between the 75% and 25% quartiles, and values outside this range are diamonds. ***=$p < 0.01$, **=$p < 0.05$, *=$p < 0.10$, and n.s.=$p \geq 0.10$.

Agreeableness Before to *Agreeableness After*); Hypothesis 7, which states that knowing Part 1 personality responses were used changes Part 3 personality responses (holding unexpected use fixed: *Agreeableness After* to *Unaware*); or Hypothesis 8, which states that both knowing and having had Part 1 personality responses used in an unexpected way changes Part 3 personality responses (*Agreeableness Before* to *Unaware*). There is no evidence that data use, expected or unexpected, of Part 1 responses influences subject responses to the Part 3 personality test. In the absence of future evidence, these results show that unexpected data use can remain an appropriate tool in experimental economics when the design requires it.

3.1.4 Agreeableness and PGG Contributions

The results suggest that the Agreeableness group formation rule is not effective in increasing contributions by creating groups with higher levels of Agreeableness. This follows from the observation that there is no difference between H and L groups, and that the *Agreeableness Never* treatment is not effective. A natural question is why this is the case, especially considering that previous studies identified a positive relationship between Agreeableness and contributions in the PGG.

To address this question, I regress an individual's Agreeableness on their contributions in the PGG, alongside other personality traits and controls.[21] Table 7 lists the relevant coefficients from this regression, and suggests that an individual's Agreeableness score has essentially no impact on their contribution behavior. This is a surprising result as it is in contrast to previous studies that find Agreeableness is positively related to prosocial actions, including in the PGG. However, it is unsurprising then that the Agreeableness group formation rule by itself proved ineffective in increasing group contributions.

In terms of the other personality traits, Table 7 suggests that higher levels of Open Mindedness increase an individual's contributions. This is a curious result as Open Mindedness has not been found to be related to prosocial behavior in previous economics experiments (e.g. Kagel and McGee (2014), Proto et al. (2019)), and it is not predicted to be related to any of the prosocial games that Thielmann et al. (2020) consider. However, Thielmann et al. (2020) do report a positive correlation of Open Mindedness and prosocial behavior in their meta-study, so more research on this trait may be warranted. Table 7 also suggests that higher levels of Conscientiousness decrease contributions. Proto et al. (2019) also find less cooperation in an RPD by higher levels of Conscientiousness although only when subjects are grouped by this trait, which is not the case in this experiment.

3.2 Exploratory Analysis

3.2.1 Pooled Treatment Comparisons

The results show that the Agreeableness group formation rule does not improve contributions by grouping people with high Agreeableness together, as was initially proposed. Rather, it seems to be that simply providing information about Agreeableness and that it will be used to form groups

[21] This analysis was originally pre-registered in the 'Conceptual Replication' appendix, because it addressed the broader research question of which individual characteristics affect prosocial contribution behavior, which was not initially of primary interest. For the exposition of the paper, I have moved this analysis to this section.

Table 7: Individual Characteristics on Contributions

Ind. Variable	Coefficient
Lagged Avg. Group Cont.	0.55***
Agreeableness	0.00
Open Mindedness	0.29***
Negative Emotionality	0.11
Extroversion	0.13
Conscientiousness	-0.26***
Honesty Humility	-0.07
Machiavellianism	-0.09
Narcissism	-0.13
Psychopathy	-0.08
Female	-0.59
2nd Year at Uni.	0.25
3rd Year at Uni.	-2.34*
4th+ Year at Uni.	-1.67
Grad. Student	-2.26
GPA	0.27
Economics	-0.07
Arts and Humanities	1.78
Natural Sciences	3.29*
Education	3.23
Engineering	4.08
Law	0.09
Social Sciences	0.63
Medicine	0.18
Other	1.05

An individual's contribution to the public good is the dependent variable. Results are from a multilevel panel tobit regression (censored at 0 and 25) with individual and group level random effects. Controls for Treatment and Period are included in the regression but their coefficients are not reported. Observations from the *Agreeableness Before* treatment are excluded. Full regression and output and details are presented in Appendix D.1.2. ***=$p < 0.01$, **=$p < 0.05$, and *=$p < 0.10$.

in the PGG is what increases contributions. Testing whether there are treatment effects under this new conjecture requires different statistical tests from what was originally pre-registered. As this new conjecture makes no distinction between H and L groups, and that empirically speaking these groups exhibit similar behavior, the observations from these groups are combined. A similar rationale applies to pooling data from all *Random* treatments.[22] I report tests of this conjecture with various combinations of pooled observations in Table 8. Table 8 provides evidence in support of the conjecture, and there appears to be an increase of around 3 or 4 tokens on average when

[22]This also has the added benefit of increasing power, which is beneficial as the actual data turned out to be noisier than what was assumed in the initial power analysis.

subjects are told about Agreeableness prior to taking part in a PGG.

Table 8: Efficiency - Exploratory Hypothesis

Pairwise Comparison	Coefficient
Agreeableness Before - Random Before	4.76**
Agreeableness Before - Random	4.93***
Agreeableness After - Random After	1.76
Agreeableness After - Random	2.84*
Agreeableness Before & After - Random	3.89***
Agreeableness Before & After - Agreeableness Never	3.51**

Reported coefficients are from a panel Tobit regression with the group's average contribution as the dependent variable, and a treatment dummy and the period as independent variables. Second group in the pair is the omitted dummy. *Agreeableness Before* and *Agreeableness After* pool observations across H and L groups. *Random* pools all observations from *Random Before*, *Random After*, and *Random Never*. $A_{\{0,1\}}$ pools observations from *Agreeableness Before* and *Agreeableness After*. ***=$p < 0.01$, **=$p < 0.05$, and *=$p < 0.10$. For space, only the treatment coefficient is reported, the full regression output is reported in Appendix D.2.2.

3.2.2 Agreeableness and PGG Contributions

Finally, I examine the robustness of the finding that Agreeableness does not affect contributions, a result that contrasts with the existing literature. For this purpose, I have conducted several robustness checks on the analysis presented in Table 7. There are a variety of different specifications, which I describe in more detail in Appendix B.2, but I briefly explain their rationale here.

Firstly, as there was no detected misrepresentation, then the initial logic behind excluding *Agreeableness Before* observations is no longer valid. Therefore, I re-run the analysis with all data included. Conversely, as Agreeableness Sorting did increase contributions, it is natural to suggest that the subset of data from both *Agreeableness Before* and *Agreeableness After* is different and should be excluded. Thirdly, Perugini et al. (2010) find a relationship between Agreeableness and contributions in the PGG, but only for men. Therefore, I re-run the analysis excluding everyone but men. Finally, Proto et al. (2019) find that Agreeableness only impacts cooperative behavior in initial decisions in the RPD. Therefore, I re-run the analysis using only early subsets of the data. Table A2 presents the results of these analyses, but they all report that the coefficient attached to Agreeableness is effectively zero. The finding that we cannot reject the null hypothesis, i.e. the coefficient on Agreeableness is zero, is robust across various specifications. This suggests that Agreeableness has little to no influence on individuals' contribution behavior in the PGG.

4 Conclusion and Discussion

Using psychometric personality testing when there may be incentives to misrepresent personality is a complex topic. A real-world example of this is job hiring, where personality tests have become integral to the process. However, there is the incentive for job-seekers to tailor their responses to align with the employers' expectations. The incentive to strategically misrepresent personality or other not directly observable traits has the potential to undermine the usefulness of psychometric tests.

To shed light on this issue, I design and conduct an incentivized laboratory experiment that aims to be analogous to real-world hiring scenarios. I first elicit Big Five characteristics through a questionnaire, much like what job-seekers have to fill out at some stage during the hiring process. I then use a standard PGG to represent a cooperative work environment. The Big Five characteristic of Agreeableness has been found to positively impact contributions in previous studies, so sorting (or 'hiring') based on this trait makes sense. By changing the timing of the revelation of the sorting rule to before or after the initial questionnaire, I am able to explore the level of misrepresentation and evaluate its subsequent impact on cooperative behavior.

I find that subjects do not misrepresent their personality in order to be placed into groups with higher levels of Agreeableness. There are two possible explanations for this finding. The most likely case is that a preference for honesty outweighs the indirect benefits of being in a high Agreeableness group.[23] The indirect benefits turned out to be effectively zero, as high and low Agreeableness groups achieved similar outcomes, and Agreeableness levels had no impact on contributions.[24] Another possibility is that subjects were unable to identify Agreeableness questions, or that they otherwise find it an impossible task to misrepresent their personality. This seems unlikely, as McGee and McGee (2024) find that subjects can misrepresent themselves from a more difficult task of inferring the personality trait in question from a job description.

I also find that the Agreeableness group formation rule increases contributions in the *Agreeableness Before* and *Agreeableness After* treatments, but not in the *Agreeableness Never* treatment. In addition, the increase in contributions occurs for both H and L groups. Therefore, in conjunction with the result that Agreeableness does not correlate with individual contributions, the higher contributions in *Agreeableness Before* and *Agreeableness After* cannot be attributed to the

[23] See Abeler, Nosenzo, and Raymond (2019) for a meta-study on truth-telling preferences.

[24] However, it is not reasonable to suggest that subjects could have known either of these things.

formation of groups with higher Agreeableness. Instead, the operative factor must be the provision of information about the Agreeableness trait and the previous research on its relationship with contributions in the PGG.

The increase in contributions observed in *Agreeableness Before* and *Agreeableness After* could be driven by group identity. Being matched with those of a similar personality score could create a group identity. Cooperation and prosocial actions have been found to increase within such 'in-groups' (Eckel & Grossman, 2005; Y. Chen & Li, 2009). Alternatively, going through the personality test could be seen as a team-building exercise, which has been found to increase contributions in the PGG (Charness et al., 2014). Another explanation follows from the literature on self-image and prosocial behavior (e.g. Bénabou and Tirole (2006, 2011)). In this experiment, I tell subjects that high Agreeableness is associated with positive traits, and that those with high Agreeableness contribute more.[25] Subjects could contribute more in order to create or maintain a positive self-image that they possess the positive personality traits encapsulated by Agreeableness. Alternatively, these positive words could have primed subjects to be more cooperative. Drouvelis et al. (2015) found that priming via a word-search task can increase contributions in a PGG.

Lastly, I find that using personality tests in an unannounced way does not affect subsequent personality tests. Withholding information about the group formation rule did not influence later behavior, suggesting that experimental control was maintained. Therefore, 'unexpected data use' (Charness et al., 2022) remains a valid methodological tool for economics experiments when required by the study design.

My paper highlights the importance of pre-registration and replication. My design is built upon the previous findings in experimental economics that Agreeableness is related to individual contributions in a PGG, and prosocial behavior more generally. This relationship has a handful of conceptual replications, however, they were conducted before the necessity of pre-registration was widely known. In a pre-registered study, I fail to replicate any substantial relationship between Agreeableness and contributions in a PGG. In light of this finding, an alternative design would be more suitable from an ex-post perspective. We should exhibit caution when building upon previous findings, and move towards open science best-practices to enable future research to build upon our own findings with more confidence.

[25]The words used were: selfless, trusting, good-natured, generous, and forgiving.

References

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for Truth-Telling. *Econometrica*, *87*(4), 1115–1153. doi: 10.3982/ECTA14673

Ahn, T. K., Isaac, R. M., & Salmon, T. C. (2009, 2). Coming and going: Experiments on endogenous group sizes for excludable public goods. *Journal of Public Economics*, *93*(1-2), 336–351. doi: 10.1016/j.jpubeco.2008.06.007

Anderson, L. R., Mellor, J. M., & Milyo, J. (2004, 2). Social Capital and Contributions in a Public-Goods Experiment. *American Economic Review: Papers & Proceedings*, *94*(2), 373–376. doi: 10.1257/0002828041302082

Ashton, M., & Lee, K. (2009, 2). The HEXACO-60: A Short Measure of the Major Dimensions of Personality. *Journal of Personality Assessment*, *91*(4), 340–345. doi: 10.1080/00223890902935878

Bardsley, N., & Moffatt, P. G. (2007, 2). The experimetrics of public goods: Inferring motivations from contributions. *Theory and Decision*, *62*(2), 161–193. doi: 10.1007/s11238-006-9013-3

Bartling, B., Fehr, E., Maréchal, M. A., & Schunk, D. (2009, 2). Egalitarianism and Competitiveness. *American Economic Review*, *99*(2), 93–98. doi: 10.1257/aer.99.2.93

Bénabou, R., & Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, *96*(5).

Bénabou, R., & Tirole, J. (2011, 2). Identity, Morals, and Taboos: Beliefs as Assets. *The Quarterly Journal of Economics*, *126*(2), 805–855. doi: 10.1093/qje/qjr002

Bock, O., Baetge, I., & Nicklisch, A. (2014, 2). hroot: Hamburg Registration and Organization Online Tool. *European Economic Review*, *71*, 117–120. doi: 10.1016/j.euroecorev.2014.07.003

Burlando, R. M., & Guala, F. (2005, 2). Heterogeneous Agents in Public Goods Experiments. *Experimental Economics*, *8*(1), 35–54. doi: 10.1007/s10683-005-0436-4

Butera, L., Grossman, P. J., Houser, D., List, J. A., & Villeval, M. C. (2020). *A New Mechanism to Alleviate the Crises of Confidence in Science With An Application to the Public Goods Game.*

Cambridge Business English Dictionary. (2023, 2). *Psychometric Test - English Meaning - Cambridge Dictionary.* Retrieved from https://dictionary.cambridge.org/dictionary/english/psychometric-test

Carpenter, J. P., Daniere, A. G., & Takahashi, L. M. (2004, 2). Cooperation, trust, and social capital in Southeast Asian urban slums. *Journal of Economic Behavior & Organization*, *55*(4), 533–551. doi: 10.1016/j.jebo.2003.11.007

Cason, T. N., & Wu, S. Y. (2019, 2). Subject Pools and Deception in Agricultural and Resource Economics Experiments. *Environmental and Resource Economics*, *73*(3), 743–758. doi: 10.1007/s10640-018-0289-x

Catola, M., D'Alessandro, S., Guarnieri, P., & Pizziol, V. (2021, 2). Personal norms in the online public good game. *Economics Letters*, *207*, 110024. doi: 10.1016/j.econlet.2021.110024

Charness, G., Cobo-Reyes, R., & Jiménez, N. (2014, 2). Identities, selection, and contributions in a public-goods game. *Games and Economic Behavior*, *87*, 322–338. doi: 10.1016/j.geb.2014.05.002

Charness, G., Samek, A., & van de Ven, J. (2022, 2). What is considered deception in experimental economics? *Experimental Economics*, *25*(2), 385–412. doi: 10.1007/s10683-021-09726-7

Charness, G., & Yang, C.-L. (2014, 6). Starting small toward voluntary formation of efficient large groups in public goods provision. *Journal of Economic Behavior & Organization*, *102*, 119–132. doi: 10.1016/j.jebo.2014.03.005

Chaudhuri, A. (2011, 2). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, *14*(1), 47–83. Retrieved from http://link.springer.com/10.1007/s10683-010-9257-1 doi: 10.1007/s10683-010-9257-1

Chen, D. L., Schonger, M., & Wickens, C. (2016, 2). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97. Retrieved from https://www.sciencedirect.com/science/article/pii/S2214635016000101?via%3Dihub doi: 10.1016/J.JBEF.2015.12.001

Chen, Y., & Li, S. X. (2009, 2). Group identity and social preferences. *American Economic Review*, *99*(1), 431–457. doi: 10.1257/aer.99.1.431

Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. Elsevier. doi: 10.1016/C2013-0-10497-7

Cooper, D. J. (2014, 8). A Note on Deception in Economic Experiments. *Journal of Wine Economics*, *9*(2), 111–114. doi: 10.1017/jwe.2014.18

Donato, K., Miller, G., Mohanan, M., Truskinovsky, Y., & Vera-Hernández, M. (2017, 2). Person-

ality Traits and Performance Contracts: Evidence from a Field Experiment among Maternity Care Providers in India. *American Economic Review: Papers & Proceedings*, *107*(5), 506–510. doi: 10.1257/aer.p20171105

Drouvelis, M., Metcalfe, R., & Powdthavee, N. (2015, 11). Can priming cooperation increase public good contributions? *Theory and Decision*, *79*(3), 479–492. doi: 10.1007/s11238-015-9481-4

Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2011, 2). The framing of games and the psychology of play. *Games and Economic Behavior*, *73*(2), 459–478. doi: 10.1016/j.geb.2011.02.003

Dylman, A. S., & Zakrisson, I. (2023, 2). The effect of language and cultural context on the BIG-5 personality inventory in bilinguals. *Journal of Multilingual and Multicultural Development*, 1–14. doi: 10.1080/01434632.2023.2186414

Eckel, C. C., & Grossman, P. J. (2005, 2). Managing diversity by creating team identity. *Journal of Economic Behavior & Organization*, *58*(3), 371–392. doi: 10.1016/j.jebo.2004.01.003

Emergen Research. (2022, 2). *Assessment Services Market, By Product Type (Psychometric Test, Aptitude Tests, Coding Tests), By Service Type, By Medium (Online, Offline), By Sectors (K-12, Higher Education, Corporate, Government), and By Region Forecast to 2030* (Tech. Rep.).

Fehr, E., & Gächter, S. (2000, 2). Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, *90*(4), 980–994. Retrieved from http://pubs.aeaweb.org/doi/10.1257/aer.90.4.980 doi: 10.1257/aer.90.4.980

Fischbacher, U., & Föllmi-Heusi, F. (2013, 2). Lies in Disguise - An Experimental Study on Cheating. *Journal of the European Economic Association*, *11*(3), 525–547. doi: 10.1111/jeea.12014

Fréchette, G. R., Schotter, A., & Trevino, I. (2017, 2). Personality, Information Acquisition, and Choice Under Uncertainty: An Experimental Study. *Economic Inquiry*, *55*(3), 1468–1488. doi: 10.1111/ecin.12438

Gächter, S., & Thöni, C. (2005, 2). Social Learning and Voluntary Cooperation among like-Minded People. *Journal of the European Economic Association*, *3*(2-3), 303–314. doi: 10.1162/jeea.2005.3.2-3.303

Gill, D., & Prowse, V. (2016, 2). Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A Level-k Analysis. *Journal of Political Economy*, *124*(6), 1619–1676. doi:

10.1086/688849

Gill, D., & Rosokha, Y. (2024, 8). Beliefs, Learning, and Personality in the Indefinitely Repeated Prisoner's Dilemma. *American Economic Journal: Microeconomics*, *16*(3), 259–283. Retrieved from https://pubs.aeaweb.org/doi/10.1257/mic.20210336 doi: 10.1257/mic.20210336

Goldberg, L. (2002). *Big Five Factor Markers.* Retrieved from https://ipip.ori.org/newBigFive5broadKey.htm

Goldberg, L., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006, 2). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96. doi: 10.1016/j.jrp.2005.08.007

Greiner, B. (2015, 2). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125. Retrieved from http://link.springer.com/10.1007/s40881-015-0004-4 doi: 10.1007/s40881-015-0004-4

Gunnthorsdottir, A., Houser, D., & McCabe, K. (2007, 2). Disposition, history and contributions in public goods experiments. *Journal of Economic Behavior & Organization*, *62*(2), 304–315. doi: 10.1016/j.jebo.2005.03.008

Hare, R. D. (1985, 2). Comparison of procedures for the assessment of psychopathy. *Journal of Consulting and Clinical Psychology*, *53*(1), 7–16. doi: 10.1037/0022-006X.53.1.7

Hawkins, T., & Monroe, M. (2021, 2). *Persona: The Dark Truth Behind Personality Tests.* HBO Max. Retrieved from https://www.imdb.com/title/tt14173880/

Holmén, M., Holzmeister, F., Kirchler, M., Stefan, M., & Wengström, E. (2023, 10). Economic Preferences and Personality Traits Among Finance Professionals and the General Population. *The Economic Journal*, *133*(656), 2949–2977. doi: 10.1093/ej/uead038

Jonason, P. K., & Webster, G. D. (2010, 2). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment*, *22*(2), 420–432. doi: 10.1037/a0019265

Kagel, J., & McGee, P. (2014, 2). Personality and cooperation in finitely repeated prisoner's dilemma games. *Economics Letters*, *124*(2), 274–277. doi: 10.1016/j.econlet.2014.05.034

Kantrowitz, T. M., Tuzinski, K. A., & Raines, J. M. (2018). *2018 Global Assessment Trends Report.*

Küfner, A. C. P., Dufner, M., & Back, M. D. (2015, 2). Das Dreckige Dutzend und die

Niederträchtigen Neun. *Diagnostica*, *61*(2), 76–91. doi: 10.1026/0012-1924/a000124

Ledyard, J. (1995). Public Goods: A Survey of Experimental Evidence. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics* (pp. 111–194). Princeton University Press.

Lee, K., & Ashton, M. C. (2009). *The HEXACO Personality Inventory - Revised: Scale Definitions.* Retrieved from https://hexaco.org/scaledescriptions

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22 140*.

Lugovskyy, V., Puzzello, D., Sorensen, A., Walker, J., & Williams, A. (2017, 2). An experimental study of finitely and infinitely repeated linear public goods games. *Games and Economic Behavior*, *102*, 286–302. doi: 10.1016/j.geb.2017.01.004

McCrae, R. R., & John, O. P. (1992, 2). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, *60*(2), 175–215. doi: 10.1111/j.1467-6494.1992.tb00970.x

McGee, A., & McGee, P. (2024, 2). Whoever you want me to be: Personality and incentives. *Economic Inquiry*, *62*(3), 1268–1291. doi: 10.1111/ecin.13220

McGee, A., & McGee, P. (2025). Gender and race differences on incentivized personality measures. *Frontiers in Behavioral Economics*, *4*. doi: 10.3389/frbhe.2025.1499464

Ones, U., & Putterman, L. (2007, 2). The ecology of collective action: A public goods and sanctions experiment with controlled group formation. *Journal of Economic Behavior & Organization*, *62*(4), 495–521. doi: 10.1016/j.jebo.2005.04.018

Paulhus, D. L., & Williams, K. M. (2002, 2). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. doi: 10.1016/S0092-6566(02)00505-6

Perugini, M., Tan, J. H. W., & Zizzo, D. J. (2010). Which is the More Predictable Gender? Public Good Contribution and Personality . *Economic Issues*, *15*(1), 83–110.

Proto, E., Rustichini, A., & Sofianos, A. (2019, 2). Intelligence, Personality, and Gains from Cooperation in Repeated Interactions. *Journal of Political Economy*, *127*(3), 1351–1390. doi: 10.1086/701355

Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2020, 2). Validation of the Short and Extra-Short Forms of the Big Five Inventory-2 (BFI-2) and Their German Adaptations. *European Journal of Psychological Assessment*, *36*(1), 149–161. doi: 10.1027/1015-5759/a000481

Raskin, R. N., & Hall, C. S. (1979, 2). A Narcissistic Personality Inventory. *Psychological Reports*,

45(2), 590. doi: 10.2466/pr0.1979.45.2.590

Soto, C. J., & John, O. P. (2017, 2). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, *68*, 69–81. doi: 10.1016/j.jrp.2017.02.004

Streib, H., & Wiedmaier, M. (2001). *German Translation of the 100-Item Lexical Big-Five Factor Markers.* Retrieved from https://ipip.ori.org/German100-ItemBig-FiveFactorMarkers.htm

Thielmann, I., Spadaro, G., & Balliet, D. (2020, 2). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, *146*(1), 30–90. doi: 10.1037/bul0000217

Villeval, M. C. (2016). Can lab experiments help design personnel policies? *IZA World of Labor*. doi: 10.15185/izawol.318

Villeval, M. C. (2020). Public goods, norms and cooperation. In C. M. Capra, R. Croson, M. Rigdon, & T. Rosenblat (Eds.), *Handbook of experimental game theory.* Edward Elgar Publishing.

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

Weber, L., & Dwoskin, E. (2014, 2). *Are Workplace Personality Tests Fair?* Retrieved from https://www.wsj.com/articles/are-workplace-personality-tests-fair-1412044257

A Deviations from the Pre-registration Document

The text of this paper differs from the original pre-registration document, which was presented as a mostly complete paper. These changes are mostly as you would expect, such as adapting the text to the results, fixing errors, incorporating suggestions, reducing the length of the manuscript, and adding citations as they were brought to my attention. Throughout this process I have attempted to minimize differences and highlight substantial things that were not pre-registered, and I also summarize major deviations in this section. However, this may not be perfect given the sheer number of changes, so to supplement this I also include a pdf that highlights all differences between the current version and the pre-registered version in the OSF project http://doi.org/10.17605/OSF.IO/MDB7A, and the reader is encouraged to use online tools like Draftable to make this comparison even easier for them to follow.

A.1 Introduction

The second paragraph of the introduction that explained how using personality testing for job hiring may have unintended consequences due to misrepresentation by job seekers was removed to reduce the length, and as this concept should be relatively straightforward for the intended audience (particularly economists) to grasp. Group identity, self-image and priming were not part of the original Introduction, as they are proposed to explain the initially unexpected finding that Agreeableness group formation was effective but not through the creation of H groups. Some discussion about personality testing and hiring was removed in the interests of space, and throughout the paper adjustments are made to make it more clear that the experiment is only intended to be an analogy of the hiring process.

A.1.1 Contribution to the Literature

I removed the paragraph on related literature on psychometric personality testing to reduce the length of the manuscript. The paragraph discussed the importance of incentives and contrasted this against psychology research, but such a general statement about that literature is inaccurate. The paragraph did not really add much to this section and so it was removed. I also emphasize the methodological contributions of the paper more by moving some discussion and explanation up from later in the paper into this section, and adjust accordingly.

A.2 Experimental Design

Treatment names were changed from being more abstract (e.g. A_0t_1 became *Agreeableness After*) here and in the rest of the paper in an attempt to improve readability. The message about the information that subjects received on Agreeableness was also included. Some more information is given about the specific Procedures, which was always able to be inferred from what was given in the pre-registration but is just made more readily available. All of these changes were based on suggestions from anonymous referees. In addition, to save space I moved some of the finer details about the Part 1 personality questions into the relevant appendix.

A.3 Results

Table 5 was added, while a table that summarized overall average contributions by treatment was removed. Reminders of what each Hypothesis stated were added. Table A3 and extended discussion about unsophisticated misrepresentation was moved to the Appendix in order to save space. Table 7 was brought up from the Appendix as it fit the narrative of the text. Discussion about the importance of pre-registration was reduced.

A.4 Appendix

The order of the appendices were changed in an attempt to reflect their relative importance. The regression specification in Appendix B.1 had an additional *Agreeableness Before* dummy added. This regression was a conceptual replication and was changed upon the advice of one of the authors of that original study. Appendix D was added to include the output of all regressions. Testing misrepresentation at the individual question level is reported in Appendix B.3.2. A test on whether the group formation rule was effective in creating High and Low Agreeableness groups was included in Appendix B.4. Summary statistics of variables not reported in Table 5 were added in Table A5.

A.5 Others

The random ordering of treatments was not able to be followed perfectly near the end of the data collection, as the number of subjects that showed up could differ from what the next treatment required. For example, 18 subjects might show up for a session that only requires 12 more observations. In this case, the next treatment that required 18 subjects was conducted in that session,

while the original assigned treatment was conducted in the next session that had 12 subjects. The treatments that were assigned to VCEE were done so as soon as the decision to employ them was made, but show-up rates also necessitated some changes. Given the large majority of sessions followed the random ordering, then the treatments that were in these later sessions were also random. Based on this, there is no reasonable threat to the randomization procedure.

B Additional Analysis and Robustness Checks

B.1 Conceptual Replication Results

The experimental design permits a conceptual replication of some elements of McGee and McGee (2024). In particular, Research Questions 1 and 2 from that paper can be partially answered.

MM Research Question 1: *How important are incentives when measuring personality?*

The incentives in McGee and McGee (2024) were a direct lump-sum payment if selected for a job. Whereas in the current study the incentive is indirect, as it is membership in the more cooperative H group that could increase earnings in the PGG. Research Question 1 is addressed by Hypothesis 1 and the comparisons in Figure 2. As there is no evidence for Hypothesis 1, I conclude that indirect incentives are not very important when measuring personality. The strength of the incentives appear to matter for misrepresentation in personality tests.

MM Research Question 2: *Are incentivized personality measures influenced by traits other than personality?*

McGee and McGee (2024) posit that traits such as intelligence, Machiavellianism, self-deception, optimism, acceptability of lying, risk aversion, and locus of control could be correlated with misrepresentation. They find that most of these characteristics are uncorrelated with misrepresentation in all treatments of their experiment.

In particular, McGee and McGee (2024) find no evidence that Machiavellianism is correlated with misrepresentation in any of their treatments. This is an interesting result, given that people high in Machiavellianism tend to be manipulative and strategically self-serving in their words and actions. In this paper, Part 3 elicits Machiavellianism using a different set of questions, and its relationship to misrepresentation of Agreeableness in Part 1 can be explored. I test this relationship with a Tobit regression censored at 26 and 130 of the following form $Agreeableness_i = \beta_0 + \beta_1 Machiavellianism + \beta_2 Machiavellianism \times AgreeablenessBefore + \beta_3 AgreeablenessBefore + \epsilon_i$. The coefficient β_1 represents the correlation between Agreeableness and Machiavellianism, and β_2 represents the increase (if > 0) in reported Agreeableness when there is an incentive to misrepresent (i.e. in *Agreeableness Before*).

Table A1 summarizes and shows that Agreeableness is increasing in Honesty Humility and decreasing in the Dark Triad traits, as expected. However, I find no evidence that any of these personality traits affect misrepresentation of Agreeableness, in line with the findings of McGee and

McGee (2024).

Table A1: Personality Traits and Agreeableness Misrepresentation

Trait	β_1	β_2	β_3
Honesty Humility	0.45**	0.03	1.82
Machiavellianism	-1.03***	0.44	-1.29
Narcissism	-0.26	0.42	-2.35
Psychopathy	-1.92***	0.52	-2.25

β_1 represents the correlation between the trait and Agreeableness, and β_2 represents the correlation between the trait and misrepresentation of Agreeableness. β_3 is the coefficient of the *Agreeableness Before* dummy. ***=$p < 0.01$, **=$p < 0.05$, and *=$p < 0.10$. Full regression output is presented in Appendix D.1.3.

B.2 Agreeableness and PGG Contributions

Table A2 presents various robustness checks on the Agreeableness coefficient in Table 7. Each robustness check follows the same analysis as in Table 7, except in the manner specified.

Robustness Check	Coefficient (Std. err.)
All Data	0.01 (0.04)
No *Agreeableness Before* or *Agreeableness After*	-0.02 (0.05)
No other Ind. Charact.	0.02 (0.04)
No Tobit	0.00 (0.02)
Only Agreeableness	0.01 (0.04)
Only Men	-0.04 (0.08)
Period ≤ 2	-0.01 (0.07)
Period ≤ 5	-0.01 (0.06)

Table A2: Robustness checks on Agreeableness and Contributions

An individual's contribution to the public good is the dependent variable. Only the relevant coefficient of Agreeableness as an independent variable is displayed. Full regression output is reported in Appendix D.2.3. ***=$p < 0.01$, **=$p < 0.05$, and *=$p < 0.10$.

'All Data' includes *Agreeableness Before* observations. 'No *Agreeableness Before* or *Agreeableness After*' drops *Agreeableness After* observations. 'No other Ind. Charact.' drops all other

personality and demographic measures as independent variables (but retains treatment and period controls). 'No Tobit' does not control for censoring (i.e. it uses xtmixed instead of xttobit). 'Only Agreeableness' has Agreeableness as the only independent variable (i.e. it drops other personality traits, demographics, and treatment controls). 'Only Men' excludes those who do not identify as men, while 'Period $\leq x$' excludes all periods greater than x.

B.3 Unsophisticated and Question-level Misrepresentation

B.3.1 Unsophisticated Misrepresentation

In pre-registered analysis, I consider whether misrepresentation is sophisticated or not. If misrepresentation is sophisticated, then subjects only misrepresent the relevant trait of Agreeableness. However, if misrepresentation is unsophisticated, then responses could also change for other Big Five characteristics. Table A3 presents the analysis of the Big Five personality traits, indicating that there are no differences in most personality traits. There is some evidence for suspicion from subjects in the $After$ & $Never$ treatments, as they report different levels of 'Open Mindedness' and 'Extraversion' than subjects in the *Agreeableness Before* and *Random Before* treatments respectively.

Table A3: All Big Five characteristics by Treatment

Characteristic	*Random Before*	*Agreeableness Before*	*After & Never*	p-values
Agreeableness	105.02	107.42	105.04	0.37, 0.81, 0.10
Open Mindedness	21.69	22.80	21.70	0.21, 0.72, 0.02
Negative Emotionality	16.48	15.64	15.70	0.54, 0.52, 0.90
Extroversion	19.44	20.05	20.91	0.30, 0.01, 0.12
Conscientiousness	22.25	21.96	21.25	0.59, 0.14, 0.29

The treatment columns report the average score of the given personality trait. Agreeableness $\in [26, 130]$ and all other personality traits $\in [6, 30]$. The p-values column reports the results from Mann-Whitney tests on the pairs: *Random Before* to *Agreeableness Before*; *Random Before* to $t > 0$; and *Agreeableness Before* to $t > 0$ respectively.

B.3.2 Question-level Misrepresentation

In exploratory analysis, I consider whether there is misrepresentation at the individual question level. Some questions may be more easily identified as being related to Agreeableness, or to cooperation in the PGG.[26] I report this analysis in Table A4 for all comparisons that had $p < 0.05$

[26] I thank an anonymous referee for this suggestion.

Question	Treatment Comparison	Big 5 Trait	Avg. Resp. G1	Avg. Resp. G2	p-value
I have few artistic interests	$A0$ **vs** $t > 0$	**Open-Mindedness**	**2.47**	**3.06**	**0.00**
I am fascinated by art, music, or literature	$A0$ vs $t > 0$	Open-Mindedness	4.03	3.66	0.00
I am dominant and act as a leader	$A0$ vs $t > 0$	Extraversion	3.18	3.52	0.00
I sympathize with other's feelings	$A0$ vs $t > 0$	Agreeableness	4.50	4.25	0.01
I show my gratitude	$A0$ vs $t > 0$	Agreeableness	4.52	4.30	0.01
I assume the best about people	$R0$ vs $A0$	Agreeableness	3.44	3.90	0.02
I tend to find fault with others	$R0$ vs $A0$	Agreeableness	3.00	2.56	0.02
I am fascinated by art, music, or literature	$R0$ vs $A0$	Open-Mindedness	3.67	4.03	0.02
I am full of energy	$R0$ vs $t > 0$	Extraversion	3.52	3.90	0.03
I think of others first	$A0$ vs $t > 0$	Agreeableness	3.46	3.21	0.03
I tend to be disorganized	$A0$ vs $t > 0$	Conscientiousness	2.44	2.75	0.04
I am sometimes rude to others	$A0$ vs $t > 0$	Agreeableness	2.05	2.33	0.04
I am full of energy	$A0$ vs $t > 0$	Extraversion	3.68	3.90	0.05
I am hard to get to know	$A0$ vs $t > 0$	Agreeableness	2.38	2.68	0.05
I am indifferent to the feelings of others	$R0$ vs $t > 0$	Agreeableness	1.81	2.01	0.05
I am outgoing and sociable	$R0$ vs $t > 0$	Extraversion	3.42	3.75	0.05

Table A4: Question-level Misrepresentation when $p < 0.05$

The p-values are from a Mann-Whitney test. Avg. Resp. G1 is the average response to the 5-point Likert scale question for how much they agree the particular statement describes them for the first treatment group listed in the Treatment Comparison column, similarly for Avg. Resp. G2. Rows in bold survive the Bonferroni-Holm correction for the 150 comparisons that were conducted in this exercise (50 questions $\times$ 3 comparison groups). $A0$ = *Agreeableness Before*, $R0$ = *Random Before*, $t > 0$ = *After & Never*.

before any correction for multiple comparisons. Table A4 shows some misrepresentation, as those in the *Agreeableness Before* treatment generally tend to report more socially desirable responses for these questions. This misrepresentation is sometimes unsophisticated, as not all questions are of the target personality trait of Agreeableness.

B.4 Effectiveness of Agreeableness Group Formation Rule

Given that high and low Agreeableness groups were formed from a silo of 6 subjects, a natural question that arises is how effective the group formation rule is in creating true high or low Agreeableness groups.[27] To that end, Figures 5 and 6 present boxplots of the distribution of Agreeableness of those assigned to high, low, or random group types for individuals and groups respectively. These figures show statistically significant separation in reported Agreeableness between group types. Furthermore, 1,000,000 simulations that randomly pair a high group to a low

[27] I thank an anonymous referee for this suggestion.

group find that the high group has a higher average Agreeableness 94.7% of the time. All this combined suggests that despite high and low Agreeableness groups being formed from a relatively small silo, the group formation rule does create groups of reasonably different levels of Agreeableness.

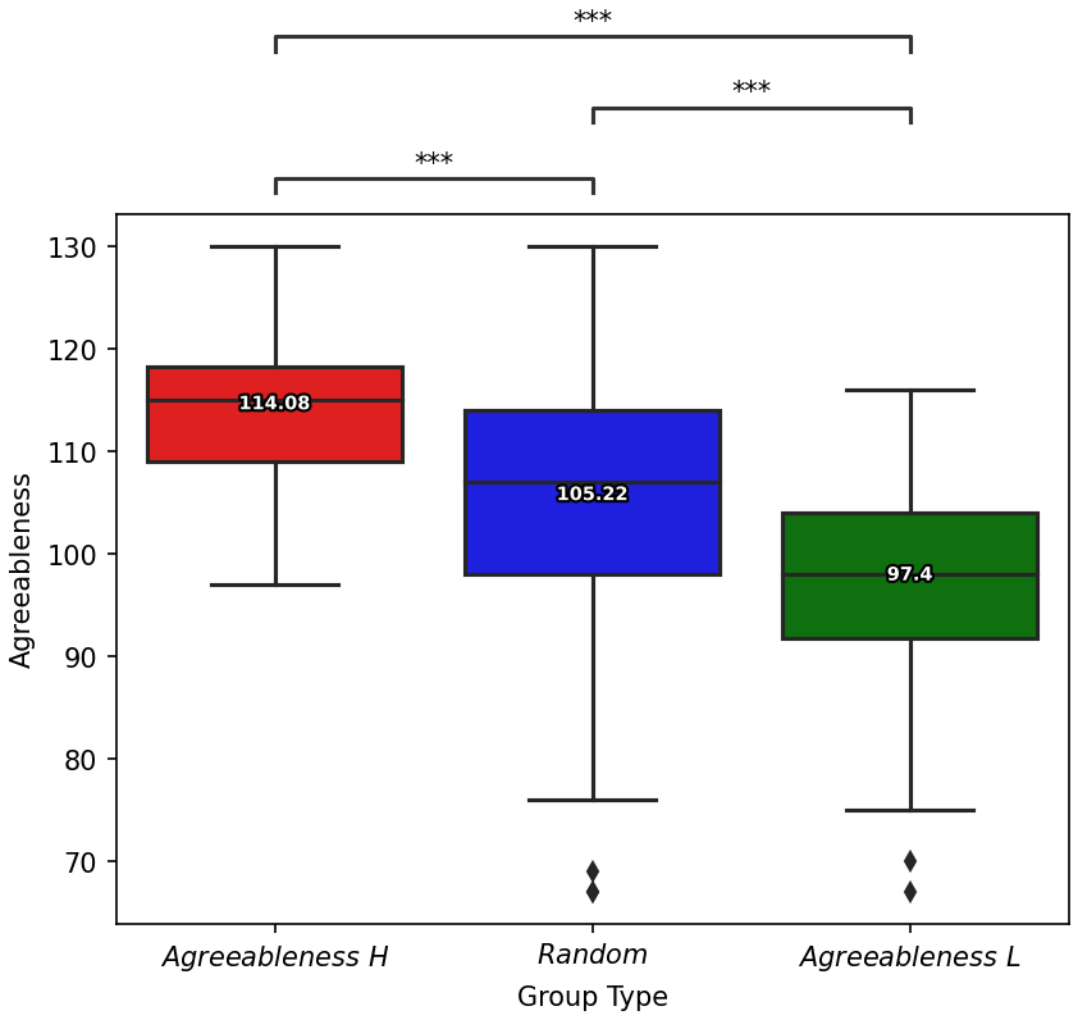


Figure 5: Agreeableness of each Individual Assigned to a Specific Group Type

Mean Agreeableness overlaid. Statistical results are based on a Mann Whitney test. The three lines in the box are the 75%, 50% and 25% quartile when going from top to bottom, the top (bottom) whisker is the largest (smallest) value that is below (above) 1.5 times the difference between the 75% and 25% quartiles, and values outside this range are presented as diamonds. ***=$p < 0.01$, **=$p < 0.05$, *=$p < 0.10$, and n.s.=$p \geq 0.10$.

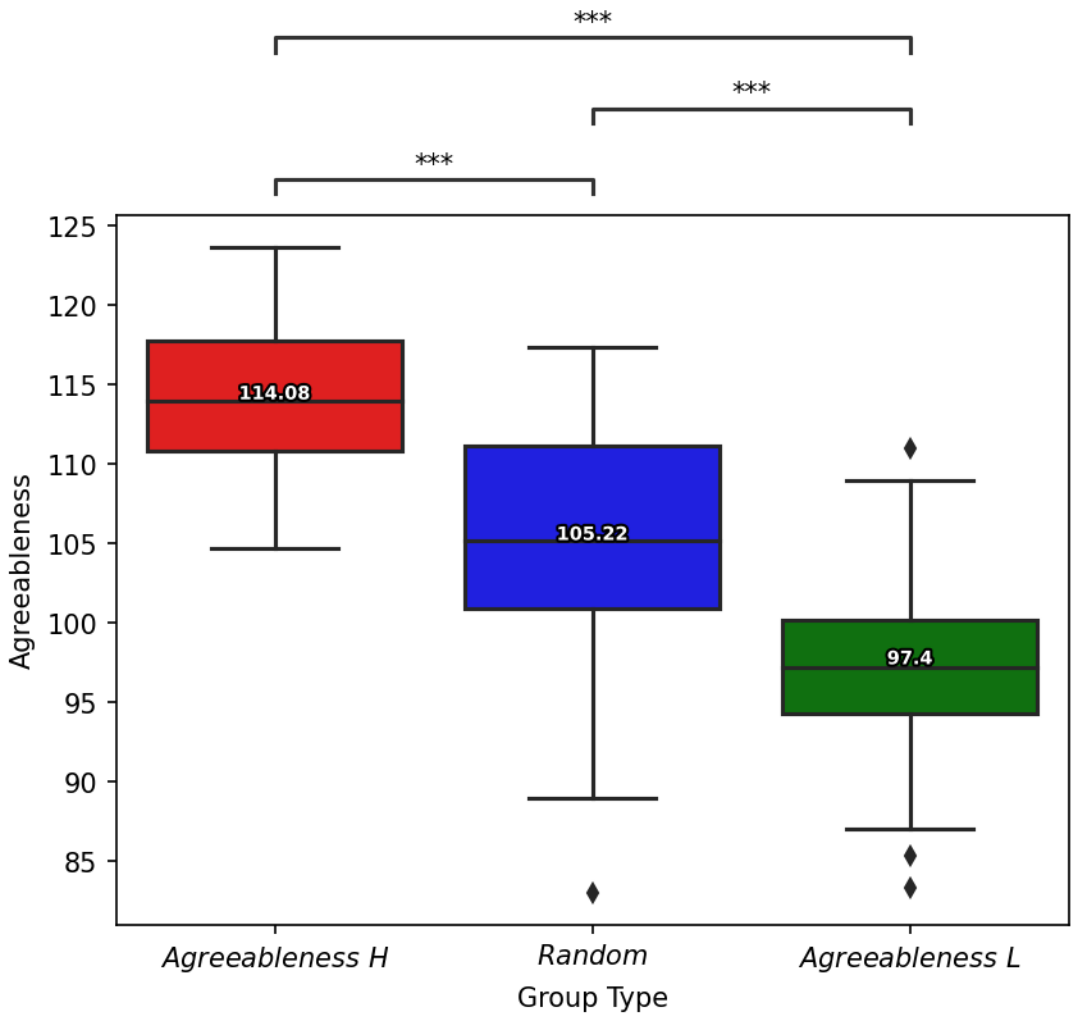


Figure 6: Average Agreeableness of each Group Assigned to a Specific Group Type

Mean Agreeableness overlaid. Statistical results are based on a Mann Whitney test. The three lines in the box are the 75%, 50% and 25% quartile when going from top to bottom, the top (bottom) whisker is the largest (smallest) value that is below (above) 1.5 times the difference between the 75% and 25% quartiles, and values outside this range are presented as diamonds. ***=$p < 0.01$, **=$p < 0.05$, *=$p < 0.10$, and n.s.=$p \geq 0.10$.

B.5 Summary Statistics

Table A5 Notes: Standard deviations in parentheses. Low and High Agreeableness groups reported separately for *Agreeableness* treatments. Categorical variables are presented as is for summary purposes only and are otherwise unadjusted unless specifically noted. PGG Total Group Contribution is reported at the group per round level. All personality traits $\in [4, 20]$. Female = 1 if reported gender was female, 0 if reported male, and not included otherwise (other data omitted for this specific statistic only). Age is reported age in years. Num. Prev. Exp. is the number of previous experiments the subject reported they had participated in. Year At Uni.: 1 = First year, 2 = Second year, 3 = Third year, 4 = Fourth year +, 5 = Graduate student. GPA: 1 = [1, 1.5], 2 = [1.51,2.5], 3 = [2.51, 3.5], 4 = [3.51, ∞), 5 = N/A. Enjoyment: 1 = Disliked the experiment a lot, 2 = Disliked ... a little, 3 = Did not like or dislike, 4 = Liked ... a little, 5 = Liked ... a lot. Num. Fail CompQ 1 is the number of times that answers with at least one error or missing value was submitted over all questions on the first page of comprehension questions, similarly for Num. Fail CompQ 2 but for the second page. Abbreviations: *Rand.* = *Random*, *Agre.* = *Agreeableness*, Tot. = Total, Contr. = Contributions, Num. = Number, Prev. = Previous, Exp. = Experiments, Uni. = University, GPA = Grade Point Average, CompQ = Comprehension Questions.

	Rand. Before	*Rand. After*	*Rand. Never*	*Agre. Before*	*Agre. After*	*Agre. Never*
PGG Tot. Group Contr. (Low Agre. High Agre.)	20.72 (7.21)	23.58 (10.30)	18.27 (8.85)	34.41 (7.71) 31.19 (1.25)	28.06 (10.74) 28.74 (5.33)	22.98 (3.61) 18.96 (5.03)
Honesty Humility	12.60 (3.40)	12.06 (3.19)	12.15 (2.76)	12.54 (3.24) 13.46 (3.09)	12.54 (3.09) 13.00 (3.02)	12.85 (3.40) 12.65 (3.46)
Machiavellianism	8.04 (3.24)	8.96 (3.94)	8.21 (3.11)	8.69 (3.56) 7.56 (3.58)	8.98 (3.34) 6.92 (3.04)	8.73 (3.85) 8.08 (3.19)
Psychopathy	8.42 (2.94)	8.04 (3.57)	7.79 (3.07)	8.48 (2.67) 7.19 (2.75)	9.27 (3.33) 6.75 (2.28)	9.46 (3.24) 7.35 (2.65)
Narcissism	11.73 (3.46)	11.85 (3.43)	10.92 (4.08)	11.58 (3.46) 11.25 (3.66)	11.27 (3.39) 10.94 (3.54)	11.60 (3.70) 12.21 (3.62)
Female	0.60 (0.50)	0.70 (0.46)	0.54 (0.50)	0.46 (0.50) 0.70 (0.46)	0.44 (0.50) 0.64 (0.49)	0.44 (0.50) 0.67 (0.48)
Age	22.88 (3.42)	22.44 (2.58)	22.38 (2.73)	23.98 (4.61) 24.15 (4.42)	24.67 (9.89) 22.65 (2.91)	23.33 (4.11) 22.27 (2.52)
Num. Prev. Exp.	5.35 (6.31)	5.88 (8.05)	5.38 (4.51)	6.52 (7.16) 6.38 (6.86)	5.73 (7.37) 3.94 (4.33)	7.04 (7.40) 5.54 (5.32)
Year At Uni.	2.98 (1.47)	2.77 (1.28)	2.81 (1.33)	2.96 (1.35) 2.79 (1.27)	2.90 (1.53) 2.62 (1.39)	3.15 (1.38) 2.88 (1.44)
GPA	2.56 (1.32)	2.71 (1.25)	2.79 (1.35)	2.79 (1.34) 2.48 (1.37)	2.88 (1.45) 2.94 (1.62)	2.62 (1.30) 2.69 (1.49)
Enjoyment	3.81 (0.91)	3.73 (0.94)	3.73 (1.07)	3.90 (0.88) 3.77 (0.93)	3.77 (1.06) 3.90 (0.86)	3.71 (1.09) 3.98 (0.89)
Num. Fail CompQ 1	1.40 (3.04)	0.73 (1.09)	1.44 (3.21)	2.02 (4.39) 1.00 (2.13)	0.90 (1.57) 1.31 (2.27)	1.04 (1.70) 1.35 (2.35)
Num. Fail CompQ 2	1.65 (3.35)	1.50 (1.83)	1.56 (2.52)	1.40 (2.22) 0.98 (1.79)	1.38 (2.10) 1.06 (1.29)	1.60 (2.35) 1.25 (1.59)

Table A5: Summary Statistics

C Personality Questions

Part 1 consists of 50 questions designed to elicit the Big Five personality traits (McCrae & John, 1992). These traits are Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. As the experiment is conducted in Austria, the questions (and the rest of the experiment)

are in German. Each Big Five characteristic is elicited using the 30 question 'BFI-2-S Inventory' (Soto & John, 2017; Rammstedt, Danner, Soto, & John, 2020). The remaining 20 questions are all on Agreeableness, and sourced from the International Personality Item Pool's (IPIP) '100-Item Lexical Big-Five Factor Markers' (Goldberg, 2002; Goldberg et al., 2006; Streib & Wiedmaier, 2001). The Agreeableness trait is disproportionately weighted (26/50) as it is of primary interest and used for group formation in Part 2 in the *Agreeableness* treatments. Agreeableness is calculated based on each subject's numerical responses on the relevant questions. The 16 questions in Part 3 are taken from the 'Dirty Dozen' (Jonason & Webster, 2010) and four Honesty-Humility questions from HEXACO's 60-item version (Lee & Ashton, 2009). Subjects are asked how much they agree each statement applies to them using a 5-point Likert scale (Likert, 1932). The 5 points are labeled: 1 = Disagree strongly, 2 = Disagree a little, 3 = Neither agree nor disagree, 4 = Agree a little, and 5 = Agree strongly. They are presented using horizontal radio buttons. Subjects face blocks of five questions on a page, and all questions are presented in a random order that differs across subjects.[28] Personality traits are scored based on each subject's numerical (i.e. 1-5) responses by the following formula: $Trait = \sum_{i \in Q}(LikertValue_{+veKey} + (6 - LikertValue_{-veKey}))$, where Q is the set of relevant questions to that trait. Appendices C.1 and C.2 report which questions are related to each trait and whether the questions are positively or negatively keyed.

As mentioned previously, the experiment is conducted in German. While the majority of the university students that make up the subject pool are fluent in English, it is important to conduct personality tests in their native language. Firstly, there will be heterogeneity in subjects' confidence or ability in using English. Secondly, there is a literature that suggests that elicited personality traits are different in bilingual speakers depending on what language is being used (see Dylman and Zakrisson (2023) for examples). I would rather observe a subject's 'regular' personality rather than one that is shaped by a foreign language. Strategic misrepresentation of personality is already likely to be difficult enough as it is, let alone with an additional levels of complexity on top of that. The question sets used in Parts 1 and 3 all have pre-existing German translations. Rammstedt et al. (2020) translate the BFI-2-S. Streib and Wiedmaier (2001) translate the 100-Item IPIP. Küfner, Dufner, and Back (2015) translate the Dirty Dozen. A translation for HEXACO is provided by Lee and Ashton (2009). A list of the questions and their translations are provided in Appendices C.1 and C.2.

[28] Technically it is possible that two subjects face exactly the same ordering, however this is unlikely as the probability of that occurring in Part 1 is $p = \frac{1}{50!}$ and in Part 3 $p = \frac{1}{16!}$.

Some questions were changed or removed. In Part 1, a question was changed slightly to avoid excessive repetition, from *'I am compassionate and soft-hearted'* to *'I am compassionate'*, as another question is *'I have a soft heart'*. Two less relevant questions from Honesty-Humility were removed in order to maintain an equal number of questions between each trait in Part 3. The removed questions are *'I'd be tempted to use counterfeit money, if I were sure I could get away with it.'* and *'If I knew that I could never get caught, I would be willing to steal a million dollars.'* It was brought to my attention that I had not included the Greed-Avoidance or Modesty facets of Honesty-Humility. I do not know why this was the case, and I attribute it to human error. I have now explicitly specified that it is the Sincerity and Fairness facets of Honesty-Humility that are asked in the main body of the text.

Some of the pre-existing translations were changed based on feedback from native German speakers. The question *'I have a soft heart'* was changed from *'Ich habe ein weiches Herz'* to *'Ich bin gutherzig'*. The question *'I have a good word for everyone'* was changed from *'Ich habe ein gutes Wort für jeden'* to *'Ich rede gut über andere'*. These two changes were implemented as the original translations were considered ambiguous and a little too literal. The question *'I make people feel at ease'* was changed from *'Ich mache andere Leute ungezwungen'* to *'Ich kann andere beruhigen'*. Ungezwungen can be interpreted as being unhinged rather than calm, and may also be grammatically incorrect. The question *'Ich habe getäuscht oder gelogen, um meinen Willen durchzusetzen'* was changed to *'Ich neige dazu, zu täuschen oder zu lügen, um meinen Willen durchzusetzen'*, and similarly the question *'Ich habe Schmeicheleien genutzt, um meinen Willen durchzusetzen'* to *'Ich neige dazu, Schmeicheleien zu benutzen, um meinen Willen durchzusetzen'*. The other questions in the Dirty Dozen all have *'Ich neige dazu'* (I have the tendency to), and I was concerned that the question about lying could be interpreted as whether they have been deceitful in the current experiment, rather than a tendency in general.

C.1 Part 1 Questions and Translations

English Questions	German Questions
Agreeableness Positively Keyed	
I am interested in people.	Ich interessiere mich für Leute.
I sympathize with other's feelings.	Ich kann die Gefühle anderer nachempfinden.
I have a soft heart.	Ich bin gutherzig.
I take time out for others.	Ich nehme mir Zeit für andere.
I feel other's emotions	Ich kann die Gefühle anderer nachfühlen.
I make people feel at ease.	Ich mache andere Leute ungezwungen.
I inquire about other's well-being.	Ich erkundige mich nach dem Wohlbefinden anderer.
I know how to comfort others.	Ich weiß wie ich andere trösten kann.
I love children.	Ich liebe Kinder.
I am on good terms with nearly everyone.	Ich komme mit fast jedem gut aus.
I have a good word for everyone.	Ich rede gut über andere.
I show my gratitude.	Ich zeige meine Dankbarkeit.
I think of others first.	Ich denke zuerst an andere.
I love to help others.	Ich liebe es anderen zu helfen.
I am compassionate.	Ich bin einfühlsam.
I assume the best about people.	Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.
I am respectful and treat others with respect.	Ich begegne anderen mit Respekt.
Agreeableness Negatively Keyed	
I insult people.	Ich beleidige Leute.
I am not interested in other people's problems.	Ich interessiere mich nicht für die Probleme anderer Leute.
I feel little concern for others.	Andere Menschen kümmern mich wenig.
I am not really interested in others.	Ich interessiere mich nicht wirklich für andere.
I am hard to get to know.	Mich kennenzulernen ist schwer.
I am indifferent to the feelings of others.	Ich bin den Gefühlen anderer gegenüber gleichgültig.
I am sometimes rude to others.	Ich bin manchmal unhöflich und schroff.
I can be cold and uncaring.	Andere sind mir eher gleichgültig, egal.
I tend to find fault with others.	Ich neige dazu, andere zu kritisieren.
Extraversion Positively Keyed	
I am dominant and act as a leader.	Ich neige dazu, die Führung zu übernehmen.
I am full of energy.	Ich bin voller Energie und Tatendrang.
I am outgoing and sociable.	Ich gehe aus mir heraus, bin gesellig.
Extraversion Negatively Keyed	
I tend to be quiet.	Ich bin eher ruhig.
I prefer to have others take charge.	In einer Gruppe überlasse ich lieber anderen die Entscheidung.
I am less active than other people.	Ich bin weniger aktiv und unternehmungslustig als andere.

Table A6: Part 1 Questions 1-32

English Questions	German Questions
Conscientiousness Positively Keyed	
I am reliable and can always be counted on.	Ich bin verlässlich, auf mich kann man zählen.
I keep things neat and tidy.	Ich mag es sauber und aufgeräumt.
I am persistent and work until a task is finished.	Ich bleibe an einer Aufgabe dran, bis sie erledigt ist.
Conscientiousness Negatively Keyed	
I tend to be disorganized.	Ich bin eher unordentlich.
I have difficulty getting started on tasks.	Ich neige dazu, Aufgaben vor mir herzuschieben.
I can be somewhat careless.	Ich bin manchmal ziemlich nachlässig.
Negative Emotionality Positively Keyed	
I worry a lot.	Ich mache mir oft Sorgen.
I tend to feel depressed and blue.	Ich bin oft deprimiert, niedergeschlagen.
I am temperamental and get emotional easily.	Ich reagiere schnell gereizt oder genervt.
Negative Emotionality Negatively Keyed	
I am emotionally stable and not easily upset.	Ich bin ausgeglichen, nicht leicht aus der Ruhe zu bringen.
I am relaxed and handle stress well.	Ich bleibe auch in stressigen Situationen gelassen.
I feel secure and comfortable with myself.	Ich bin selbstsicher, mit mir zufrieden.
Open-mindedness Positively Keyed	
I am fascinated by art, music, or literature.	Ich kann mich für Kunst, Musik und Literatur begeistern.
I am original and come up with new ideas.	Ich bin originell, entwickle neue Ideen.
I am complex and a deep thinker.	Es macht mir Spaß, gründlich über komplexe Dinge nachzudenken und sie zu verstehen.
Open-mindedness Negatively Keyed	
I have little interest in abstract ideas.	Mich interessieren abstrakte Überlegungen wenig.
I have few artistic interests.	Ich bin nicht sonderlich kunstinteressiert.
I have little creativity.	Ich bin nicht besonders einfallsreich.

Table A7: Part 2 Questions 33-50

C.2 Part 3 Questions and Translations

English Questions	**German Questions**
Narcissism Positively Keyed	
I tend to want others to admire me.	Ich neige dazu, von anderen bewundert werden zu wollen.
I tend to want others to pay attention to me.	Ich neige dazu, von anderen beachtet werden zu wollen.
I tend to expect special favors from others.	Ich neige dazu, besondere Gefälligkeiten von anderen zu erwarten.
I tend to seek prestige or status.	Ich neige dazu, nach Ansehen oder Status zu streben.
Psychopathy Positively Keyed	
I tend to lack remorse.	Ich neige dazu, keine Gewissensbisse zu haben.
I tend to be callous or insensitive.	Ich neige dazu, gefühllos oder unsensibel zu sein
I tend to not be too concerned with morality or the morality of my actions.	Ich neige dazu, mich nicht um die Moral meiner Handlungen zu kümmern.
I tend to be cynical.	Ich neige dazu, zynisch zu sein.
Machiavellianism Positively Keyed	
I have used deceit or lied to get my way.	Ich neige dazu, zu täuschen oder zu lügen, um meinen Willen durchzusetzen.
I tend to manipulate others to get my way.	Ich neige dazu, andere zu manipulieren, um meinen Willen durchzusetzen.
I have used flattery to get my way.	Ich neige dazu, Schmeicheleien zu benutzen, um meinen Willen durchzusetzen.
I tend to exploit others towards my own end.	Ich neige dazu, andere für meine Zwecke auszunutzen.
Honesty-Humility Positively Keyed	
I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.	Ich würde keine Schmeicheleien benutzen, um eine Gehaltserhöhung zu bekommen oder befördert zu werden, auch wenn ich wüsste, dass es erfolgreich wäre.
I wouldn't pretend to like someone just to get that person to do favors for me.	Ich würde nicht vortäuschen, jemanden zu mögen, nur um diese Person dazu zu bringen, mir Gefälligkeiten zu erweisen.
I would never accept a bribe, even if it were very large.	Ich würde niemals Bestechungsgeld annehmen, auch wenn es sehr viel wäre.
Honesty-Humility Negatively Keyed	
If I want something from someone, I will laugh at that person's worst jokes.	Wenn ich von jemandem etwas will, lache ich auch noch über dessen schlechteste Witze.

Table A8: Part 3 Questions

D Full Regression Output

D.1 Pre-registered Analysis

D.1.1 Table 6

To test for the statistical significance of treatment effects, I use a panel data approach with random effects at the group level that accounts for the possibility of decreasing contributions over time. I use the group's average contribution in a period as the dependent variable, and a treatment dummy and the period as the independent variables. That is, the regression is of the form $AverageGroupContribution_{g,t} = \beta_0 + \beta_1 Treatment_g + \beta_2 t + v_g + \epsilon_{g,t}$. I use a Tobit regression to account for the censoring that can occur at 0 and 25 tokens for lower and upper limits respectively. For each relevant comparison between two treatments (or group types within a treatment), I run the regression using data only from the two treatments under consideration.

Table A9: Test of Hypothesis 3 for *Before - H* vs *Random*

	Avg. Contr.	
Agr. Before H	4.344	*
	(2.530)	
Round Num.	-0.712	***
	(0.047)	
Intercept	11.916	***
	(1.824)	
σ_u	7.063	***
	(0.913)	
σ_e	4.171	***
	(0.160)	
ρ	0.741	
	(0.051)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *Before H* and *Random Before*. *Random Before* is the omitted dummy.

Table A10: Test of Hypothesis 3 for *Before* - *L* vs *Random*

	Avg. Contr.	
Agr. Before L	5.194	**
	(2.413)	
Round Num.	-0.720	***
	(0.049)	
Intercept	11.955	***
	(1.746)	
σ_u	6.719	***
	(0.872)	
σ_e	4.354	***
	(0.167)	
ρ	0.704	
	(0.056)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *Before L* and *Random Before*. *Random Before* is the omitted dummy.

Table A11: Test of Hypothesis 3 for *Before* - *H* vs *L*

	Avg. Contr.	
Agr. Before H	-0.812	
	(2.884)	
Round Num.	-0.632	***
	(0.114)	
Intercept	16.452	***
	(1.945)	
σ_u	7.801	***
	(1.030)	
σ_e	4.210	***
	(0.329)	
ρ	0.774	
	(0.045)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *Before H* and *Agreeableness Before L*. *Agreeableness Before L* is the omitted dummy.

Table A12: Test of Hypothesis 3 for *After - H* vs *Random*

	Avg. Contr.	
Agr. After H	2.179	
	(2.540)	
Round Num.	-0.781	***
	(0.047)	
Intercept	13.352	***
	(1.831)	
σ_u	7.088	***
	(0.915)	
σ_e	4.217	***
	(0.160)	
ρ	0.739	
	(0.051)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *After H* and *Random After*. *Random After* is the omitted dummy.

Table A13: Test of Hypothesis 3 for *After - L* vs *Random*

	Avg. Contr.	
Agr. After L	1.356	
	(2.610)	
Round Num.	-0.846	***
	(0.050)	
Intercept	13.806	***
	(1.885)	
σ_u	7.280	***
	(0.941)	
σ_e	4.414	***
	(0.170)	
ρ	0.731	
	(0.052)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *After L* and *Random After*. *Random After* is the omitted dummy.

Table A14: Test of Hypothesis 3 for *After* - *H* vs *L*

	Avg. Contr.	
Agr. After H	0.853	
	(2.536)	
Round Num.	-0.768	***
	(0.046)	
Intercept	14.579	***
	(1.829)	
σ_u	7.083	***
	(0.913)	
σ_e	4.185	***
	(0.158)	
ρ	0.741	
	(0.051)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *After H* and *Agreeableness After L*. *Agreeableness After L* is the omitted dummy.

Table A15: Test of Hypothesis 3 for *Never* - *H* vs *Random*

	Avg. Contr.	
Agr. Never H	1.654	
	(1.630)	
Round Num.	-0.648	***
	(0.040)	
Intercept	10.077	***
	(1.194)	
σ_u	4.500	***
	(0.602)	
σ_e	3.651	***
	(0.132)	
ρ	0.603	
	(0.066)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *Never H* and *Random After*. *Random After* is the omitted dummy.

Table A16: Test of Hypothesis 3 for *Never - L* vs *Random*

	Avg. Contr.	
Agr. Never L	1.590	
	(2.006)	
Round Num.	-0.715	***
	(0.047)	
Intercept	10.487	***
	(1.463)	
σ_u	5.555	***
	(0.737)	
σ_e	4.123	***
	(0.155)	
ρ	0.645	
	(0.063)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *Never L* and *Random After*. *Random After* is the omitted dummy.

Table A17: Test of Hypothesis 3 for *Never - H* vs *L*

	Avg. Contr.	
Agr. Never H	0.159	
	(1.927)	
Round Num.	-0.712	***
	(0.048)	
Intercept	12.028	***
	(1.410)	
σ_u	5.326	***
	(0.699)	
σ_e	4.288	***
	(0.158)	
ρ	0.607	
	(0.065)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *Never H* and *Agreeableness Never L*. *Agreeableness Never L* is the omitted dummy.

Table A18: Test of Hypothesis 4 for *H* - *Before* vs *After*

	Avg. Contr.	
Agr. Before H	1.260	
	(2.690)	
Round Num.	-0.670	***
	(0.044)	
Intercept	14.661	***
	(1.933)	
σ_u	7.529	***
	(0.967)	
σ_e	4.014	***
	(0.149)	
ρ	0.779	
	(0.046)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *Before H* and *Agreeableness After H*. *Agreeableness After H* is the omitted dummy.

Table A19: Test of Hypothesis 4 for *L* - *Before* vs *After*

	Avg. Contr.	
Agr. Before L	2.966	
	(2.643)	
Round Num.	-0.735	***
	(0.049)	
Intercept	14.304	***
	(1.907)	
σ_u	7.379	***
	(0.952)	
σ_e	4.399	***
	(0.167)	
ρ	0.738	
	(0.051)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *Before L* and *Agreeableness After L*. *Agreeableness After L* is the omitted dummy.

Table A20: Test of Hypothesis 5 for *H* - *After* vs *Never*

	Avg. Contr.	
Agr. After H	2.777	
	(2.063)	
Round Num.	-0.680	***
	(0.043)	
Intercept	11.965	***
	(1.496)	
σ_u	5.742	***
	(0.745)	
σ_e	3.914	***
	(0.142)	
ρ	0.683	
	(0.058)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *After H* and *Agreeableness Never H*. *Agreeableness Never H* is the omitted dummy.

Table A21: Test of Hypothesis 5 for *L* - *After* vs *Never*

	Avg. Contr.	
Agr. After L	2.186	
	(2.448)	
Round Num.	-0.812	***
	(0.052)	
Intercept	12.702	***
	(1.774)	
σ_u	6.809	***
	(0.884)	
σ_e	4.569	***
	(0.175)	
ρ	0.689	
	(0.057)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *After L* and *Agreeableness Never L*. *Agreeableness Never L* is the omitted dummy.

Table A22: Test of Hypothesis 4 for *Random - Before* vs *After*

	Avg. Contr.	
Random Before	-0.919	
	(2.374)	
Round Num.	-0.832	***
	(0.050)	
Intercept	13.717	***
	(1.721)	
σ_u	6.603	***
	(0.859)	
σ_e	4.376	***
	(0.171)	
ρ	0.695	
	(0.057)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Random Before* and *Random After*. *Random After* is the omitted dummy.

Table A23: Test of Hypothesis 5 for *Random - After* vs *Never*

	Avg. Contr.	
Random After	2.416	
	(2.200)	
Round Num.	-0.751	***
	(0.045)	
Intercept	10.764	***
	(1.593)	
σ_u	6.118	***
	(0.803)	
σ_e	3.965	***
	(0.150)	
ρ	0.704	
	(0.056)	
N	480	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Random After* and *Random Never*. *Random Never* is the omitted dummy.

D.1.2 Table 7

I use a multilevel panel Tobit regression (censored at 0 and 25) with random effects at the individual and group levels. This regression incorporates all elicited personality characteristics and demographic data, along with controls for treatment effects and the average group contribution from the previous period.[29] I exclude the *Agreeableness Before* treatment data from this regression, as I anticipated misrepresentation in Agreeableness in this treatment.[30]

[29]Following Bardsley and Moffatt (2007), the initial lagged contribution in period 1 is found using a grid search.

[30]Despite the absence of evidence for misrepresentation, I adhere to the pre-registered analysis. However, I relax this approach in the exploratory Section 3.2.

Table A24: Individual Contributions Regression

	Contribution	
Random After	0.814	
	(2.976)	
Random Never	-2.653	
	(2.972)	
Agr. After H	2.183	
	(2.977)	
Agr. Never H	0.159	
	(2.978)	
Agr. After L	0.877	
	(3.016)	
Agr. Never L	-0.063	
	(2.969)	
Round Num.	-0.929	***
	(0.037)	
Avg.Gr.Cont.Prev.Rnd	0.550	***
	(0.050)	
Agreeableness	0.004	
	(0.043)	
Open Mindedness	0.286	***
	(0.092)	
Negative Emotionality	0.112	
	(0.089)	
Extraversion	0.130	
	(0.093)	
Conscientiousness	-0.265	***
	(0.090)	
Honesty Humility	-0.069	
	(0.125)	
Machiavellianism	-0.093	
	(0.152)	
Narcissism	-0.129	
	(0.115)	
Psychopathy	-0.085	
	(0.153)	
Female	-0.586	
	(0.819)	
Year At Uni		
2nd Year	0.246	
	(1.422)	
3rd Year	-2.336	*
	(1.397)	
4th Year+	-1.670	
	(1.411)	
Grad. Student	-2.261	
	(1.526)	
GPA	0.266	
	(0.389)	
Field Of Study		
Economics	-0.065	
	(1.479)	
Arts & Hum.	1.778	
	(1.968)	
Nat. Science	3.287	*
	(1.698)	
Education	3.226	
	(2.264)	
Engineering	4.079	
	(2.676)	
Law	0.093	
	(2.308)	
Soc. Science	0.628	
	(1.847)	
Medicine	0.178	
	(2.361)	
Other	1.047	
	(1.636)	
Intercept	9.061	
	(7.443)	
Var(Group)	60.346	
	(9.776)	
Var(Subject)	20.559	
	(2.860)	
Var(Contribution)	75.658	
	(2.328)	
N	5040	

*** p<.01, ** p<.05, * p<.1. Notes: Multi-level Tobit Regression on contributions to the public good at the subject level censored at 0 and 25, with subject- and group-level random effects. *Agreeableness* treatments abbreviated as *Agr.* Avg.Gr.Cont.Prev.Rnd. is the average contributions of the other two subjects in the group in the previous round, where the first round value is found by a gridsearch that maximizes the log-likelihood (Bardsley and Moffatt, 2007). Grad., Hum., Nat., and Soc. abbreviate Graduate, Humanities, Natural, and Social respectively. All treatments except for *Agreeableness Before* (as they could exhibit misrepresentation) are included in this regression.

D.1.3 Table A1

	Agreeableness	
Honesty Humility	0.449	**
	(0.200)	
*Agr. Before*xHH	0.030	
	(0.426)	
Agr. Before	1.825	
	(5.661)	
Intercept	99.414	***
	(2.590)	
Var(Agreeableness)	135.857	
	(9.278)	
N	432	

*** p<.01, ** p<.05, * p<.1

Notes: Tobit Regression on Agreeableness at the subject level censored at 26 and 130. *Agr. Before* X HH is an interaction term between the *Agreeableness* (*Agr.*) *Before* treatment and the Honesty Humility personality trait. All treatments are included in this regression.

	Agreeableness	
Machiavellianism	-1.028	***
	(0.179)	
Agr. Before X MM	0.437	
	(0.369)	
Agr. Before	-1.292	
	(3.281)	
Intercept	113.556	***
	(1.605)	
Var(Agreeableness)	127.198	
	(8.687)	
N	432	

*** p<.01, ** p<.05, * p<.1

Notes: Tobit Regression on Agreeableness at the subject level censored at 26 and 130. *Agr. Before* X MM is an interaction term between the *Agreeableness* (*Agr.*) *Before* treatment and the Machiavellianism personality trait. All treatments are included in this regression.

	Agreeableness	
Narcissism	-0.255	
	(0.178)	
Agr. Before X NN	0.415	
	(0.383)	
Agr. Before	-2.347	
	(4.579)	
Intercept	107.986	***
	(2.140)	
Var(Agreeableness)	137.196	
	(9.370)	
N	432	

*** p<.01, ** p<.05, * p<.1

Notes: Tobit Regression on Agreeableness at the subject level censored at 26 and 130. *Agr. Before* X NN is an interaction term between the *Agreeableness* (*Agr.*) *Before* treatment and the Narcissism personality trait. All treatments are included in this regression.

	Agreeableness	
Psychopath	-1.918	***
	(0.179)	
*Agr. Before*xPP	0.516	
	(0.422)	
Agr. Before	-2.250	
	(3.538)	
Intercept	120.689	***
	(1.566)	
Var(Agreeableness)	106.318	
	(7.261)	
N	432	

*** p<.01, ** p<.05, * p<.1

Notes: Tobit Regression on Agreeableness at the subject level censored at 26 and 130. *Agr. Before* X PP is an interaction term between the *Agreeableness* (*Agr.*) *Before* treatment and the Psychopathy personality trait. All treatments are included in this regression.

D.2 Exploratory Analysis

D.2.1 Misrepresentation and Suspicion Robustness Checks

Table A25: Tobit Robustness Regressions on Misrepresentation and Suspicion Treatment Effects

	Agr. Before *Random Before*		*Agr. Before* *After & Never*		*Random Before* *After & Never*	
Agr. Before	2.441		2.410	*		
	(2.098)		(1.311)			
Random Before					-0.032	
					(1.875)	
Intercept	105.021	***	105.053	***	105.053	***
	(1.765)		(0.702)		(0.694)	
Var(Agreeableness)	131.805		136.884		141.777	
	(15.191)		(10.566)		(11.940)	
N	144		384		336	

*** p<.01, ** p<.05, * p<.1

Notes: Tobit Regression on Agreeableness at the subject level censored at 26 and 130, with bootstrapped standard errors. Regression includes only a treatment dummy, and drops all observations from other treatment groups. *Agreeableness* treatments abbreviated as *Agr.*

D.2.2 Table 8

Table A26: Regression with Pooled H and L Groups - *Before*

	Avg. Contr.	
Agr. Before	4.760	**
	(2.239)	
Round Num.	-0.688	***
	(0.039)	
Intercept	11.722	***
	(1.851)	
σ_u	7.218	***
	(0.761)	
σ_e	4.247	***
	(0.132)	
ρ	0.743	
	(0.041)	
N	720	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *Before* and *Random Before*. *Random Before* is the omitted dummy.

Table A27: Regression with Pooled *H*, *L*, and *Random* Groups - *Before*

	Avg. Contr.	
Agr. Before	4.935	***
	(1.582)	
Round Num.	-0.713	***
	(0.030)	
Intercept	11.742	***
	(1.026)	
σ_u	6.837	***
	(0.561)	
σ_e	4.139	***
	(0.099)	
ρ	0.732	
	(0.033)	
N	1200	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *Before* and all *Random* treatments. *Random* is the omitted dummy.

Table A28: Regression with Pooled *H* and *L* Groups - *After*

	Avg. Contr.	
Agr. After	1.762	
	(2.221)	
Round Num.	-0.797	***
	(0.039)	
Intercept	13.465	***
	(1.837)	
σ_u	7.155	***
	(0.754)	
σ_e	4.272	***
	(0.133)	
ρ	0.737	
	(0.042)	
N	720	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *After* and *Random After*. *Random After* is the omitted dummy.

Table A29: Regression with Pooled *H*, *L*, and *Random* Groups - *After*

	Avg. Contr.	
Agr. After	2.842	*
	(1.511)	
Round Num.	-0.768	***
	(0.029)	
Intercept	12.168	***
	(0.982)	
σ_u	6.523	***
	(0.536)	
σ_e	4.120	***
	(0.099)	
ρ	0.715	
	(0.035)	
N	1200	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *After* and all *Random* treatments. *Random* is the omitted dummy.

Table A30: Regression with Pooled *H*, *L*, and *Random* Groups, and Pooled *Agreeableness Before & After* Treatments

	Avg. Contr.	
Agr. Before & After	3.893	***
	(1.347)	
Round Num.	-0.729	***
	(0.025)	
Intercept	11.861	***
	(1.036)	
σ_u	6.961	***
	(0.482)	
σ_e	4.154	***
	(0.084)	
ρ	0.737	
	(0.028)	
N	1680	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *Before*, *Agreeableness After*, and all *Random* treatments. *Random* is the omitted dummy.

Table A31: Regression with Pooled H and L Groups, and Pooled *Agreeableness Before & After* Treatments

	Avg. Contr.	
Agr. Before & After	3.515	**
	(1.510)	
Round Num.	-0.705	***
	(0.027)	
Intercept	12.047	***
	(1.250)	
σ_u	6.881	***
	(0.514)	
σ_e	4.234	***
	(0.091)	
ρ	0.725	
	(0.031)	
N	1440	

*** p<.01, ** p<.05, * p<.1

Notes: xttobit regression on the Average Group Contribution in Tokens (Avg. Contr.) with random-effects at the group level, censored at 0 and 25. Included treatments are *Agreeableness* (*Agr.*) *Before*, *After*, and *Never*. *Agreeableness Never* is the omitted dummy.

D.2.3 Table A2

Table A32: Individual Contributions Regression with All Data

	Contribution				
Random After	0.509		Year At Uni		
	(3.134)		2nd Year	0.100	
Random Never	-2.754			(1.228)	
	(3.134)		3rd Year	-0.753	
Agr. Before H	4.897			(1.234)	
	(3.149)		4th Year+	-0.924	
Agr. Before L	6.079	*		(1.191)	
	(3.142)		Grad. Student	-1.851	
Agr. After H	2.180			(1.368)	
	(3.136)		GPA	0.501	
Agr. Never H	-0.165			(0.338)	
	(3.137)		Field Of Study		
Agr. After L	1.020		Economics	0.996	
	(3.169)			(1.387)	
Agr. Never L	-0.085		Arts & Hum.	2.728	
	(3.130)			(1.756)	
Round Num.	-0.891	***	Nat. Science	3.849	**
	(0.032)			(1.564)	
Avg.Gr.Cont.Prev.Rnd	0.535	***	Education	2.946	
	(0.043)			(1.929)	
Agreeableness	0.007		Engineering	4.732	*
	(0.040)			(2.511)	
Open Mindedness	0.254	***	Law	1.197	
	(0.081)			(2.090)	
Negative Emotionality	0.151	**	Soc. Science	1.436	
	(0.076)			(1.683)	
Extraversion	0.152	*	Medicine	1.368	
	(0.083)			(1.921)	
Conscientiousness	-0.275	***	Other	2.030	
	(0.079)			(1.495)	
Honesty Humility	-0.043		Intercept	5.100	
	(0.109)			(6.729)	
Machiavellianism	-0.108		Var(Group)	68.330	
	(0.125)			(9.563)	
Narcissism	-0.051		Var(Subject)	20.800	
	(0.098)			(2.530)	
Psychopathy	-0.113		Var(Contribution)	74.913	
	(0.135)			(2.015)	
Female	-0.297		N	6480	
	(0.709)				

*** p<.01, ** p<.05, * p<.1. Notes: Multi-level Tobit Regression on contributions to the public good at the subject level censored at 0 and 25, with subject- and group-level random effects. *Agreeableness* treatments abbreviated as *Agr.* Avg.Gr.Cont.Prev.Rnd. is the average contributions of the other two subjects in the group in the previous round, where the first round value is found by a gridsearch that maximizes the log-likelihood (Bardsley and Moffatt, 2007). Grad., Hum., Nat., and Soc. abbreviate Graduate, Humanities, Natural, and Social respectively. All treatments are included in this regression.

Table A33: Individual Contributions Regression without *Agreeableness Before* & *After*

	Contribution	
Random After	0.746	
	(2.847)	
Random Never	-2.476	
	(2.837)	
Agr. Never H	0.443	
	(2.846)	
Agr. Never L	0.285	
	(2.838)	
Round Num.	-0.927	***
	(0.045)	
Avg.Gr.Cont.Prev.Rnd	0.510	***
	(0.058)	
Agreeableness	-0.016	
	(0.046)	
Open Mindedness	0.181	*
	(0.104)	
Negative Emotionality	-0.028	
	(0.098)	
Extraversion	-0.019	
	(0.111)	
Conscientiousness	-0.228	**
	(0.103)	
Honesty Humility	-0.159	
	(0.131)	
Machiavellianism	0.014	
	(0.162)	
Narcissism	-0.038	
	(0.119)	
Psychopathy	-0.206	
	(0.163)	
Female	0.291	
	(0.942)	
Year At Uni		
2nd Year	0.342	
	(1.539)	
3rd Year	-2.447	
	(1.527)	
4th Year+	-1.667	
	(1.530)	
Grad. Student	-1.645	
	(1.618)	
GPA	0.318	
	(0.420)	
Field Of Study		
Economics	0.722	
	(1.678)	
Arts & Hum.	4.996	**
	(2.238)	
Nat. Science	3.255	
	(1.997)	
Education	4.255	
	(2.624)	
Engineering	2.071	
	(3.549)	
Law	1.258	
	(2.442)	
Soc. Science	1.555	
	(2.034)	
Medicine	1.675	
	(2.982)	
Other	2.363	
	(1.838)	
Intercept	15.472	**
	(7.573)	
Var(Group)	55.507	
	(10.602)	
Var(Subject)	15.782	
	(2.809)	
Var(Contribution)	77.813	
	(2.846)	
N	3600	

*** p<.01, ** p<.05, * p<.1. Notes: Multi-level Tobit Regression on contributions to the public good at the subject level censored at 0 and 25, with subject- and group-level random effects. *Agreeableness* treatments abbreviated as *Agr.* Avg.Gr.Cont.Prev.Rnd. is the average contributions of the other two subjects in the group in the previous round, where the first round value is found by a gridsearch that maximizes the log-likelihood (Bardsley and Moffatt, 2007). Grad., Hum., Nat., and Soc. abbreviate Graduate, Humanities, Natural, and Social respectively. All treatments except for *Agreeableness Before* and *Agreeableness After* are included in this regression.

Table A34: Individual Contributions Regression without other Big 5 Traits

	Contribution	
Random After	-1.173	
	(2.440)	
Random Never	-4.706	*
	(2.444)	
Agr. Never H	-2.110	
	(2.465)	
Agr. Never L	-2.357	
	(2.443)	
Round Num.	-0.924	***
	(0.037)	
Avg.Gr.Cont.Prev.Rnd	0.564	***
	(0.050)	
Agreeableness	0.017	
	(0.039)	
Intercept	10.103	**
	(4.220)	
Var(Group)	59.980	
	(9.901)	
Var(Subject)	26.611	
	(3.464)	
Var(Contribution)	75.616	
	(2.327)	
N	5040	

*** p<.01, ** p<.05, * p<.1. Notes: Multi-level Tobit Regression on contributions to the public good at the subject level censored at 0 and 25, with subject- and group-level random effects. *Agreeableness* treatments abbreviated as *Agr.* Avg.Gr.Cont.Prev.Rnd. is the average contributions of the other two subjects in the group in the previous round, where the first round value is found by a gridsearch that maximizes the log-likelihood (Bardsley and Moffatt, 2007). Grad., Hum., Nat., and Soc. abbreviate Graduate, Humanities, Natural, and Social respectively. All treatments except for *Agreeableness Before* are included in this regression.

Table A35: Individual Contributions Regression without Tobit Censoring

	Contribution	
Random After	0.237	
	(1.369)	
Random Never	-1.812	
	(1.365)	
Agr. Never H	-0.844	
	(1.377)	
Agr. Never L	-0.317	
	(1.366)	
Round Num.	-0.501	***
	(0.020)	
Avg.Gr.Cont.Prev.Rnd	0.346	***
	(0.027)	
Agreeableness	0.002	
	(0.023)	
Open Mindedness	0.152	***
	(0.050)	
Negative Emotionality	0.040	
	(0.048)	
Extraversion	0.057	
	(0.050)	
Conscientiousness	-0.159	***
	(0.049)	
Honesty Humility	0.000	
	(0.068)	
Machiavellianism	-0.020	
	(0.082)	
Narcissism	-0.040	
	(0.062)	
Psychopathy	-0.083	
	(0.083)	
Female	-0.468	
	(0.444)	
Year At Uni		
2nd Year	0.105	
	(0.773)	
3rd Year	-1.233	
	(0.757)	
4th Year+	-1.082	
	(0.770)	
Grad. Student	-0.924	
	(0.832)	
GPA	0.041	
	(0.212)	
Field Of Study		
Economics	-0.102	
	(0.815)	
Arts & Hum.	1.081	
	(1.080)	
Nat. Science	1.616	*
	(0.929)	
Education	1.576	
	(1.236)	
Engineering	2.822	*
	(1.467)	
Law	0.439	
	(1.260)	
Soc. Science	0.305	
	(1.011)	
Medicine	1.346	
	(1.275)	
Other	0.658	
	(0.900)	
Intercept	10.895	***
	(3.842)	
Var(Group)	1.481	
	(0.078)	
Var(Subject)	0.901	
	(0.066)	
Var(Contribution)	1.715	
	(0.010)	
N	5040	

*** p<.01, ** p<.05, * p<.1. Notes: Multi-level Regression on contributions to the public good at the subject level with subject- and group-level random effects. *Agreeableness* treatments abbreviated as *Agr.* Avg.Gr.Cont.Prev.Rnd. is the average contributions of the other two subjects in the group in the previous round, where the first round value is found by a gridsearch that maximizes the log-likelihood (Bardsley and Moffatt, 2007). Grad., Hum., Nat., and Soc. abbreviate Graduate, Humanities, Natural, and Social respectively. All treatments except for *Agreeableness Before* are included in this regression.

Table A36: Individual Contributions Regression without Other Controls

	Contribution
Agreeableness	0.013
	(0.036)
Intercept	3.384
	(3.923)
Var(Group)	83.141
	(12.690)
Var(Subject)	19.583
	(2.916)
Var(Contribution)	105.722
	(3.286)
N	5040

*** p<.01, ** p<.05, * p<.1. Notes: Multi-level Regression on contributions to the public good at the subject level with subject- and group-level random effects. All treatments except for *Agreeableness Before* are included in this regression.

Table A37: Individual Contributions Regression without Women

	Contribution	
Random After	2.779	
	(4.598)	
Random Never	-9.406	**
	(3.683)	
Agr. Never H	0.229	
	(3.933)	
Agr. Never L	-1.652	
	(3.563)	
Round Num.	-1.280	***
	(0.076)	
Avg.Gr.Cont.Prev.Rnd	0.676	***
	(0.097)	
Agreeableness	-0.037	
	(0.085)	
Open Mindedness	0.615	***
	(0.197)	
Negative Emotionality	-0.057	
	(0.195)	
Extraversion	-0.075	
	(0.213)	
Conscientiousness	-0.648	***
	(0.192)	
Honesty Humility	-0.322	
	(0.300)	
Machiavellianism	0.209	
	(0.292)	
Narcissism	-0.504	**
	(0.249)	
Psychopathy	0.033	
	(0.302)	
Year At Uni		
2nd Year	1.614	
	(3.481)	
3rd Year	-2.682	
	(3.426)	
4th Year+	-1.454	
	(3.544)	
Grad. Student	-2.930	
	(4.019)	
GPA	0.598	
	(1.041)	
Field Of Study		
Economics	2.574	
	(2.600)	
Arts & Hum.	1.214	
	(4.451)	
Nat. Science	8.641	***
	(3.262)	
Education	-1.967	
	(5.222)	
Engineering	11.282	**
	(4.804)	
Law	5.992	
	(4.911)	
Soc. Science	-8.667	*
	(4.575)	
Medicine	0.057	
	(4.995)	
Other	6.732	**
	(3.198)	
Intercept	23.958	*
	(14.267)	
Var(Group)	97.150	
	(21.642)	
Var(Subject)	20.301	
	(6.721)	
Var(Contribution)	122.211	
	(6.521)	
N	2130	

*** p<.01, ** p<.05, * p<.1. Notes: Multi-level Tobit Regression on contributions to the public good at the subject level censored at 0 and 25, with subject- and group-level random effects. *Agreeableness* treatments abbreviated as *Agr.* Avg.Gr.Cont.Prev.Rnd. is the average contributions of the other two subjects in the group in the previous round, where the first round value is found by a gridsearch that maximizes the log-likelihood (Bardsley and Moffatt, 2007). Grad., Hum., Nat., and Soc. abbreviate Graduate, Humanities, Natural, and Social respectively. All treatments except for *Agreeableness Before* are included in this regression. Subjects who reported their gender as female are not included in this regression.

Table A38: Individual Contributions Regression in the First Two Rounds

	Contribution	
Random After	1.116	
	(2.327)	
Random Never	-4.195	*
	(2.286)	
Agr. Never H	0.623	
	(2.356)	
Agr. Never L	0.106	
	(2.290)	
Round Num.	2.160	
	(2.402)	
Avg.Gr.Cont.Prev.Rnd	0.175	
	(0.116)	
Agreeableness	-0.005	
	(0.073)	
Open Mindedness	0.444	***
	(0.168)	
Negative Emotionality	0.019	
	(0.163)	
Extraversion	0.180	
	(0.171)	
Conscientiousness	-0.312	*
	(0.169)	
Honesty Humility	-0.013	
	(0.230)	
Machiavellianism	-0.028	
	(0.275)	
Narcissism	-0.193	
	(0.211)	
Psychopathy	0.028	
	(0.279)	
Female	-3.928	***
	(1.468)	
Year At Uni		
2nd Year	1.747	
	(2.618)	
3rd Year	0.995	
	(2.557)	
4th Year+	1.404	
	(2.563)	
Grad. Student	1.635	
	(2.810)	
GPA	0.154	
	(0.697)	
Field Of Study		
Economics	3.083	
	(2.793)	
Arts & Hum.	8.925	**
	(3.797)	
Nat. Science	6.896	**
	(3.161)	
Education	2.705	
	(4.365)	
Engineering	9.979	*
	(5.277)	
Law	2.744	
	(4.291)	
Soc. Science	3.793	
	(3.327)	
Medicine	1.586	
	(4.387)	
Other	2.454	
	(3.018)	
Intercept	4.930	
	(13.465)	
Var(Group)	19.561	
	(9.428)	
Var(Subject)	84.202	
	(13.407)	
Var(Contribution)	53.197	
	(5.210)	
N	672	

*** p<.01, ** p<.05, * p<.1. Notes: Multi-level Tobit Regression on contributions to the public good at the subject level censored at 0 and 25, with subject- and group-level random effects. *Agreeableness* treatments abbreviated as *Agr.* Avg.Gr.Cont.Prev.Rnd. is the average contributions of the other two subjects in the group in the previous round, where the first round value is found by a gridsearch that maximizes the log-likelihood (Bardsley and Moffatt, 2007). Grad., Hum., Nat., and Soc. abbreviate Graduate, Humanities, Natural, and Social respectively. All treatments except for *Agreeableness Before* are included in this regression. Obserations with a Round Num. > 2 are not included in this regression.

Table A39: Individual Contributions Regression in the First Five Rounds

	Contribution	
Random After	1.144	
	(2.478)	
Random Never	-5.225	**
	(2.458)	
Agr. Never H	-1.852	
	(2.490)	
Agr. Never L	-1.341	
	(2.453)	
Round Num.	-1.864	***
	(0.159)	
Avg.Gr.Cont.Prev.Rnd	0.279	***
	(0.075)	
Agreeableness	-0.013	
	(0.057)	
Open Mindedness	0.356	***
	(0.125)	
Negative Emotionality	0.037	
	(0.121)	
Extraversion	0.068	
	(0.126)	
Conscientiousness	-0.267	**
	(0.124)	
Honesty Humility	-0.129	
	(0.170)	
Machiavellianism	-0.226	
	(0.207)	
Narcissism	-0.042	
	(0.157)	
Psychopathy	-0.076	
	(0.206)	
Female	-1.785	
	(1.105)	
Year At Uni		
2nd Year	0.069	
	(1.927)	
3rd Year	-1.926	
	(1.895)	
4th Year+	-0.534	
	(1.913)	
Grad. Student	-1.571	
	(2.074)	
GPA	0.196	
	(0.522)	
Field Of Study		
Economics	1.715	
	(2.030)	
Arts & Hum.	5.071	*
	(2.695)	
Nat. Science	4.829	**
	(2.319)	
Education	3.372	
	(3.105)	
Engineering	8.428	**
	(3.755)	
Law	2.008	
	(3.132)	
Soc. Science	1.145	
	(2.493)	
Medicine	0.855	
	(3.210)	
Other	1.884	
	(2.228)	
Intercept	18.449	*
	(9.415)	
Var(Group)	52.087	
	(10.216)	
Var(Subject)	37.136	
	(5.843)	
Var(Contribution)	68.705	
	(3.565)	
N	1680	

*** p<.01, ** p<.05, * p<.1. Notes: Multi-level Tobit Regression on contributions to the public good at the subject level censored at 0 and 25, with subject- and group-level random effects. *Agreeableness* treatments abbreviated as *Agr.* Avg.Gr.Cont.Prev.Rnd. is the average contributions of the other two subjects in the group in the previous round, where the first round value is found by a gridsearch that maximizes the log-likelihood (Bardsley and Moffatt, 2007). Grad., Hum., Nat., and Soc. abbreviate Graduate, Humanities, Natural, and Social respectively. All treatments except for *Agreeableness Before* are included in this regression. Obserations with a Round Num. > 5 are not included in this regression.

E Instructions

Full Instructions are provided in the OSF project and/or replication packet - either in the oTree code or a .docx file for the paper instructions. Selected parts of the Instructions that are particularly relevant for the understanding of the experiment are presented below.

E.1 Part 1

E.1.1 First Screen

This experiment will have two parts.

Part 1 will be a set of questions about yourself. We ask that you answer these questions accurately.

Part 2 has 15 decision rounds. A brief summary of one decision round follows:

- Subjects are in groups of 3
- Each subject has 25 tokens that they divide between their Private Account or a Cooperation Account
- Each token placed in their Private Account earns that subject 10 points.
- Each token placed in the Cooperation Account earns the entire group 12 points.
- Everyone in the group receives an equal portion of the earnings from the Cooperation account, that is, they earn 12*1/3=4 points per token in the Cooperation Account.

This is only a basic outline of Part 2. More instructions will be provided before starting Part 2.

[$t > 0$ Treatments:]

We will now start with Part 1 - the set of questions about yourself.

E.1.2 Second Screen

[Second Screen only in *Before* Treatments]

[*Random* Treatments:] For Part 2, you will be assigned to a group of three **randomly**.

[*Agreeableness* Treatments:] For Part 2, you will be assigned to a group of three **based on your 'Agreeableness' score. Your Agreeableness score is determined by your responses to particular questions in Part 1.**

Agreeableness is a personality trait where people high in Agreeableness are often described as *selfless, trusting, good-natured, generous, and forgiving.* (Costa, McCrae, & Dembroski, 1989)

In scientific studies, **a high level of Agreeableness has been found to have a positive effect on group cooperation decisions** similar to the type in Part 2.

[References button with pop-up window that states:

Perugini, Tan, & Zizzo in Economic Issues, Volume 15, Part 1, 2010.

Volk, Thöni, & Ruigrok in the Journal of Economic Behavior & Organization, Volume 81, Issue 2, 2012.

Kagel & McGee in Economics Letters, Volume 124, Issue 2, 2014.

Thielmann, Spadaro, & Balliet in Psychological Bulletin, Volume 146, Issue 1, 2020.]

Each group of three is formed from six randomly selected subjects. **The three subjects with the highest Agreeableness scores will be assigned to one group, and the remaining three subjects to the other group.**

[All *Before* treatments:] Each group of three will remain together for all 15 decisions in Part 2.

We will now proceed with Part 1 - the set of questions about yourself.

E.2 Part 2

[$t \leq 1$ Treatments:]

[*Agreeableness* treatments:] For Part 2, you will be assigned to a group of three **based on your 'Agreeableness' score. Your Agreeableness score is determined by your responses to particular questions in Part 1.** Agreeableness is a personality trait where people high in Agreeableness are often described as *selfless, trusting, good-natured, generous, and forgiving.* (Costa, McCrae, & Dembroski, 1989) In scientific studies, **a high level of Agreeableness has been found to have a positive effect on group cooperation decisions** similar to the type in Part 2.

Each group of three is formed from six randomly selected subjects. **The three subjects with the highest Agreeableness scores will be assigned to one group, and the remaining three subjects to the other group.**

[*Random* treatments:] For Part 2, you will be assigned to a group of three **randomly**.

E.3 Part 3

Parts 1 and 2 of the experiment are now complete.

We ask you to fill out a final short survey, before your final earnings are displayed. Your final earnings have already been calculated and set.

There are no further parts to the experiment after this final survey.

F Power Analysis

For each statistical test, I conducted a simulation-based power analysis in order to determine the minimum detectable effect size that attains 80% power given a $\alpha = 0.05$ rejection threshold. The simulations were conducted in the same code as the statistical analysis that was attached to the pre-registration. I report the minimum effect size for each test below.

F.1 Part 1

The outcome of interest is each subject's Agreeableness score, calculated from their responses to the Part 1 questions. Each of the three groups of treatments is tested using a Mann-Whitney test, with each subject being an independent observation. The power analysis for the *Random Before* to *Agreeableness Before* comparison suggests a minimum detectable effect size of 7.7 units, and for the *Agreeableness Before* to *Others* comparison it is 5.1. These minimum detectable effect sizes seem reasonable given they represent an average change of one or two out of the 26 Agreeableness questions being flipped from 1 to 5.

F.2 Part 2

I use the group's average contribution in a period as the dependent variable, and a treatment dummy alongside the period for the independent variables. I use a panel-data Tobit regression for the possible censoring that occurs at 0 and 25 tokens for upper and lower limits respectively. For each relevant comparison between two treatments (or group types within a treatment), I run the regression using only data from the pair that is being considered. I use the simulated data-set that was used in the initial pre-registration to get rough estimates of the period effect, σ_e, and σ_u. Using a Tobit regression as the underlying DGP, a simulation-based power analysis suggests a minimum detectable treatment effect size of around 1.7 tokens.

F.3 Part 3

I combine all of the characteristics elicited in Part 3 into one measure based on how likely it is they would be positively perceived by an observer. That is, I reverse code the Dark Triad as these traits are negative, and leave the coding for Honest-Humility as it is. I call this combined measure 'Positive Perception'. As there are 16 questions elicited on a 5-point Likert scale, it can take a

minimum value of 16, and a maximum of 80. I use a Mann-Whitney test to test for differences between the three relevant subject groups. A power analysis suggests minimum detectable effect sizes of 6.3 and 5.2 respectively.

F.4 Conceptual Replication - MM Research Question 2

A simulation-based power analysis suggests a minimum detectable effect size of around 1 unit.