Gender Differences
in the Development of Other-Regarding Preferences

Katrin John[*]
Leibniz Universität Hannover

Stephan L. Thomsen[†]
Leibniz Universität Hannover, ZEW Mannheim, and IZA Bonn

September 14, 2017

Abstract

We use data from a gender-neutral dictator and public goods game setting to analyze differences in other-regarding preferences between boys and girls aged 10 to 17. The results indicate a higher mean of dictator giving, degree of egalitarian decisions and lower frequency of selfish decisions, free-riding and efficiency concerns for girls. Gender differences are already established at approximately age 10. They cannot be explained by gender-specific increases in other-regarding preferences, differences in dispositions or the impact of personality traits. We conclude that genes and early social learning are the sources of gender differences in other-regarding preferences.

Keywords: gender, other-regarding preferences, personality traits, dictator game, public goods game

JEL Classification: C91, D03, J16

---

[*]email: katrin.john@outlook.com

[†]Contact Address: Leibniz Universität Hannover, Institut für Wirtschaftspolitik, Königsworther Platz 1, D-30167 Hannover, Germany, email: thomsen@wipol.uni-hannover.de, phone: +49 511 762-14596, fax: +49 511 762-4574.

1 Introduction

Gender differences can be observed in many economic situations. Besides differences in human capital, gender-specific preferences and attitudes are relevant factors or explanations for this.[1] One important group of factors are other-regarding preferences. They influence economic decision-making beyond the effect of "hard", measurable outcomes such as interests, returns, wages and so forth that have been used as inputs for the assumed maximization of individual payoffs (see, e.g., Fehr and Fischbacher, 2002). These preferences, therefore, provide an explanation for heterogeneous choices by similar individuals in terms of measurable characteristics when confronted with a decision.

While different inclinations towards risk and competition have been shown to exist for girls and boys at young ages (see, for example, Sutter and Rützler, 2014), this is less obvious for other-regarding preferences. There is evidence both in support of the existence of gender differences in children and adolescents (see, for example, Fehr et al., 2013, or Angerer et al., 2015) and against it (see, for example, Fehr et al., 2008, or Eckel et al., 2011). Furthermore, although adult women have been reported to be more other-regarding than men (Eckel and Grossman, 1998), this result has been shown to be context-specific while maintaining the general conclusion that women are more inequality averse and men are more efficiency-oriented (Croson and Gneezy, 2009).[2] Different expressions of other-regardedness and preferences have economic implications. Whereas the link between labor market outcomes and preferences for risk or competition is quite direct, other-regarding preferences also relate to economic outcomes but in a more subtle way. A high inclination towards cooperation can offer more options for economic interaction compared with a low expectation of reciprocation. Someone who is more generous or altruistic may have a different evaluation of a trade-off between individual and public welfare with consequences for his or her own wealth.[3]

Despite that gender differences in preferences exist and can explain some part of the differences in economic situations, there is only a very limited amount of research attempting to identify the origin and timing of these differences. Beyond genetic disposition, social norms and learning, biological and cultural development during childhood and adolescence may contribute to the distribution of social preferences in adult men and women. Childhood and adolescence constitute sensitive periods for the formation of other-regarding preferences. In particular, understanding how children and adolescents reason about choices and make (allocation) decisions incorporating themselves and others can help to design policies to improve the educational

---

[1]In the labor market, for example, the gender gap in earnings is a widely discussed and analyzed phenomenon. In recent years, researchers have started to consider psychological and socio-psychological factors, e.g., risk aversion and risk attitudes, attitudes towards competition, or the existence of social norms, for explanations in addition to human capital differences and discrimination (see Bertrand, 2011, for an overview).

[2]The ambiguity of earlier research may hint at cultural differences between the samples studied or at a varying and ongoing developmental path of gender differences over the ages studied (with a maximum of 17 years).

[3]Related evidence shows that women have stronger redistributive preferences (that may also interfere with women's success in the labor market); see Bertrand (2011). The literature further reports gender differences in attitudes towards negotiation, where negotiation can be viewed as competition over resource distribution and is therefore influenced by social preferences. Bowles et al. (2005) find that the gender gap in negotiation outcomes depends on whether women negotiate for themselves or for others. In the latter case, women show significantly better performance than in the former.

environment. Moreover, doing so contributes to apprehending (the sources of) adult decision-making.

The aim of this paper therefore is to shed more light on the development of gender differences in other-regarding preferences. For this purpose, we study gender-neutral dictator and public good decisions for a sample of almost 1,000 German children and adolescents at different developmental stages until adulthood (ages 10 to 17 years). The choice of the games was made to analyze two domains of other-regarding preferences. To uncover variations in the propensity to favor payoff maximization (egoism) and fairness considerations (preferences for altruism, egalitarianism or inequality aversion) – the first domain of other-regarding preferences to be analyzed – we played a dictator game (DG).[4] To analyze the second domain of other-regarding preferences, the incentives to free-ride and the inclination towards cooperation or social welfare preferences, we played a public goods game (PGG).

Empirically, we employ a threefold approach. First, the prominent features of the study design include gender neutrality in terms of decisions made, simple choices while maintaining a larger choice set (such that children from 10 to 17 are able to understand), and the use of a large sample. This ensures that a comprehensive database is used for a more in-depth look at the development of gender differences. Moreover, since we conducted one-shot games and the children and adolescents had no experience with strategic behavior in the sense of game theory before, our design should allow us to distinguish altruistic behavior from strategic behavior. Second, we analyze gender differences in other-regarding preferences over a broad age group. Any gender differences obtained from different age groups in the same experimental setting would make it possible to draw a picture of the gender-specific development of other-regarding preferences. Since increasing pro-sociality with age is a stylized fact (see Fehr et al., 2013 or Brocas et al., 2017), the composition of girls and boys and the age group involved may lead to different results in the event that boys and girls develop in a time-displaced manner. Our large sample allows us to test whether increases from one age group to another differ across males and females. Third, since gender differences in personality are already present at the youngest age studied (see Else-Quest et al., 2006), these could in principle lead to the observed gender differences in other-regarding preferences. Using self-reported personality traits, we test whether individual personality contributes to the variance in distributional choices.

Our results from the DG show that girls, on average, give more as dictators and less frequently choose to give nothing. We find no mean difference in public goods contributions between genders. Allocations that involve keeping or contributing a full endowment for the public good are, however, less frequently selected by girls. These results indicate an early emergence of more pronounced levels of generosity and egalitarian preferences among girls and a more pronounced inclination towards efficiency among boys, which are in line with what is known for adults. Regarding the possible link between personality and other-regarding preferences, we advance two hypotheses. One is that underlying differences in personality traits

---

[4]There is a general consensus that distribution decisions in a DG provide a measure of generosity, while the relevance of inequality aversion in these decisions has been disputed (see, e.g., List, 2007; Engelmann and Strobel, 2004; Fehr et al., 2006).

between boys and girls together with personality influencing experimental decisions lead to the observed gender differences in measures of other-regarding preferences. The other is that personality aspects influence distributional decisions differently between genders. Our results show that there is neither evidence for gender-specific increases in other-regardedness over the age interval studied nor for an influence of individual personality. Consequently, gender differences in other-regarding preferences are established early and likely relate to genetic differences or to differences in social norms learned by girls and boys.[5]

The remainder of the paper is organized as follows. Section 2 contains the presentation of the data and experimental design used for the analysis. Section 3 provides the experimental results. A comprehensive robustness analysis is provided in section 4. It contains a regression analysis, considers potential bias due to experiment design and reports on the investigation of potential age and gender differences. Section 5 analyzes to what extent personality and personality differences between genders can explain the findings. Further aspects and explanations are discussed in section 6. The final section provides a short conclusion.

2 Data Set and Experiment Design

2.1 Data Set

The data base for the analysis contains experimental choices of German pupils enrolled in six public secondary schools. Data were collected at two points in time. The first round (12/2011 and 01/2012) encompassed four public schools, two in the federal state of Saxony-Anhalt and two in the federal state of Lower Saxony. In both federal states, we chose non-academic track schools (intermediate secondary track, *Realschule*) and academic track schools (highest-level secondary track, *Gymnasium*). The second round (01/2014) contained two further academic track schools in Lower Saxony. The approximate distance between schools is 150km in the bordering federal states.

For the analysis, we considered pupils enrolled in two to three classrooms in each school of grades 5, 7, 9 and 11 (academic tracks only).[6] Average class sizes vary between 20 and 32 pupils, and classes usually remain intact until pupils leave school. Hence, we expect better knowledge of classmates in the higher grades. Despite sampling from two different federal states, all six schools considered have a very similar institutional setting. All schools are public and non-specialized schools. In addition, the school districts are very similar in regional economic characteristics, e.g., size of region, unemployment rate, share of employees subject to social security insurance contributions.[7] We therefore aggregate the data from both states and both

---

[5]Since heritability has been shown to account only partly for the transmission of other-regarding preferences (Cesarini et al., 2009), and moreover, the social learning of gender differences is known to start very early (Bussey and Bandura, 1999), the latter transmission mechanism is deemed more important.

[6]In non-academic track schools, pupils graduate in grade 10 and fulfill the requirements to enter apprenticeship training.

[7]Three of the four places considered in the first round are classified as semi-urban with 24,000 to 53,000 inhabitants, while the fourth is a city with approximately 100,000 inhabitants. Unemployment rates differ only marginally, with three places having a rate of 6% and one having 4%. The share of employees subject to social insurance contributions also shows little variation, ranging from 30 to 35%. The two schools from the second round

rounds.

The analysis sample includes 985 pupils. The composition by gender and grade is presented in Table 1. To sketch some further characteristics of the sample, in total, 16% are only children with school means ranging from 10 to 20%. The mean share of pupils living with a single parent is 16%. Five of the six school means are in the range of shares between 9 and 18%; one school has a share of 33%.

< Include Table 1 about here >

Relying on pupils for the study of other-regarding preferences has several advantages. First, pupils in our sample were unaware of the experimental procedure. Second, they were also highly unlikely to have been introduced to economic game theory. This makes potential bias in the results due to informed strategic behavior negligible. Moreover, we do not face distortions in the measurement of other-regarding preferences expressed in experimental outcomes, as can be the case for university student samples (see Anderson et al., 2013). Third, our sample is non-selective with respect to incentives, as pupils were unaware of the incentives until the beginning of the experiments. Most samples used to study other-regarding preferences rely on university students, who are likely to self-select due to monetary incentives (see Abeler and Nosenzo, 2015).

2.2 Experiment Design

Pupils participated in one-shot dictator and public goods games that were carried out during normal school hours.[8] The sequence of games varied randomly across sessions. Experiments were conducted by the same persons at all six schools, ensuring that conditions were comparable. In the second round of the experiments, two additional experimenters were involved. Experimental sessions were conducted in one grade at a time by randomly allocating pupils from all classrooms to three to four sessions of approximately 45 minutes in length with a mean size of 16 pupils.[9] Due to the random assignment of pupils from the grade, each session's group consisted of pupils from several classes. Sessions took place in classrooms with individual seating such that decisions could be made in private. Teachers were sometimes present during the experiment. Participation rates ranged from roughly 60 to 90% across the six schools, and non-consent from parents cannot be distinguished from cases of illness.[10] Before the experimental sessions, all pupils completed a questionnaire on personality and a background questionnaire

---

are located in the major city of the county where the two schools from the first round are located. The major city has approximately 514,000 inhabitants. The share of employees subject to social insurance contributions is 35%. The unemployment share is 5%. (All numbers refer to the sampling years. Sources: Federal Statistical Office Germany, 2012b, for the population data, Federal Statistical Office Germany, 2012a, for the number of employees, and German Federal Employment Agency, 2012, for the number of unemployed.)

[8]Detailed instructions for both games can be found in appendix B.

[9]Session sizes ranged from 5 to 23 pupils, with a median session size of 16, and fewer than 4% of the participants were in a session with fewer than 10 pupils. In our analysis, we control for the differing session sizes.

[10]The only way that there may be a systematic influence of the participation decision is via parental background. More educated parents may be more concerned about data confidentiality. Similarly, less educated parents may not support the research and therefore refuse their consent. Despite these hypothetical concerns, a sample of pupils from public schools will suffer from selection to a much lesser extent than a university student sample, where, for example, the educational background is more homogeneous.

that covered questions on pupils' socio-demographic characteristics, such as gender, date of birth and the type of household in which they lived.

Dictator Game (DG): To uncover variations in the propensity to favor payoff maximization (egoism) and fairness considerations (preferences for altruism, egalitarianism or inequality aversion) – the first domain of other-regarding preferences to be analyzed – we played a DG. This game was administered as follows: half of the pupils in a session were given 10 tokens. Pupils in grade 5 (grade 7, grade 9, and grade 11) were told that each token was worth 0.10 EUR (0.15 EUR, 0.30 EUR, and 0.50 EUR, respectively). The increase in incentives was thus oriented to mirror the increase in pupils' pocket money with age. The pupils were then asked to allocate each endowment as desired between themselves and a powerless receiver. After tokens were exchanged for cash, those who were receiver in the first round became dictators but based on a new random assignment of receivers. Role reversal in the DG was employed to establish some equity in payoffs for the children. Especially for the younger children, we expected only small shares given in the DG. Then, playing only one role would have led to larger differences in individual payoffs between those who were dictators only and those who were receivers only.[11]

Importantly, the dictators were not given any information about the receivers. In particular, pupils were therefore not aware of the receiver's gender, which may affect behavior. This design is intended to result in gender-neutral decisions being made. However, the assumption of gender-neutrality may not hold if the shares of girls and boys in a school were very unequal.[12] Since the shares of girls in all considered schools vary between 41 and 55%, we do not consider this an obstacle to the analysis at hand. A further caveat would arise if girls and boys treated the ambiguity of the receiver's gender differently. Then, gender-neutrality may, again, not hold. Nevertheless, it is not possible to give a prior on the direction or size of the potential bias in a dictator giving due to differences in responses to gender ambiguity, particularly with respect to the age groups in consideration. We therefore perceive our measure of fairness considerations to be plausibly gender-neutral.

Public Goods Game (PGG): To analyze the second domain of other-regarding preferences, the incentives to free-ride and the inclination towards cooperation or social welfare preferences, we played a PGG. In this game, all pupils in a session received an endowment of ten tokens and were told to determine how to allocate them. They were allowed to keep tokens in exchange for cash as indicated above, or they could contribute tokens to a public good. Contributions made to the public good were doubled and distributed evenly among all participants. Feedback in the

---

[11]Charness and Rabin (2002) apply role reversal and argue that this does not influence their results. Burks et al. (2003) investigate role reversal in the trust game and conclude that playing both roles reduces trust and reciprocity. However, giving was still positive. Therefore, playing both roles provides lower bound measurements of trust and reciprocity in the population.

[12]Akerlof and Kranton (2000) and Akerlof and Kranton (2002) are the first to address how the concept of identity affects behavior in economics. In particular, gender identity is associated with specific behavioral descriptions for men and women that reduce individual utility if violated. The schooling environment has been shown to affect gender identity, where adolescent girls in a co-educational environment could see their traditional female identity reinforced as they attempt to be attractive to the surrounding boys and compete with other girls. For this reason, gender composition has to be considered.

PGG is implemented by publicly announcing the contribution at the group level.[13] Since we do not repeat the PGG, feedback can only have an influence on giving in the DG if the PGG was played as the first game. We will account for clustering at the session level to correct standard errors for this and related issues.

The two simple games were chosen to be easily explained to pupils in a few words, even to the youngest ones. The experimental design was driven by the practicality of being played in classrooms rather than in a laboratory. Several aspects such as playing both roles in the DG, changing the session size, the public announcement of the groups' contribution in the PGG or paying pupils after each game originate from this. Payment at the end of all decisions would have strained the patience of the children and could well have influenced decision-making according to the level of patience, which is also age-related (Bettinger and Slonim, 2007). Bauer et al. (2014) argue that paying children after each decision ensures concentration and motivation. Therefore, payment after each decision was implemented to ensure pupils of real payoffs. On average, pupils received 25.7 EUR in pocket money per month, but the means vary across schools from 21.2 to 29.5 EUR. Fifth graders on average receive 16.8 EUR, seventh graders 22.6 EUR, ninth graders 31.5 EUR and eleventh graders 44.8 EUR per month. With respect to the available endowment of 1 EUR, 1.5 EUR, 3 EUR and 5 EUR this results in a slightly decreasing relationship, i.e., fifth graders, on average, had higher levels of monthly pocket money relative to their available endowment. The experiments were further not double blind. Engel (2011), however, reports that in DGs, double blind protocols do not significantly change giving.

Ultimately, the experimental design was adapted to the requirements of the analysis group and therefore differs partly from previous research. Nevertheless, these aspects do not pose a problem in light of our research question, as the experimental conditions were the same in all schools and for all children and adolescents.

3 Experimental Results

Table 2 presents the results of the two experiments. For the DG, we report the mean number of tokens allocated by dictators, the tendency to give nothing (DG=0) and the share of egalitarian allocations (DG=5). For the PGG, we present the mean contribution, the share of free-riders (PGG=0) and the share of socially efficient allocations (PGG=10). In particular, shares of DG=0 and DG=5 denote the importance of completely selfish and egalitarian decisions, whereas a share of PGG=10 can be interpreted as an indicator of efficiency concerns.

< Include Table 2 about here >

DG: Girls show a higher mean rate of dictator giving than boys (+0.32 tokens, p=0.02, Wilcoxon rank sum test). This difference is driven by a much lower share of girls who give

---

[13]Information feedback is usually only used in experiments where repetition of the PGG is used. In those designs, the effect of the feedback is not large. For example, Weimann (1994) observes no difference between when the average contribution (comparable to our group contribution) is announced and when it is not.

nothing to the recipient. Whereas almost every fourth boy keeps the full endowment for himself, this is the case only for every tenth girl (p=0.00). Egalitarian allocations are selected equally often by both genders (p=0.90). These results therefore suggest a lower incidence of selfishness among girls. Furthermore, the egalitarian preferences are similar across genders.

PGG: Mean public good contributions are not found to vary by gender (p=0.20). By contrast, the share of free-riders is more than twice as large for boys (at 9%) as it is for girls (4%, p=0.00). Additional gender differences are observable in the share of those selecting socially efficient allocations of contributing the full endowment to the public good. This is twice as common for boys as for girls (18 and 9%, p=0.00). Hence, despite boys presenting a lower preference for cooperation, they place more emphasis on efficiency concerns.

< Include Figure 1 about here >

The distribution of choices over the total range of possible allocations in the DG (see Figure 1) shows that gender differences pertain only to some selected aspects and do not span the entire distribution. Testing the equality of distributions with a Mann-Whitney test results in gender-specific distributions of dictator giving and giving to the public good ($p = 0.00$ for both). Regarding the PGG, girls place much less emphasis on the tails of the distribution compared to boys. A likely explanation for the large discrepancy between genders in the shares contributing the full endowment to the public good relates to risk aversion. Girls have been shown to be more risk-averse than boys on average (Booth and Nolen, 2012b). In PGGs, conditional cooperation has been shown to affect the level of cooperation. Even in a one-shot game, a large contribution can indicate the expectation that the other person will contribute substantially. More risk-averse individuals will expect less conditional cooperation and therefore will have a lower inclination to full and unconditional cooperation.

4 Robustness Analysis

4.1 Multivariate Regression Framework

To check the robustness of the observed gender differences against other factors that may influence the decisions made, we analyze them within a multivariate regression framework. We focus the analysis on outcomes that have a significant mean difference in the experiments: hence, we consider DG, DG=0, PGG=0, and PGG=10 in the following. Table 3 provides coefficient estimates from linear probability models of the girls' indicator for different specifications.[14] For each outcome of interest, we present results from four different specifications.

The first specification (1) controls for grade indicators in addition to the gender indicator. These are expected to capture the development over adolescence. The second specification (2)

---

[14]The marginal effects from a logistic regression and coefficients from the linear probability model are very similar.

8

further adds controls for the experimental design (the session size, sequence of games, and sequence of roles in the DG) and for school fixed effects.[15] Specification (3) augments the first model by an indicator for being in the academic school track. This robustness check is intended to uncover whether school track choice matters for the development of gender differences. School tracks have been related to the development of other-regarding preferences (see John and Thomsen, 2015). Therefore, gender differences in preferences may also relate to influences of the school type. The last specification (4) is based on specification (2) but adds an indicator for the receiver's gender in the DG. Since interaction in the PGG was at the level of the group, receiver gender does not matter; PGG results are therefore not displayed. Because in the DG gender information was not revealed to the subjects, it should not influence the contribution results. Specification (4) accordingly provides a test of the gender neutrality of the experimental design.

< Include Table 3 about here >

In the four model specifications, the estimated gender difference in mean dictator giving varies between 0.38 and 0.44, compared to 0.32 when only experimental results are considered. The results of the most parsimonious specification (1) confirm all of the gender differences found in the experimental analysis above. Girls give on average roughly 0.38 tokens more as dictators than do boys, and boys are more likely to give nothing than are girls (by approximately 14 percentage points). The probabilities of contributing either nothing or the complete endowment to the public good are significantly lower for girls (-5 and -9 percentage points) than they are for boys. These findings are robust to the inclusion of further control variables in the estimation model (specification (2)). The robustness check shown in specification (3) extends specification (1) to account for possible differences between school types and the related school and peer-group environments (academic track vs. non-academic track). The corresponding coefficient estimates differ little from those in specification (1). We can therefore conclude that gender differences do not vary across school tracks. We also estimated a variant of specification (3) with an additional interaction term between gender and the school track indicator (not shown here). It reveals that the interaction term does not have a significant coefficient estimate, i.e., there is also no school-track-specific gender difference. The last specification (4) tests whether the experimental design was indeed gender neutral. Including a binary indicator for having a female receiver when playing the DG shows that gender differences do not change, and the indicator for the receiver's gender itself is not significant for the DG. For DG=0, we observe that having a female receiver reduces the probability of giving nothing by approximately 5 percentage points (significant at the 10% level). As there is a very large gender discrepancy for DG=0, actual shares of gender pairing may play a role. While 53% of the girls acting as dictators were randomly paired with a girl, only 43% of the boys were. The barely significant negative effect of having a female receiver may also relate to the unequal pairing and, therefore, should be interpreted with caution.

---

[15]Table A.1 in the appendix displays full coefficient estimates for all control variables considered.

4.2 Aspects of the Experimental Design

Aspects of the experimental design are also included in the robustness check of gender differences. The DG was played twice, with students playing each role once. The sequence of roles could matter and distort gender effects. Therefore, we analyzed the results in greater detail for the two subsamples: those who were receivers first and those who were dictators first. We used specification (2) from Table 3 for this robustness check. Table 4 shows the coefficient estimates for the gender dummy. A Welch test for differences in the coefficient estimates indicates that those for the DG are different at a $p = 0.00$ significance level. Coefficients for $DG = 0$ do not differ significantly ($p = 0.12$). Economically, however, the differences between genders for dictator giving depending on the role played first can be neglected. Therefore, we can conclude that the sequence of roles in the DG is not (substantially) important for the gender differences in giving decisions.

< Include Table 4 about here >

4.3 Age and Gender

Increasing other-regarding preferences with age have been shown in the literature (see, for example, Fehr et al., 2013 or, more recently, Brocas et al., 2017).[16] In contrast, different gender effects across the developmental path have been analyzed to a lesser extent. One exception is Martinsson et al. (2011), who test for gender effects across different age groups. Similar to their analysis, we present results from testing for gender differences across age groups. The results obtained from regression analysis including interaction terms between gender and age groups are displayed in Table 5. We further report the corresponding p-values from testing for gender and age effects with F-tests of joint significance.

< Include Table 5 about here >

The results show that pupils become more generous with age, confirming the general result from the literature. Regarding the gender differences, the results are somewhat different from those in Table 3, where no interactions are considered. While giving nothing in both the DG and PGG is still significantly different between girls and boys (but at a lower significance level), the difference for tokens sent as a dictator and the choice to contribute all of the endowment to the public good become insignificant. The interaction terms of being female and grade are themselves not relevant. This can be interpreted as indicating that the gender difference does not change over the analyzed age span. One exception is the interaction between female and grade 9, which shows a marginally significant effect with respect to free-riding on the public good. Regarding the coefficient estimates of the other interaction terms, nevertheless, does not reveal any gender-specific trend in the development of other-regarding preferences. Testing for an overall gender effect shows significant p-values for all four outcomes, albeit only at the

---

[16]Brocas et al. (2017) show that increasing cooperation during childhood and adolescence results from the combination of three factors: the evolution of altruism, the evolution of strategic thinking, and the effect of the group in which subjects are located.

margin ($p = 0.10$) for tokens sent as dictator. As additional robustness check, we estimated another model controlling for potential school-track differences by including grade indicators, an indicator for being in the academic track, and the interaction thereof (results not displayed). A comparison with the results in Table 3 indicates negligible differences in coefficient estimates for being female. Hence, tracking pupils does not impact the gender differences in other-regarding preferences.

For the moment, we can conclude that gender differences tend to already be present and robust at approximately the age of 10 years. Moreover, they do not diverge or converge from age 10 to age 17. Further support for this conclusion is obtained from the test results displayed in Table 5. There are significant gender differences across all grades regarding egoistic choices as dictator. For free-riding on the public good, such differences appear for grades 5 and 11. Males and females in grades 7 and 9 significantly differ in their propensity to choose the efficient contribution to the public good (PGG=10). Finally, regarding tokens sent as a dictator, only seventh graders display significantly different values between genders.

Importantly, it should be kept in mind that the subsamples of grades are much smaller. Therefore, the overall tendency to observe statistically significant grade-specific gender differences provides evidence in support of the main findings. Moreover, the presence of individual age effects for males and females for (almost) all outcomes further underlines the independence of age and gender effects in the development of other-regarding preferences.[17] This result – together with the insignificant interaction terms – clearly underlines that the developmental increases in the other-regardedness of males and females are similar between 10 and 17 years of age.

5   Identifying Potential Reasons: Personality and Gender

Given the intermediate conclusion that gender differences are already observable at age 10 and do not change thereafter raises the question of the determinants of these gender differences. One potential source are gender-specific dispositions in terms of personality. Croson and Gneezy (2009) report systematic gender differences in emotional or affective reaction to risk. Moreover, women appear to experience more stress, fear or dread in situations that involve the risk of a negative outcome. Both suggest gender-specific personality dispositions or mechanisms related to preferences.

As mentioned above, before the experimental sessions were played, pupils completed a self-report of personality traits. It refers to the Junior Temperament and Character Inventory (JTCI); see Goth and Schmeck (2009). For the test battery, pupils had to evaluate 103 statements with the help of a 5-point Likert scale. Completion required approximately 20 minutes. The JTCI is based on the bio-psychological personality model of Cloninger (1994) that distinguishes four aspects of temperament and three aspects of character (see Table 6 for a summary). In this model, temperament traits are seen as quasi-automatic emotional reaction tendencies and habits in response to stimuli. Character traits are seen as expressing individual aims and

---

[17]The only exception refers to a female age effect regarding free-riding on the public good.

values, which also depend on socio-cultural learning processes.

< Include Table 6 about here >

Personality is often mapped via the Big Five inventory. De Fruyt et al. (2000) show, in a sample of adults, that all scales from the Temperament and Character Inventory (TCI) "*correlate higher than |0.40| with at least one NEO-PI-R domain scale* (Big Five)". Hence, there is a substantial overlap between the two scales. This allows for a reliable mapping of children's and adolescents' personalities and comparability with results obtained for the Big Five is possible. Moreover, the overlap makes it possible to advance hypotheses with regard to personality traits and other-regarding preferences. Previous experiments with adults suggest a positive association between agreeableness and altruistic giving or cooperation. Since De Fruyt et al. (2000) report a correlation coefficient of |0.51| for agreeableness and cooperativeness, we expect to find a positive relationship between cooperativeness and giving in both games. For the other traits, there is no clear prediction, especially not for a children's sample.

We make use of the JTCI measures to test whether differences in personality between genders exist and whether these can provide an explanation for the overall gender differences in other-regarding preferences. First, we compare mean percentile ranks of the traits for both genders to check for potential gender differences in personality traits (Table 6). The results show distinct means for harm avoidance, reward dependence, self-directedness, cooperativeness, and self-transcendence. With the exception of harm avoidance, these higher values for girls can be interpreted in the direction of having more socially desirable traits than boys. For instance, high values of cooperativeness relate to being empathic, tolerant, and helpful, while low values indicate that someone is rather revengeful and intolerant.

There are two possible channels by which gender differences in distributional outcomes can be related to personality. One is that existing differences in personality traits together with an influence of personality on other-regarding preferences lead to the observed gender differences in outcomes. The other is that males' and females' personalities differently influence distributional choices. In an interacted regression model of personality traits and gender, the first hypothesized channel would result in significant trait coefficients together with differences in trait characteristics of males and females. The second channel would instead result in significant interaction coefficients.

A fully interacted model, however, suffers from multicollinearity. The corresponding variance inflation factors for the interaction terms range from 7 to 11 and thus clearly exceed the usually accepted threshold of 4. To mitigate these problems, we conducted a factor analysis to reduce the number of correlated independent variables. Based on information criteria, three unique factors were retained. The related factor loadings after Varimax rotation are shown below in Table 7.

< Include Table 7 about here >

The first factor loads highest and positively on reward dependence and to a slightly smaller extent on persistence and cooperativeness. The second factor loads highest and positively

12

on harm avoidance and negatively on self-directedness. The third factor loads positively on novelty seeking, albeit to a lesser extent compared with the highest loadings of factors 1 and 2. Personality traits have been shown to significantly differ between boys and girls in five of the seven cases (Table 6). The three factors similarly show differences between girls and boys. For all three factors, boys depart negatively from overall zero means, whereas girls do so positively. Gender differences are largest for factor 1 and smallest for factor 3 but significantly different according to a two-group mean-comparison test (t-test) in all three cases.

Although interpreting these factors is more difficult than interpreting single personality traits, we consider the factors instead of single traits in the regression models. Thereby, we are able to test the two possible channels proposed: whether gender differences in personality combined with an influence of personality affect giving decisions (first channel), and/or whether personality traits have a gender-specific influence on giving decisions (second channel). Table 8 depicts the main and interaction effects of gender and personality measured by the three factors. With respect to our central question, the first important finding is that there is a significant gender influence for all four outcomes when personality and gender interactions are explicitly taken into account.

< Include Table 8 about here >

Giving decisions as a dictator seem to be the most susceptible to influence from personality. Here, factor 1 indicates a positive relationship; factor 3 has a negative correlation. Consequently, reward dependence, persistence and cooperativeness positively correlate with giving in the DG, while novelty seeking negatively correlates with it. Referring to the differences in disposition between the two sexes, girls have a much higher level of cooperativeness. While girls have a mean percentile rank of 55.9, the boys' mean is 43.2. Hence, there is a mean difference of more than 10 percentile ranks. Since girls, on average, display higher dictator giving than boys, the positive association with cooperativeness can be interpreted in favor of the hypothesized transmission of gender differences via a personality differential. Similarly, factor 3 is negatively correlated with dictator giving, which can be related to a gender difference in novelty seeking. Boys have lower scores for that trait and give less as dictators. These results for dictator giving seem to support the hypothesis that gender differences in personality drive gender differences in other-regarding preferences. Nevertheless, such support cannot be extended to giving nothing as a dictator. Therefore, a consistent conclusion of personality differences between genders being the driver of differences in other-regardedness seems to hold only partly.

Regarding the second hypothesis that males' and females' personalities differently influence distributional choices, there are few significant coefficient estimates of the interaction terms in Table 8 and only for the PGG. Girls with high scores on factor 1 free-ride less often. Similarly, girls with high scores on factor 3 also free-ride less often. Full contribution to the public good is less frequently observed for girls scoring high on factor 2. Turning to the tests of overall effects, the p-values of F-tests on the joint significance of coefficients are displayed below the regression results in Table 8. They indicate that there is an effect of gender for all outcomes

expect for dictator giving. An overall effect of personality can only be observed for dictator giving resulting from a significant personality effect for males.

In summary, although there seems to be some relationship between personality and gender that could help to explain gender differences in experimental decisions, the relationship is far from universal. The evidence provided offers limited hints in favor of the first hypothesis, i.e., gender differences in personality drive gender differences in outcomes. The results further provide indications in support of the second hypothesis, i.e., personality differently affects decisions across genders. Hence, there is no clear-cut conclusion supporting or refuting either of the hypotheses concerning the role of personality.

6 Discussion

The analysis thus far has revealed that gender differences in other-regarding preferences as measured by dictator giving and contributions to a public good are robust and exist already for pupils of approximately 10 years of age. Since differences between boys' and girls' experimental decisions neither decrease nor increase over the age span studied, variance in developmental paths does not provide an explanation for the overall gender difference. Moreover although genders differ in personality, these differences cannot be directly linked to observed gender differences in other-regarding preferences. Specifically, neither does personality differently influence experimental decisions by gender, nor do gender differences in personality provide a transmission mechanism for gender differences in outcomes.

This suggests that gender differences in other-regarding preferences develop prior to age 10. Here, genetic factors and non-dispositional factors such as being taught gender-specific norms and values can provide sources of early differences. Existing knowledge regarding genetic sources of other-regarding preferences indicate a partial heritability of these traits (Cesarini et al., 2009). Social learning is another source of early-developing gender differences in other-regarding preferences. As the literature on gender development highlights, gender identities can be observed already at very young ages (see, for example, Bussey and Bandura, 1999). Furthermore, Booth and Nolen (2012a) and (2012b) show that preferences for risk and competition for girls from single-sex schools do not differ from those of boys from coeducational schools. However, they are higher than those of girls from coeducational schools. This could be interpreted as an indication of social learning of preferences. Hence, the social environment can be related to preferences, and genetic disposition is unlikely to be the only source of gender differences in preferences.

As noted above, prior research on gender differences in other-regarding preferences is ambiguous, for children as well as adults. Croson and Gneezy (2009, p. 455) argue that the cause of these inconsistent results is that *"women are more sensitive to social cues in determining appropriate behaviour than are men"*. Assuming that this explanation is also valid for children, our design features of a very simple experimental design and gender neutrality in decisions helps to further validate the gender differences found. Moreover, our study demonstrates that the summarized findings for adults provided by Croson and Gneezy (2009, p. 457) already exist in later childhood and adolescence: *"women are more concerned with equalizing earnings between*

the parties, while men are more concerned with maximizing efficiency". This exactly matches what we find for dictator and public goods allocations. In the DG, efficiency concerns are absent, but a social norm of fairness is evoked. Here, girls have a higher mean giving as dictators compared with boys. In the PGG, efficiency considerations are present, and there is no obvious social norm to contribute the complete endowment that could serve as a "social cue" for girls. Conveniently, boys more often choose the highly efficient allocation of contributing the complete endowment to the public good. Moreover, our findings clearly underline that girls indeed more frequently choose more equal allocations between themselves and the public good than boys do (see Figure 1 above). Thus, the ambiguity of previous research for children may also be explained by design features.

Gender differences in a social context have often been discussed with respect to social desirability (see, for example, Dalton and Ortegren, 2011). A more other-regarding choice in the experiments simultaneously reflects a more socially desirable choice. For the DG, the social norm of fairness is salient. Hence, individuals who are more prone to social desirability may wish to prevent to feeling social disapproval when choosing to deviate from the fairness norm. This mechanism results from internalizing norms and therefore needs no explicit sanctioning. If females respond in a more socially desirable way, this could explain gender differences such as a lower propensity to give nothing as a dictator or to free-ride. The above-mentioned differences in personality traits further support such an interpretation, as girls score higher on socially desirable trait values in their self-reported personality dimensions.

However, an influence from the propensity to respond in a socially desirable way should have manifested itself in a distinct interaction between personality and gender on measures of other-regarding preferences. This was not the case and cannot be confirmed by our results. Moreover, Platow (1994) analyzes self-other choices in light of social desirability bias and does not find a relationship between the two concepts. Even if prosocial choices were obviously socially desirable, subjects' choices were rather reflections of their true social value orientation. Given this result and our findings, we do not expect that there is a social desirability bias in the gender differences we observed.

Another objection to the hypothesis of socially learned gender differences in other-regarding preferences may result from knowledge about gender differences in risk preferences. Women are known to be more risk-averse than men (Croson and Gneezy, 2009), and this disparity has already developed at a young age; see Sutter et al. (2013). The decisions of whether and how much to contribute to the public good are related to risk preferences even in a one-shot game. In particular, the contribution of the full endowment reflects a risky choice to maximize one's own payoff. Since boys choose this socially efficient allocation far more often than girls do in our study, gender differences in risk preferences are an obvious argument to propose.

We do not have the direct information on risk preferences required to directly analyze this potential objection. However, we can use the information on personality traits to approximate risk preferences. Wills et al. (1994) demonstrate for adolescents that novelty seeking highly correlates with the Eysenck scale of risk taking ($\rho = 0.59$). Rao et al. (2011) study how traits of the JTCI relate to a real-stakes decision task to measure risk preferences. They report a

medium correlation ($\rho = 0.31$) for harm avoidance and risky choices in an adolescent sample. Hence, personality traits can be related to risk preferences, although the association seems to be unstable across samples. As shown above, novelty seeking did not differ substantially between boys and girls in our sample, and neither novelty seeking nor harm avoidance have been related to any of the experimental outcomes. Therefore, we conclude that risk preferences are not likely to be particularly important for the observed gender differences in our analysis.

7 Conclusion

We have studied dictator and public goods allocations for boys and girls aged 10 to 17 years. The main findings highlight robust gender differences with respect to mean dictator giving, egoistic dictator choices and decisions to keep or contribute a complete endowment in a PGG. The simple and gender-neutral design of our experiments applied to a large sample of pupils offers a clean opportunity to explore such questions in detail. It shows that already by age 10, differences in choices occur and that these resemble those known for adults. The analysis using interactions between preference outcomes and age as well as personality measures has shown that these differences can neither be explained by age-specific differences in gender nor by differences in individual disposition, as measured by personality traits. Having discussed possible other explanations, it seems most probable that genetic disposition together with early social learning are the predominant sources of the gender differences in other-regarding preferences. Future research should analyze when gender differences in other-regarding preferences emerge and how social learning takes place.

References

Abeler, J. and D. Nosenzo (2015). Self-selection into laboratory experiments: Pro-social motives versus monetary incentives. *Experimental Economics 18*(2), 195–214.

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *Quarterly Journal of Economics 115*(3), 715–753.

Akerlof, G. A. and R. E. Kranton (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature 40*(4), 1167–1201.

Anderson, J., S. V. Burks, J. Carpenter, L. Götte, K. Maurer, D. Nosenzo, R. Potter, K. Rocha, and A. Rustichini (2013). Self-selection and variations in the laboratory measurement of other-

regarding preferences across subject pools: Evidence from one college student and two adult samples. *Experimental Economics 16*(2), 170–189.

Angerer, S., D. Glätzle-Rützler, P. Lergetporer, and M. Sutter (2015). Donations, risk attitudes and time preferences: A study on altruism in primary school children. *Journal of Economic Behavior & Organization 115*, 67–74.

Bauer, M., J. Chytilova, and B. Pertold-Gebicka (2014). Parental background and other-regarding preferences in children. *Experimental Economics 17*(1), 24–46.

Bertrand, M. (2011). New perspectives on gender. *Handbook of Labor Economics 4b*, 1543 – 1590.

Bettinger, E. and R. Slonim (2007). Patience among children. *Journal of Public Economics 91*(1), 343–363.

Booth, A. L. and P. J. Nolen (2012a). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization 81*(2), 542–555.

Booth, A. L. and P. J. Nolen (2012b). Gender differences in risk behavior: Does nurture matter? *The Economic Journal 122*(558), F56–F78.

Bowles, H. R., L. Babcock, and K. L. McGinn (2005). Constraints and triggers: situational mechanics of gender in negotiation. *Journal of Personality and Social Psychology 89*, 951–965.

Brocas, I., J. D. Carillo, and N. Kodaverian (2017). Altruism and strategic giving in children and adolescents. Discussion Paper, University of Southern California.

Burks, S. V., J. P. Carpenter, and E. Verhoogen (2003). Playing both roles in the trust game. *Journal of Economic Behavior and Organization 51*(2), 195–216.

Bussey, K. and A. Bandura (1999). Social cognitive theory of gender development and differentiation. *Psychological Review 106*, 676–713.

Cesarini, D., C. T. Dawes, M. Johannesson, P. Lichtenstein, and B. Wallace (2009). Genetic variation in preferences for giving and risk taking. *The Quarterly Journal of Economics 124*(2), 809–842.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics 117*(3), 817–869.

Cloninger, C. R. (1994). Temperament and personality. *Current Opinion in Neurobiology 4*, 266–273.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–474.

Dalton, D. and M. Ortegren (2011). Gender differences in ethics research: The importance of controlling for the social desirability response bias. *Journal of Business Ethics 103*, 77–93.

De Fruyt, F., L. Van De Wiele, and C. Van Heeringen (2000). Cloninger's psychobiological model of temperament and character and the five-factor model of personality. *Personality and Individual Differences 29*, 441–452.

Eckel, C. C. and P. J. Grossman (1998). Are women less selfish than men?: Evidence from dictator experiments. *The Economic Journal 108*, 726–735.

Else-Quest, N. M., J. S. Hyde, H. H. Goldsmith, and C. A. Van Hulle (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin 132*, 33–72.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics 14*(4), 583–610.

Engelmann, D. and M. Strobel (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution games. *The American Economic Review 94*(4), 857–869.

Federal Statistical Office Germany (2012a). Regionalstatistik, Beschäftigte. Wiesbaden.

Federal Statistical Office Germany (2012b). Regionalstatistik, Bevölkerung. Wiesbaden.

Fehr, E. and U. Fischbacher (2002). Why social preferences matter - the impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal 112*, C1–C33.

Fehr, E., M. Naef, and K. M. Schmidt (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *The American Economic Review 96*, 1912–1917.

Fehr, E., D. Rützler, and M. Sutter (2013). The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. *European Economic Review 64*, 369–383.

German Federal Employment Agency (2012). Arbeitsmarkt in Zahlen, Arbeitsmarktstatistik, Arbeitslose nach Gemeinden. Nürnberg.

Goth, K. and K. Schmeck (2009). *Das Junior Temperament und Charakter Inventar*. Hogrefe.

John, K. and S. L. Thomsen (2015). School-track environment or endowment: What determines different other-regarding behavior across peer groups? *Games and Economic Behavior 94*, 122–141.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy 115*, 482–493.

Martinsson, P., K. Nordblom, D. Rützler, and M. Sutter (2011). Social preferences during childhood and the role of gender and age – An experiment in Austria and Sweden. *Economics Letters 110*, 248–251.

Platow, M. J. (1994). An evaluation of the social desirability of prosocial self-other allocation choices. *The Journal of Social Psychology 134*, 61–68.

Rao, U., T. Sidhartha, K. R. Harker, A. S. Bidesi, L.-A. Chen, and M. Ernst (2011). Relationship between adolescent risk preferences on a laboratory task and behavioral measures of risk-taking. *Journal of Adolescent Health 48*, 151–158.

Schmeck, K., K. Goth, F. Poustka, and R. C. Cloninger (2006). Reliability and validity of the Junior Temperament and Character Inventory. *International Journal of Methods in Psychiatric Research 10*(4), 172–182.

Sutter, M., M. G. Kocher, D. Rützler, and S. T. Trautmann (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *The American Economic Review 103*(1), 510–531.

Sutter, M. and D. Rützler (2014). Gender differences in the willingness to compete emerge early in life and persist. *Management Science 61*(10), 2339–2354.

Weimann, J. (1994). Individual behavior in a free-riding experiment. *Journal of Public Economics 54*(2), 185–200.

Wills, T. A., D. Vaccaro, and G. McNamara (1994). Novelty seeking, risk taking, and related constructs as predictors of adolescent substance use: An application of Cloninger's theory. *Journal of Substance Abuse 6*, 1–20.

Tables and Figures

Table 1: Number of Observations by Gender and Grade

	Grade 5	Grade 7	Grade 9	Grade 11	Total
Girls	126 (45%)	172 (54%)	121 (48%)	57 (43%)	476 (48%)
Boys	155 (55%)	144 (46%)	133 (52%)	77 (57%)	509 (52%)
Total	291 (29%)	318 (32%)	271 (26%)	137 (14%)	985

Table 2: Mean Outcomes in DG and PGG by Gender

	Dictator Game			Public Goods Game			
	DG	DG=0	DG=5	PGG	PGG=0	PGG=10	N
Boys	2.47	0.24	0.13	4.79	0.09	0.18	
	(2.24)	(0.43)	(0.34)	(3.33)	(0.29)	(0.39)	509
Girls	2.79	0.10	0.14	4.54	0.04	0.09	
	(2.02)	(0.30)	(0.34)	(2.57)	(0.19)	(0.28)	476
All	2.62	0.17	0.13	4.67	0.07	0.14	
	(2.14)	(0.38)	(0.34)	(2.99)	(0.25)	(0.34)	985

The table displays means with standard deviations in parentheses. Number of observations refers to individuals having participated in at least one of the DG and PGG.

Figure 1: Tokens sent as dictator and tokens contributed to public good by gender
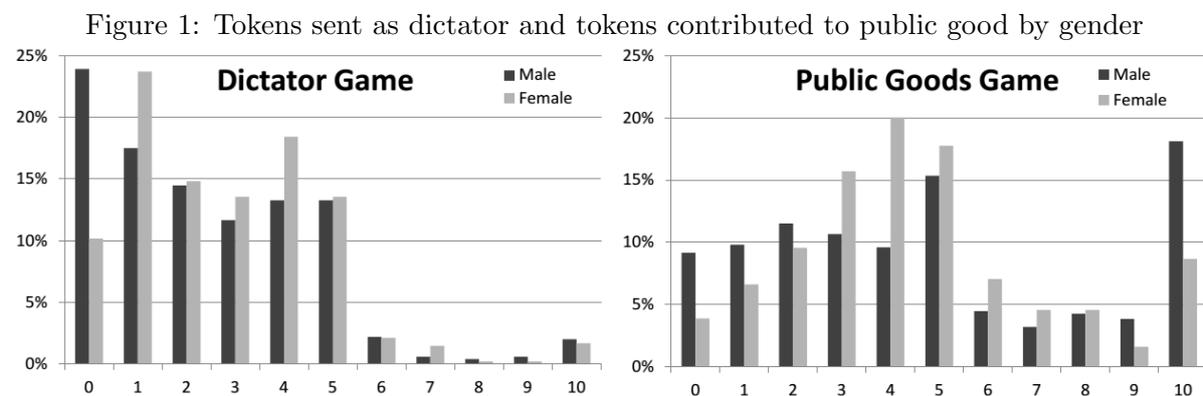
The table displays shares of tokens sent as dictator (left figure, N=969) and shares of contributions to the public good (right figure, N=908) by gender.

Table 3: Gender Effects in Multivariate Analysis

Outcome	Specification			
	(1)	**(2)**	**(3)**	**(4)**
DG (tokens)	0.377**	0.442***	0.408**	0.416**
	(0.163)	(0.159)	(0.159)	(0.166)
DG=0	-0.142***	-0.144***	-0.145***	-0.145***
	(0.032)	(0.032)	(0.032)	(0.035)
PGG=0	-0.053***	-0.053***	-0.055***	-0.060***
	(0.019)	(0.016)	(0.019)	(0.017)
PGG=10	-0.086***	-0.080***	-0.090***	-0.079***
	(0.023)	(0.023)	(0.023)	(0.023)
Control Variables				
Grade indicators	x	x	x	x
Experiment design		x		x
School fixed effects		x		x
Indicator higher track			x	
Gender of receiver in DG				x
N (DG/PGG)	969/908	969/892	969/908	898

The table displays OLS coefficient estimates for an indicator of being female for different outcomes (rows) and specifications (columns) with standard errors in parentheses. Standard errors are clustered at the session level. Specification (4) is displayed for the DG only, as receiver gender is not relevant for PGG decisions.

Table 4: Gender Effects According to Sequence of Roles in the DG

Outcome	**First Receiver**	**First Dictator**
DG (tokens)	0.474**	0.412**
	(0.193)	(0.194)
DG=0	-0.147***	-0.143***
	(0.044)	(0.035)
N	474	495

The table displays OLS coefficient estimates for an indicator of being female for different outcomes (rows) and subsamples (columns) with standard errors in parentheses. Standard errors are clustered at the session level.

Further variables included in the estimates are grade indicators, session size, an indicator for game sequence and indicators for school fixed effects.

Table 5: Interaction Effects Gender and Age

	Dictator Game		Public Goods Game	
	DG	DG=0	PGG=0	PGG=10
Coefficient Estimates from Regressions				
Female	0.279	-0.097*	-0.104**	-0.027
	(0.223)	(0.058)	(0.044)	(0.020)
Grade 7	0.208	0.018	-0.108**	0.124*
	(0.351)	(0.081)	(0.051)	(0.070)
Grade 9	0.704**	-0.127**	-0.144***	0.127*
	(0.309)	(0.060)	(0.044)	(0.067)
Grade 11	1.302***	-0.205***	-0.075	0.299***
	(0.480)	(0.070)	(0.051)	(0.091)
Female*Grade 7	0.219	-0.104	0.069	-0.059
	(0.339)	(0.093)	(0.053)	(0.037)
Female*Grade 9	0.189	-0.045	0.085*	-0.085
	(0.406)	(0.068)	(0.047)	(0.054)
Female*Grade 11	0.323	-0.013	0.028	-0.072
	(0.568)	(0.071)	(0.051)	(0.101)
Testing Gender and Age Effects				
H_0: no overall gender effect[a]	0.10	0.00	0.00	0.01
H_0: no gender effect in grade 5[b]	0.22	0.10	0.02	0.19
H_0: no gender effect in grade 7[c]	0.06	0.01	0.24	0.01
H_0: no gender effect in grade 9[d]	0.18	0.00	0.30	0.03
H_0: no gender effect in grade 11[e]	0.26	0.02	0.00	0.32
H_0: no overall age effect[f]	0.00	0.01	0.01	0.00
H_0: no male age effect[g]	0.03	0.01	0.00	0.00
H_0: no female age effect[h]	0.00	0.00	0.21	0.01

The table displays OLS coefficient estimates with standard errors (clustered at the session level) in parentheses and p-values for testing the following hypothesis after regressions.

[a] $\beta_{female} = \beta_{female*grade7} = \beta_{female*grade9} = \beta_{female*grade11} = 0$

[b] $\beta_{female} = 0$

[c] $\beta_{female*grade7} = 0$

[d] $\beta_{female*grade9} = 0$

[e] $\beta_{female*grade11} = 0$

[f] $\beta_{female} = \beta_{female*grade7} = \beta_{female*grade9} = \beta_{female*grade11}$

[g] $\beta_{grade7} = \beta_{grade9} = \beta_{grade11} = 0$

[h] $\beta_{grade7} + \beta_{female*grade7} = \beta_{grade9} + \beta_{female*grade9} = \beta_{grade11} + \beta_{female*grade11} = 0$

Further control variables are grade indicators, school indicators, session size, an indicator for playing the dictator role before being the receiver, and an indicator for playing DG before PGG.

Table 6: Description of Personality Dimensions in the JTCI

Trait	Concept	Adjectives (low vs. high rank)	Mean Boys	Mean Girls	p-value
TEMPERAMENT					
Novelty Seeking	relates to the behavioral activation system; extent of exploratory activity, approach to novel stimuli, and avoidance of aversive stimuli	reflective, orderly versus impulsive, exploratory, curious	47.0	49.3	0.21
Harm Avoidance	relates to the behavioral inhibition system; extent of being shy, fatiguable, fearful, or pessimistic	confident, outgoing, optimistic versus shy, worrying, pessimistic	44.0	56.3	0.00
Reward Dependence	relates to the behavioral dependence system; extent of regulating one's behavior due to social cues	detached, withdrawn, cold versus attached, sentimental, warm	42.4	61.9	0.00
Persistence	relates to the behavioral persistence system; the intrinsic maintenance of a behavior	gives up easily, inactive versus hard-working, ambitious	54.7	53.1	0.38
CHARACTER					
Self-Directedness	self-concept; relates to responsibility and resourcefulness to achieve personal goals	immature, unreliable versus responsible, reliable	55.9	48.2	0.00
Cooperativeness	extent of feeling a part of the social system; comprises social tolerance, empathy, helpfulness, compassion and moral principles	revengeful, intolerant versus empathic, tolerant, helpful	43.2	55.9	0.00
Self-Transcendence	extent of feeling as part of one's surroundings as a unitive whole; tendency of creative inspiration and spirituality	impatient, self-conscious versus patient, creative, self-forgetful	38.7	45.8	0.00

The description follows Goth and Schmeck (2009, pp. 16-20) and Cloninger (1994, pp. 269-270). Adjectives in the third column are cited from Schmeck et al. (2006), Table 1. Means refer to mean percentile ranks. p-values refer to tests with H_0: no mean difference and H_a: mean difference is different from zero.

Table 7: Factor loadings for Personality Traits

Personality Trait	Factor 1	Factor 2	Factor 3	Uniqueness
Novelty Seeking	-0.20	0.10	0.55	0.65
Harm Avoidance	0.00	0.71	-0.03	0.50
Reward Dependence	0.60	-0.00	0.22	0.59
Persistence	0.56	-0.28	-0.27	0.54
Cooperativeness	0.75	-0.09	-0.17	0.40
Self-Directedness	0.22	-0.69	-0.15	0.45
Self-Transcendence	0.30	0.32	0.19	0.77

The table displays factor loadings after Varimax rotation. N=985.

Table 8: Interaction Effects Gender and Personality

	Dictator Game		Public Goods Game	
	DG	DG=0	PGG=0	PGG=10
Coefficient Estimates from Regressions				
Female	0.329*	-0.140***	-0.055***	-0.087***
	(0.180)	(0.035)	(0.018)	(0.023)
Factor1	0.281**	-0.030	0.033*	-0.018
	(0.136)	(0.027)	(0.018)	(0.024)
Factor2	0.047	0.010	-0.018	0.047**
	(0.115)	(0.020)	(0.021)	(0.019)
Factor3	-0.288*	0.027	0.035	-0.011
	(0.156)	(0.031)	(0.025)	(0.031)
Female*Factor1	-0.007	0.016	-0.045**	0.027
	(0.156)	(0.032)	(0.022)	(0.034)
Female*Factor2	0.017	-0.011	0.009	-0.044**
	(0.165)	(0.023)	(0.025)	(0.022)
Female*Factor3	0.162	-0.002	-0.044*	0.026
	(0.195)	(0.037)	(0.024)	(0.043)
Testing Gender and Age Effects				
H_0: no overall gender effect[a]	0.29	0.00	0.03	0.01
H_0: no overall personality effect[b]	0.02	0.58	0.50	0.93
H_0: no male personality effect[c]	0.06	0.55	0.25	0.07
H_0: no female personality effect[d]	0.85	0.93	0.16	0.16

The table displays OLS coefficient estimates with standard errors (clustered at the session level) in parentheses and p-values for testing the following hypothesis after regressions with F-tests.

[a] $\beta_{female} = \beta_{female*factor1} = \beta_{female*factor2} = \beta_{female*factor3} = 0$

[b] $\beta_{factor1} + \beta_{female*factor1} = \beta_{factor2} + \beta_{female*factor2} = \beta_{factor3} + \beta_{female*factor3} = 0$

[c] $\beta_{factor1} = \beta_{factor2} = \beta_{factor3} = 0$

[d] $\beta_{female*factor1} = \beta_{female*factor2} = \beta_{female*factor3} = 0$

Further control variables are grade indicators, school indicators, session size, an indicator for playing dictator role before being the receiver, indicator for playing DG before PGG.

A   Appendix

Table A.1: Control Variables in Multivariate Analysis

	DG	**DG=0**	**PGG=0**	**PGG=10**
Female	0.442***	-0.144***	-0.053***	-0.080***
	(0.159)	(0.032)	(0.016)	(0.023)
Grade 7	0.311	-0.034	-0.074**	0.097
	(0.279)	(0.055)	(0.032)	(0.059)
Grade 9	0.794***	-0.148***	-0.104***	0.087
	(0.288)	(0.053)	(0.028)	(0.060)
Grade 11	1.449***	-0.214***	-0.060*	0.266***
	(0.347)	(0.057)	(0.035)	(0.065)
DG played before PGG	0.453**	-0.114***	0.053***	0.057
	(0.226)	(0.039)	(0.018)	(0.037)
Dictator role first	0.550***	-0.105***	-0.011	0.006
	(0.121)	(0.029)	(0.014)	(0.014)
Session size	-0.071	-0.008	0.005	-0.027**
	(0.048)	(0.008)	(0.005)	(0.012)
School 1	-0.130	0.060	-0.078**	0.347***
	(0.471)	(0.094)	(0.036)	(0.119)
School 3	-0.224	0.149*	-0.053	0.221**
	(0.451)	(0.087)	(0.046)	(0.091)
School 4	-0.542	0.157*	0.052	0.350**
	(0.454)	(0.087)	(0.051)	(0.165)
School 5	0.839**	-0.033	-0.014	0.168**
	(0.390)	(0.064)	(0.031)	(0.068)
School 6	0.668*	-0.036	-0.065*	0.261***
	(0.378)	(0.061)	(0.038)	(0.074)
Constant	2.296***	0.525***	0.068	0.220
	(0.727)	(0.119)	(0.069)	(0.139)
Adj. R^2	0.208	0.138	0.094	0.173
N	969	969	892	892

The table displays OLS coefficient estimates with standard errors in parentheses for specification (2) from Table 3. Standard errors are clustered at the session level. School 2 is used as a baseline.

B Experiment Course and Instructions

Experiments took place in classrooms where each pupil sat down at a table turned towards the wall or window. Tables were separated with the help of large boards or with baffles such that participants could make their decisions unobserved. Each table had a number to assign pairs in the DG. Experiment sessions took approximately 45 minutes including two dictator games and one public goods game.

General Instructions:

"Please sit down at any place and put your identity card next to the number you find at your desk. Then you can turn towards the middle. You can keep all of your things in your bag, you don't need them, the desks should be empty. Please be quiet so that I can tell you what we will do now. We will play five (three) little games. This will take approximately 45 minutes. During the games, you can earn money. You get the money at the end of each of the games. How much money you earn depends on the decisions you make during the games. Your participation is voluntary. The games do not require any specific skills such as cunning, knowledge or quickness.

Here are some important rules: from now on, you are not allowed to talk to each other. Everyone plays on his or her own. During the games, you'll be equipped with plastic chips. These chips can be converted into money. Each chip is worth (10, 15, 30 or 50) euro cents to you. [*Note: the respective amount for each grade was used.*] At the end of each game, your plastic chips will be exchanged for real money that you can keep. But before that, you need to decide how many you will give to unknown partners. The conditions for that decision change in each game. You'll get an explanation of the conditions at the beginning of each of the games.

With the help of the identity number that you have, we will collect your decisions when we pay you the money. At the end, we will have a large table with the decisions of all pupils from your school who participated. That means we cannot trace back your individual decision because we do not know all the numbers. If you have questions please ask them now. If you have questions during the games, please raise your hand, I will come and answer them privately. Ok, let's start. Please turn around. From now on, communication is not allowed."

Dictator game

"Some of you have two envelopes in front of you. Those who received an envelope are going to make a decision right now; all the others will make the same decision afterwards. To all of you with envelopes, please write down your table number on the darker envelope. Inside the white envelope you will find 10 plastic chips.

Please remember: each chip is worth (10, 15, 30 or 50) euro cents. You are playing together with another person in this room, but you do not know with whom. Please decide now how many chips you would like to keep at your desk and how many you would like to send to your unknown partner. Your unknown partner receives what you decided to allocate without making a decision and without being able to influence the division of chips.

The allocation of partners is random and fixed in advance. Please note, that you will be playing with a different partner every round. You can choose any division of chips you like; there are no restrictions. Now please put the amount that you would like to send to your partner in the empty darker envelope. Please leave the amount that you would like to keep on your desk. If you need more time, please raise you hand.

We will now collect all your darker envelopes and will give them to your partners. Please keep quiet and do not turn around. To all of you who just received an envelope: please open it and put the contents in front of you on your desk. We will now come around to exchange your chips for money. Please do not turn around until I ask you to."

Public goods game

"All of you have two envelopes in front of you. In the white one there are ten green plastic chips. Each of the chips is worth (10, 15, 30, or 50) euro cents. Each of you now has to decide how to divide his or her chips. You can put chips in the darker, empty envelope. Everything that everyone puts in there will be collected from all pupils in that room. The sum of the chips will then be doubled and distributed equally among all of you. Once more, all chips that you put in the darker envelope will be gathered together from all pupils here, doubled, and then distributed equally among you.

That means that if, in the end, we have 40 chips from all of you, for example, we will add another 40 chips. These 80 chips will then be divided by the number of participants in the room. You'll get this amount in addition to what you keep from the white envelope in cash at the end. That is to say, you'll get the amount from the "community pot" no matter how much you put into it. Everyone gets the same share. Please decide now how many chips you will keep at your desk and how many you will put into the community pot. You can choose any division you like. Place the amount you would like to be put into the community pot into the darker, empty envelope. If you need more time to decide, please raise your hand.

We will now collect all your darker envelopes. You can turn around now. Please do not talk. We will now pour out all the chips collected and count them in front of you. (Experimenter counts and announces sum.) We will add the same number of chips and divide them by (number of people in room). For (number of people in room) pupils, each of you will get XX (number of individual) chips from the community pot.[*Note: the number is rounded to an integer.*] Please turn around so that we can come and exchange the chips for money. You'll get the amount from the community pot as well as from the chips that are at your desk."